



Walden University  
**ScholarWorks**

---

Walden Dissertations and Doctoral Studies

Walden Dissertations and Doctoral Studies  
Collection


---

2020

## Ethics of Collection and Use of Consumer Information on the Internet

Thanh M. Pham  
*Walden University*

Follow this and additional works at: <https://scholarworks.waldenu.edu/dissertations>

 Part of the [Databases and Information Systems Commons](#), and the [Library and Information Science Commons](#)

---

This Dissertation is brought to you for free and open access by the Walden Dissertations and Doctoral Studies Collection at ScholarWorks. It has been accepted for inclusion in Walden Dissertations and Doctoral Studies by an authorized administrator of ScholarWorks. For more information, please contact [ScholarWorks@waldenu.edu](mailto:ScholarWorks@waldenu.edu).

# Walden University

College of Management and Technology

This is to certify that the doctoral dissertation by

Thanh M. Pham

has been found to be complete and satisfactory in all respects,  
and that any and all revisions required by  
the review committee have been made.

Review Committee

Dr. Roger Wells, Committee Chairperson,  
Applied Management and Decision Sciences Faculty

Dr. David Bouvin, Committee Member,  
Applied Management and Decision Sciences Faculty

Dr. Daphne Halkias, University Reviewer  
Applied Management and Decision Sciences Faculty

Chief Academic Officer and Provost  
Sue Subocz, Ph.D.

Walden University  
2020

Abstract

Ethics of Collection and Use of Consumer Information on the Internet

by

Thanh M. Pham

MS, Walden University, 2006

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Applied Management and Decision Science

Walden University

November 2020

## Abstract

Consumer online activities can generate massive volumes of data that private companies may collect and use for business purposes. Consumer personal data need to be protected from unauthorized access and misuse. The specific problem is that consumers have little control regarding their data being collected and used by private companies. The purpose of this qualitative archival research was to explore business practices involving collection and use of consumer data without an individual's consent. This study used the big data ethical conceptual framework to focus on various privacy issues, including those related to ownership, transparency, ethics, and consumer privacy laws. Archival documents were collected from United States government agencies and private companies. Collected archives included transcripts of U.S. Congressional hearings, witness statements, federal agency reports, privacy policy statements, data catalogs, and panel discussions on data privacy. Thematic data analysis was selected to code and identify common themes from collected documents. Findings revealed business practices of collecting and using consumer data were not transparent, third party companies obtained personal data without individuals' knowledge, and consumers had limited control over their personal data being collected by private companies. The implications for positive social change may help consumers better understand the benefits and risks involving sharing personal information on the Internet. Company leaders may recognize concerns with respect to consumers' ability to control their data and give them choices to manage their personal information.

Ethics of Collection and Use of Consumer Information on the Internet

by

Thanh M. Pham

MS, Walden University, 2006

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Applied Management and Decision Science

Walden University

November 2020

## Dedication

I would like to dedicate this study to God for His blessings and goodness in my life. Thank you for giving me strength, wisdom, and patience to learn and become a good person. I would also like to dedicate this work to my Dad. I know you are always around to protect and guide me through life.

## Acknowledgments

I would like to thank my supervisory committees, Dr. Wells, as the Committee Chair; Dr. Bouvin, as the Committee Member; and Dr. Halkias, as the University Research Reviewer. Thank you for everything during the time I have spent and learned from you. You always supported, encouraged, and guided me through this long study journey. I would also like to thank my Mom, parents-in-law, brothers, and sisters, who are always ready to help me. I would also not forget to thank my four little angles. You are beautiful, and I am so proud of you. Final, I would like to thank my wife, who has always supported and helped me take care of the kids and family while I was busy working and studying. I am fortunate and blessed to have all of you in my life.

## Table of Contents

List of Tables .....	vi
List of Figures .....	vii
Chapter 1: Introduction to the Study.....	1
Background of the Study .....	2
Problem Statement .....	6
Purpose of the Study .....	8
Research Question .....	8
Conceptual Framework.....	9
Nature of the Study .....	11
Definitions.....	13
Assumptions.....	15
Scope and Delimitations .....	16
Limitations .....	16
Significance of the Study .....	18
Significance to Practice.....	19
Significance to Theory .....	19
Significance to Social Change .....	19
Summary and Transition.....	20
Chapter 2: Literature Review .....	22
Literature Search Strategy.....	23
Conceptual Framework.....	24



Identity .....	26
Ownership .....	28
Reputation .....	28
Privacy in the Digital Age.....	29
Big Data .....	32
Big Data Ethics .....	36
Information in the Digital Ecosystem .....	37
Data-Driven Business Model .....	44
Data Sources .....	46
Data Providers.....	50
Data Buyers.....	51
Consumers in the Data-Driven Ecosystem .....	52
Benefits of Data .....	52
Risks of Collecting and using Consumer Data .....	53
Privacy Risks in Digital Marketing.....	57
Sharing Personal Information on the Internet .....	58
Consumers Managing and Controlling Their Personal Data .....	59
Current Data Protection Laws.....	60
Transparency of Processing Personal Information .....	65
Purpose of Using Consumer Data.....	66
Big Data Analytic Quality .....	67
Consumer Consent .....	68

Current Approaches to Personal Data Protection .....	70
Gaps and Study Findings Related to the Literature .....	72
Summary and Conclusions .....	85
Chapter 3: Research Method.....	86
Research Design and Rationale .....	86
Role of the Researcher .....	90
Methodology .....	92
Document Selection Logic.....	92
Instrumentation .....	95
Procedures for Data Collection.....	97
Data Analysis Plan.....	101
Issues of Trustworthiness.....	106
Credibility .....	106
Transferability.....	108
Dependability .....	108
Confirmability.....	109
Ethical Procedures .....	110
Summary .....	111
Chapter 4: Results.....	113
Research Setting.....	114
Demographics .....	114
Data Collection .....	115

Data Analysis .....	118
Evidence of Trustworthiness.....	125
Credibility .....	125
Transferability.....	126
Dependability .....	127
Confirmability.....	127
Study Results .....	128
Finding 1: Potential Benefits .....	128
Finding 2: Data from Multiple Sources .....	130
Finding 3: Identity Information .....	135
Finding 4: Transparency in Data-driven Businesses .....	136
Finding 5: Data Products and Services .....	138
Finding 6: Affiliates and Partner Networks .....	141
Finding 7: Purposes of Using Consumer Data.....	142
Finding 8: Complexity of the Data-Driven Industry.....	144
Finding 9: Ownership and Consent.....	146
Finding 10: Access and Control Over Personal Information.....	150
Finding 11: Risks and Ethics of Using Personal Data .....	153
Finding 12: Data Protection Regulations .....	161
Summary .....	163
Chapter 5: Discussion, Conclusions, and Recommendations .....	164
Interpretation of Findings .....	165

Limitations of the Study.....	179
Recommendations.....	181
Implications.....	191
Conclusions.....	195
References.....	198
Appendix A: Archived Document Selection Protocol.....	223
Appendix B: Permission Letter for Use of ITU Figures.....	228
Appendix C: Permission Letter for Use of IDC Figures.....	229

## List of Tables

Table 1. U.S. Federal Data Protection Laws.....	61
Table 2. Archival Library .....	94
Table 3. Data Provider Company.....	95
Table 4. Archival Material Metadata .....	99
Table 5. Reliability Evaluated Criteria .....	100
Table 6. Number of Documents Collected from Archival Libraries .....	116
Table 7. Number of Documents Collected from Each Data Provider Company .....	118
Table 8. Examples of Personal Data Elements .....	132
Table 9. Examples of Industry Data .....	141
Table 10. Purposes of Using Consumer Data .....	143
Table A1. Archival Library.....	223
Table A2. Data Provider Company.....	224

## List of Figures

Figure 1. Graphical model of big data ethical framework .....	10
Figure 2. Annual size of global data .....	45
Figure 3. Future data storage location.....	47
Figure 4. Number of Internet users .....	48
Figure 5. Number of telephone subscriptions .....	49
Figure 6. Percentage of households accessing the Internet at home.....	49

## Chapter 1: Introduction to the Study

New technologies such as the Internet, social media networks, smartphones, Internet-of-things (IoT) devices, and big data have been changing rapidly. Many companies have been transforming their traditional brick-and-mortar business models into online business models (Kramer & Wohlfarth, 2018). Companies increasingly use consumer data to improve customer experiences, discover new opportunities, and expand to new market shares (Harting, Reichstein, & Schad, 2018). Public and private organizations are collecting, using, and sharing consumer information so that they are able to make their businesses more productive (Zaki, 2019). Innovative technologies, such as social media networks, Internet websites, and mobile devices provide new data sources for private companies (Hartmann, Zaki, Feldmann, & Neely, 2016; Kramer & Wohlfarth, 2018; Zaki, 2019). Internet-connected devices and applications generate huge amounts of personal data in detail, including information such as name, age, gender, location, and financial transactions (Matthias, Fouweather, Gregory, & Vernon, 2017).

According to Bradlow, Gangwar, Kopalle, and Voleti (2017), advanced analytic technologies, such as big data, artificial intelligence, and machine learning algorithms provide businesses with the capability to analyze and generate new information, including information about consumer behaviors and characteristics. The capacity of computer systems has also been improving, and the cost of data storage has been decreasing. This allows companies to collect and use large amounts of data more efficiently and at a lower cost.

Altman, Wood, O'Brien, and Gasser (2018) noted that the collection and use of consumer data is the subject of legal debate in terms of personal privacy rights and ethics. In collecting large volumes of consumers' personal information, businesses may cause harm to individuals by exposing sensitive data to unauthorized third party organizations. The risks of using personal data without consumer knowledge include misuse of data by businesses, loss of consumer trust in businesses, and price discrimination in the marketplace. Private companies need to develop stronger privacy laws and policies to protect personal data from unauthorized access and misuse.

This study focuses on exploring business practices of collection and use of consumer information without an individual's consent. This chapter includes an introduction to the study, which provides general information. I introduce background information, problem statement, purpose of the study, and the research question. Next, I present the conceptual framework, which is the research foundation for this study. I discuss the rationale for research and selected research methods. The next sections include key concepts, assumptions, and limitations of the study. Finally, I show how this study contributes to filling a gap in the literature as well as addressing positive social change.

### **Background of the Study**

Consumers rely on Internet-connected devices and applications to conduct daily business as well as communicate with friends and loved ones. They use Internet websites to search for information or purchase goods. Innovative technologies such as social networks or smartphones allow online users to share personal information with friends



and relatives at any time. The convenience of these technologies offers consumers new ways to search and explore information, encouraging them to do more business online. Daily activities that consumers perform on the Internet generate a large amount of data revealing personal information, such as information about identity, location, behavior, and characteristics (Baesens, Bapna, Marsden, Vanthienen, & Zhao, 2016). At the same time, emerging technologies such as big data, artificial intelligence, and advanced predictive data analytics provide companies the capability to collect consumer data and analyze it to extract hidden information (Baesens et al., 2016; George, Haas, & Pentland, 2014).

According to Altman et al. (2018), both private and public organizations have been collecting consumer data for long periods of time. A report about benefits and opportunities of big data by the White House (2014) said that analyzing personal data allows organizations to operate business processes efficiently, make the right business decisions, and understand customer preferences as well as behaviors. Consumer data are not only valuable for improving products and services, but also for creating potential benefits for third party data providers. Data providers can extract new insights from customer data and sell them for profit. Many companies build business models that are based on consumer data. Technology and e-commerce companies may consider consumer data to be crucial business assets when competing in the digital economy.

Consumers may not understand fully how third-party companies collect, manage, and share their personal information in the digital ecosystem (Altman et al., 2018). Private companies collect, store, and analyze consumer data for many business purposes.

They collect consumer data from third party data providers and combine data from multiple sources to build comprehensive datasets. Companies also share customer data with partners and third parties without the knowledge and consent of customers (Altman et al., 2018).

There was a lack of studies regarding collecting and using consumer data without individual consent within private companies. Consumers need to know how businesses collect and use their personal data so that they can make appropriate decisions about sharing their data (Prince, 2018). Additionally, exploring issues related to private companies using consumer information may not only protect consumer privacy but also build trust in the Internet environment and thereby promote the digital economy (Department of Commerce Internet Policy Task Force, 2010).

After online users post data, it may be impossible for them to remove all copies of their information completely from the Internet (Kissell, 2019). Consumers can delete data belonging to their online accounts, but that data may be still stored in database systems. In some cases, online users may need to pay to delete their personal data from Internet websites (Petrow, 2018).

Many business sectors, such as online retail, education, transportation, and healthcare may find potential value in consumer data. Online users may also get the advantage of customized products and services that companies recommend based on new insights from customer information. At the same time, issues associated with the use of personal data need to be discussed, especially when consumers lack understanding about how their personal data is being collected and used by private companies (Marreiros,

Tonin, Vlassopoulos, & Schraefel, 2017; Zhu & Chang, 2016). Despite risks involved with consumer privacy violations, there is a lack of studies that address issues associated with these business practices. This research may help consumers better understand how their personal data are collected and used by companies so that they can make appropriate decisions when interacting with Internet-connected applications. Knowledge gained from these research findings may mitigate risks of unauthorized access and misuse of consumer data.

Innovative technologies such as the Internet, smartphones, IoT devices, social media networks, and big data have been changing many aspects of daily life. These technologies transform the ways individuals communicate, entertain, pay bills, or buy goods. IoT technology provides the capability to connect electronic devices together and allows them to communicate and transfer data over the Internet (Atlam & Wills, 2020). At the same time, IoT systems might pose serious cybersecurity threats related to unauthorized access due to the dynamic connectivity of the entire system (Atlam & Wills, 2020). With the increasing use of Internet-connected devices in society, private companies have been able to collect massive amounts of information about online users (Mulligan, Linebaugh, & Freeman, 2019a). In many cases, companies aggregate, analyze, and sell consumer data to other companies (Crain, 2018). Business practices of collecting and using consumer data to improve product quality and enhance customer services may initially be harmless to individuals. However, when companies collect large volumes of consumer information and segregate them into profiles to target individuals for specific

purposes related to price discrimination or political influence, these business practices may cause harm to individuals (Savage, 2019; Schudy & Utikal, 2017; Spence, 2020).

According to Borgesius (2016), Facebook collected the personal data of more than one billion online users. Acxiom is a data-broker company that has headquarters in Little Rock, Arkansas and operates over 23,000 computer servers to manage and analyze over 50 trillion data transactions a year (Roderick, 2014). This company collected more than 190 million consumer records as well as information from 144 million households in the United States (Roderick, 2014). These collected data sets hold thousands of data points on customers, such as contact information, financial data, medical records, age, gender, education, marital status, and other personal information (Anthes, 2015; Roderick, 2014).

### **Problem Statement**

In the data-driven ecosystem, data provider companies such as Oracle, Acxiom, CoreLogic, and Datalogix have a critical role in providing data services for other companies (Mendelson & Mendelson, 2017). They collect massive amounts of personal information from various data sources and sell it to other companies to use for multiple business purposes, such as marketing, financial assessment, or fraud detection (Crain, 2018). Business and marketing firms can buy these data services to understand and target potential customers (Oracle, 2018). These business practices of collecting and then selling consumer data to other companies may occur without the knowledge and permission of customers (Yeh, 2018). Big data technology has been providing companies with various potential opportunities as well as transforming business models in the data-

driven economy (Zaki, 2019). The social problem is that private companies use big data technology to collect and use consumer data without their consent (Estrada-Jimenez, Parra-Arnau, Rodriguez-Hoyos, & Forne, 2017; Mamonov & Benbunan-Fich, 2018; Marreiros et al., 2017; Yeh, 2018).

The specific problem is that consumers have little control regarding their data being collected and used by private companies, and such unauthorized use of personal data may result in harm to consumers (Crain, 2018; Prince, 2018). For instance, consumers were not aware that data-provider companies such as Experian, Acxiom, Equifax, and Facebook collected their personal information (Anthes, 2015; Yeh, 2018). In many cases, collected data sets include sensitive information, such as health records or financial data, which should not be available to the public. Furthermore, data-provider companies have limited customers from viewing their own information (Crain, 2018). Data provider companies have also not disclosed the clients who buy the consumer data from them (Anthes, 2015; Crain, 2018; U.S. Senate, Committee on Commerce, Science, and Transportation, 2013). Another problem associated with the use of sensitive consumer data is that data breach incidents that might cause serious harm to individual privacy have repeatedly been happening at major corporations. Smith (2018) noted that a total of 159,700 data breach incidents happened across major corporations in 2017. For instance, the data breach at Equifax exposed 145 million personal financial records to the public (Smith, 2018). Thus, consumers need to know and understand precisely how their personal data have been collected and used by data broker companies such as Experian, Acxiom, Equifax, or Facebook.

### **Purpose of the Study**

The purpose of this qualitative archival research is to explore private company practices of collection and use of consumer data without an individual's consent. While big data technology transforms business models in the e-commerce economy, collecting and using consumer data without an individual's consent may put privacy information at risk (Chua, Herbland, Wong, & Chang, 2017; Prince, 2018).

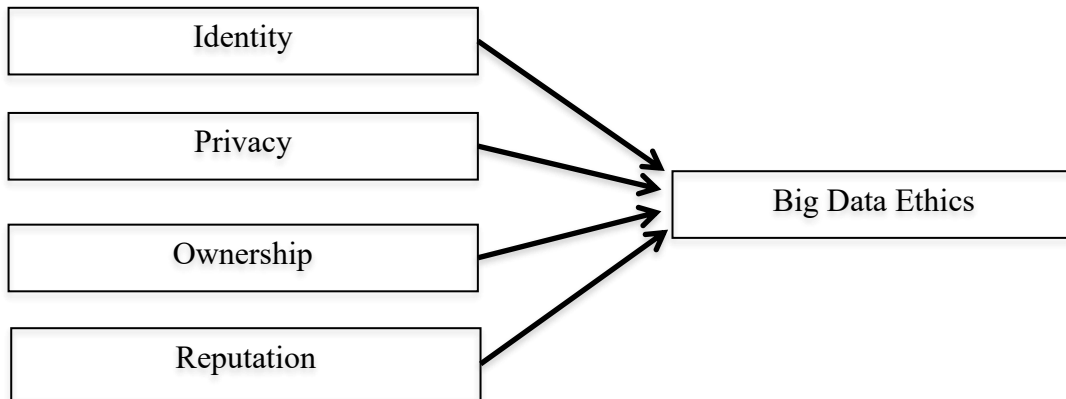
### **Research Question**

The advantage of big data technology is that it aggregates huge amounts of data to discover insights that may be impossible to come to via traditional data management systems (Lee, 2017). Consumer data become valuable for businesses to engage potential customers and improve business services. I explored these business practices of collection and use of consumer data in the digital age. Private companies have initiated new privacy policies to protect consumer personal information. Governments introduced data protection laws to regulate companies to ensure that consumer data processing complies with the law. At the same time, concerns about using private information without consumer knowledge and consent need to be addressed. The central research question of this study is:

How do private company practices of collection and use of consumer data without an individual's consent align with ownership, transparency, ethics, and consumer privacy laws?

## **Conceptual Framework**

To explore the use of consumer data in the digital businesses, this research adapted the Patterson and Davis (2012) conceptual framework of big data ethics and the United States Consumer Privacy Bill of Rights initiated by the White House (2013). Patterson and Davis (2012) developed a theory that provided a conceptual framework to inquire about the ethics of leveraging private information to promote economic growth. This big data ethical framework contains four elements: identity, privacy, ownership, and reputation (Patterson & Davis, 2012). The identity component refers to the relationship between people and information existing in both digital and physical environments. The privacy component relates to the concept of who controls and gives access to the data. The ownership component represents who has the authorization to use, store, and share the data. The reputation component refers to the ability to determine the origin and truth of the information. Figure 1 is a graphical depiction of the Patterson and Davis (2012) conceptual framework of big data ethics.



*Figure 1.* Graphical model of big data ethical framework. Adapted from “Ethics of Big Data,” by D. Patterson, and K. Davis, 2012.

Big data technology refers to the ability to generate, capture, store, and analyze large amounts of data (Halaweh & Massry, 2015). Big data technology has three core characteristics: volume, velocity, and variety (Akhtar, 2018; Halaweh & Massry, 2015; Lee, 2017; Matthias et al., 2017). The volume element of big data technology refers to large amounts of data. The velocity element represents the speed at which data is generated, which increases rapidly over time. The variety element represents a variety of data types. Akhtar (2018) discussed additional dimensions, including veracity, variability, and value, to define in detail the characteristics of big data technology. The veracity element refers to data quality as well as noise in data. The variability element refers to the inconsistency of big data. The value element refers to new information extracting from big data (Akhtar, 2018).

Applying big data technology to consumer data might touch many aspects of people’s lives. Exploring the potential benefits of big data and respecting consumer privacy protection laws can generate benefits for both businesses and customers. Both companies and consumers might be interested in understanding how to use and manage



private information in digital businesses. While big data technology brings new benefits to businesses, such as improving products or customer services, or bringing in new revenue, the study by Marreiros et al. (2017) showed that there are a number of issues related to business ethics and privacy violation when using the private information of consumers without an individual's consent. The Patterson and Davis (2012) conceptual framework provides a foundation for exploring big data ethics with respect to ownership, identity, reputation, and privacy. The White House (2013) also introduced the United States Consumer Privacy Bill of Rights to ensure that businesses comply with current data privacy laws when handling consumer data.

### **Nature of the Study**

The purpose of this qualitative archival research is to explore private company practices of collection and use of consumer data without an individual's consent. To address the research problem, archival research method was considered for this research. Ventresca and Mohr (2017) noted that archival research methods allow investigators to study historical data so that they can explore company practices at different times. Corti (2004) defined archival research as a systematic form of knowledge inquiry that is used to search, analyze, and draw inferences from archival data. Corti (2004) mentioned that investigators can use archival research methods to answer new research questions, evaluate existing conclusions, discover emerging issues, and generalize findings by aggregating archival data from multiple sources. Stan (2010) pointed out that investigators can rely on archival data to study organizational practices. For the purpose of this study, I used archival research methods to explore organizational practices of

using consumer data without an individual's consent with respect to ownership, transparency, and ethics. Archival data can contain important company information which investigators can study to understand the impact of company practices at the present time. Investigators can develop systematic methods to gain knowledge from archival documents. They can find past information and associate it with current events so that they may understand missions, objectives, and operations of organizations.

Ventresca and Mohr (2017) noticed that modern organizations use written documents as one of the primary communication methods to record official business operations. Many organizations in both public and private sectors generate and archive documents in digital formats. Das, Jain, and Mishra (2018) mentioned that archival documents contain rich information such as organizational policies, annual reports, press releases, transcripts of the United States House of Representatives hearings, and subject matter expert interviews. Historical documents may be archived as text, images, websites, or media records. These documents generated in the past allow investigators to access and study organizations (Ventresca & Mohr, 2017). In this study, archival documents helped to explore the insights of companies when they collect and use consumer data without an individual's consent.

Falkenberg (2010) mentioned that ethical problems are often complex, requiring different points of view to assess and interpret issues. It is therefore critical to gather data from different sources to have sufficient information regarding the topic of interest. Regnell, Rainer, Host, and Runeson (2012) noted that collecting data from multiple

sources could limit the effect of drawing conclusions and reduce response biases from a single data source.

I used archival research methods to collect historical data from multiple sources, such as hearing transcripts from the U.S. House of Representatives, archived documents from the U.S. Federal Trade Commission (FTC) testimonies, and transcripts of interviews with business executive officers. These documents are archived in digital formats and accessible through digital libraries, Internet web sites, and government digital archives. Conclusions can also be triangulated from different data sources. When interpretations of data from multiple sources lead to the same conclusion, the conclusion may have enough substantial evidence to allow the development of an initial theory that explains the topic of interest.

### **Definitions**

*Artificial intelligence:* A study field within computer science that applies mathematical and computational algorithms to calculate, learn, and adjust automatically based on data inputs.

*Cloud computing:* Virtual computing resources, such as networking, databases, or software platforms that are provided over the Internet by technology companies.

*Data analytics:* A process of cleansing, transforming, aggregating, and extracting insights from raw data for various business purposes.

*Database:* A collection of data that is stored and managed through relational database management software and run in a computer system.

*Data mining:* A process involving various computing algorithms and mathematic calculations to analyze and reveal hidden patterns existing in data sets.

*Data provider:* Company that collects data from multiple sources, aggregates them to discover hidden values, and resells them to other organizations.

*Ecommerce:* Commerce transactions that involve buying and selling goods over the Internet.

*Information management system:* Computing software that manages, analyzes, and displays business information for decision-making purposes.

*Internet-of-things (IoT):* Electronic devices connect to the Internet to collect, communicate, and exchange digital data with other devices or systems.

*Internet-connected device:* An electronic device that connects with other systems to exchange digital information over the Internet.

*Internet protocol (IP) address:* A unique numerical identifier that is assigned to each device to send and receive data in the computing network.

*Machine learning:* A computer science field involving computing algorithms and mathematic calculations to analyze, learn, and adjust automatically based on data inputs.

*Microsoft SQL server:* Relational database management software developed by Microsoft.

*Online application:* Computer software, mobile application, or web-based application that allows users to create accounts, submit forms, or search information over the Internet.

*Online platform:* Computer software, mobile application, or web-based application that allows businesses and consumers to conduct business transactions over the Internet.

*Online User:* An individual who uses web browsers or mobile applications to access the Internet.

*Oracle database:* Relational database management software developed by Oracle.

*Smartphone:* A cellular or mobile phone installed with various software applications that allow users to make phone calls, send text messages, browse the Internet, take pictures, and watch online streaming videos.

*Social Network:* A web site through which online users create online social profiles and stay connected with friends or family members.

### **Assumptions**

This study is based on three assumptions. The first assumption is that consumers use various privacy protection strategies to protect their personal information in the Internet environment. They try to limit third party businesses collecting personal information from their Internet-connected devices. For instance, social network users restrict unknown friends from accessing their profiles, comments, and photos. They adjust privacy settings provided by online applications to restrict access and to limit data sharing with business providers (Kokolakis, 2017). The second assumption is that consumers often need to agree to privacy statements or terms of use documents before being able to create personal accounts or to use online applications. They disclose their personal information to business providers, expecting companies to use their data in order

to provide better business. The third assumption is that consumers conduct business transactions online through various Internet-connected devices, websites, or online applications provided by business providers. These online platforms provide consumers the capability to search for information, create personal accounts, write comments, and share personal information.

### **Scope and Delimitations**

This study focuses on private companies currently operating in the United States. These U.S. companies collect and aggregate consumer data so that they disclose private information without consumer knowledge. They also collect personal data from different sources and resell it to other companies. These companies might gather data from consumers directly. They might also purchase processed data from third party data providers. Consumers may share their personal data through various Internet-connected devices, applications, and websites. In order to gain a better understanding of these business practices, existing digital materials from both public and private organizations were collected and used as primary data in this study.

### **Limitations**

I used qualitative archival research methods to conduct this study. Digital archives from different sources were the primary data sources in this study. The limitation of using digital archives is that researchers might not be able to measure the volume, size, and content of archival materials (Gaillet, 2012). Researchers cannot control archival repositories and determine what documents they find from archival collections. Since archival collections are originally documented and used for other purposes, researchers

cannot control how information was collected and documented. In addition, as archival documents may not directly respond to the research questions of the current study, researchers need to compile and re-code them to find answers.

In some cases, digital repositories might not keep archives up to date due to reasons such as lack of budget, personnel, or technology infrastructure. Collecting data from web sites or digital repositories might constrain researchers' ability to access physical materials or find aid from librarians. Gaillet (2012) mentioned the limitation of keeping long-term access to digital archives. Organizations might update their websites due to changing technologies, modifying links to documents, or making archival collections inaccessible from the outside. Another limitation of digital archives regarding proprietary documents is that private companies might restrict outsiders from accessing business operational records due to trademarks, contracts, or business competition.

This research focuses on exploring general business practices of collection and use of consumer data. It did not provide a specific privacy solution for a particular Internet application or business. Individuals may have different perspectives about their privacy and personal information. Consumers may lose the convenience of innovative technologies that make information easy to search, share, and store when they refrain from connecting to the Internet. At the same time, personal information may be misused by unknown entities when consumers share their personal information without knowledge of how companies use it.

### **Significance of the Study**

Data provider companies such as Acxiom, Facebook, Datalogix, CoreLogic, and many other firms in the consumer data-providing industry have been collecting millions of individual consumer records from a wide range of data sources (Anthes, 2015; Ramirez, Brill, Ohlhausen, & McSweeney, 2014a). Many data provider companies have not been collecting data directly from the customers, but instead have been gathering customer data from multiple sources, such as the federal government, local governments, public data available on the Internet, and commercial data (Ramirez et al., 2014a). Data firms have aggregated information from these data sources to create a complete and comprehensive profile of millions of individuals (U.S. Senate, Committee on Commerce, Science, and Transportation, 2013). I explore these business practices with respect to ownership, transparency, ethics, and consumer privacy laws. Patterson and Davis (2012) pointed out that businesses need to align between benefits and risks when using big data technology to collect and use consumer data.

Consumer data is an essential asset in ecommerce. Lee (2017) showed that data-driven businesses could take advantage of information to improve products and stay ahead of competitors. Using big data technology to manage and process large volumes of data also benefits consumers. It allows consumers to easily search, retrieve, and share information over various Internet-connected devices and applications. While technologies and consumer data lead to potential benefits for both businesses and consumers, companies need to consider how to protect private information from misuse, unauthorized access, and data breaches. Innovative technologies drive companies to



progressively use consumer data in their business operations (Jobs, Aukers, & Gilfoil, 2015). Without appropriate regulation and management, this may lead to privacy violations and cause harm to society. Feri, Giannetti, and Jentzsch (2016) pointed out that consumers are not surprised by acknowledging data breach events in the news due to data spillage incidents happen frequently.

### **Significance to Practice**

Understanding issues regarding using consumer data may help companies develop an appropriate privacy protection framework and best practices for protecting consumer data. Companies may also use results from this study to maximize benefits of using consumer data while mitigating the ethical risks of exposing consumer privacy.

### **Significance to Theory**

The results of this study may contribute to existing information management research as I explore organizational practices of collection and use of consumer data without an individual's consent. Collecting and using consumer data without an individual's consent may compromise customer privacy and result in harm to the consumer society (Crain, 2018).

### **Significance to Social Change**

The results of this study may contribute to positive social change. Specifically, consumers can secure their personal information better with more knowledge of business collection processes taking place through social networks, online web sites, smart devices, and other Internet-connected platforms. While consumers may be familiar with using technologies such as smartphones, mobile applications, and social networks, they

may not fully understand the complexity of data processing technologies that businesses use to collect, analyze, and extract information from online activities. Companies may sell consumer data to other organizations for profit. The results of this research may help consumers understand the benefits and risks of sharing their personal data with private companies so that they can make appropriate decisions to protect their privacy. Personal data breaches frequently happen in many industries, including technology, healthcare, banking, and education (Verizon, 2018). Consumers may need to know and understand personal data being collected and used by third party data providers. Fuster and Gutwirth (2013) pointed out that consumers have the right to access their personal data as well as protect their privacy.

### **Summary and Transition**

This chapter includes an overview of the study, showing issues in business practices of collection and use of consumer information without an individual's consent. I discuss the research problem, the purpose of the study, and the research question along with the qualitative archival research design, which is selected to explore and gain a better understanding of this topic of interest. This chapter also includes the big data ethical conceptual framework and its components which provide foundational concept for the research problem. This chapter also includes the rationale for selecting archival research method that is used to conduct this study. I present the study's contribution to filling gaps in the literature and positive social change. Chapter 2 includes an overview of existing studies regarding using consumer information in private businesses as well as current data privacy laws. Chapter 3 includes a discussion of the archival research

method, including archival data sources, research instrumentation, data collection procedures, and the data analysis process. Chapter 3 also includes the role of the researcher, issues of trustworthiness, and ethical procedures to ensure that the credibility of this study is preserved.

## Chapter 2: Literature Review

In the Internet ecosystem, consumer data becomes valuable for businesses to understand consumer preferences and market products to potential customers. Consumer online activities, including using mobile phone applications, browsing the Internet, shopping online, searching for services, and posting messages on social networks may be collected and sold by data provider companies (Ramirez, Wright, Brill, Ohlhausen, & McSweeney, 2014b). Companies collect consumer data from various sources and use it for many business purposes. Many businesses might not get data directly from customers; instead, they buy processed data from data provider companies. The problem is that consumers may not be aware of their personal information being collected and used by private companies. Lack of transparency regarding these business practices may result in harm to consumers. The purpose of this qualitative archival research is to explore private company practices of collection and use of consumer data without an individual's consent.

This chapter includes an overview of the big data ethical conceptual framework and its components, including identity, ownership, reputation, privacy, and ethics of big data. This big data ethical conceptual framework provides a foundation to conduct the literature review involving using consumer data in the data-driven business ecosystem. Characteristics of data sources, data providers, benefits of data, and privacy risks are discussed to provide an in-depth understanding of the data-driven business model in contemporary studies. This chapter also includes current data protection laws that regulate consumer data processing in the Internet economy.

### **Literature Search Strategy**

I used the Walden University Library to collect and review existing literature regarding information systems and technology research. I used the following databases in this study: EBSCOhost, ProQuest, Emerald Insight, and ScienceDirect. These databases have many options that allow filtering search results to most relevant articles. I used several options, including keywords, full text, publication date, peer review, source type, and document type, to reduce numbers of articles returning from each search keyword or terms.

The keywords and terms used to search relevant articles were: *consumer privacy, consumer data, personal data, privacy information, data privacy, online privacy, data-driven business, data provider, consumer data in online marketing, big data, ethics of big data, data ethics, privacy law, privacy protection, regulation control, and data breach*. I included relevant articles which were published within the last 5 years. Some peer-reviewed articles published earlier were also reviewed due to essential information to support this study. The objective of reviewing prior studies is to gain a better understanding of the research problem, recognize strengths and weaknesses in the literature, define theories and findings, and identify knowledge gaps in existing studies.

Beside scholar articles collecting from the Walden University Library, I also used Google to review reports from web sites of both public agencies and industries. Reports were collected from the U.S. Congress, FTC, World Economic Forum, International Telecommunications Union, IBM, and Pew Research Center. These organizations study social trends, technology, and policy. They conduct public polling and report data privacy

issues that might have an impact on consumers. I collected articles involving Internet users, IoT devices, privacy data, digital data trends, and social networks.

### **Conceptual Framework**

Patterson and Davis (2012) argued that there are different perspectives regarding businesses utilizing big data technology to collect, store, and analyze large amounts of data from business processes and consumers. George et al. (2014) noted that big data technology provides businesses the capability to understand customer preferences, improve product and service qualities, and enhance marketing campaigns. While businesses are excited about the potential benefits of big data technology, there are also potential risks to using big data technology to aggregate and analyze consumer data without their consent. Patterson and Davis (2012) argued that the potential risks may be associated with ownership, transparency, ethics, and consumer privacy laws. Patterson and Davis (2012) pointed out that businesses need to balance benefits and risks when using big data technology to collect, store, and analyze consumer data. Using consumer data to improve products and services is a good thing for businesses. At the same time, companies also need to consider protecting consumer privacy and preventing the misuse of sensitive personal information.

While technologies such as big data, data analytics, and machine learning have been changing rapidly in terms of functionality and capacity, data protection legislations have not been revised quickly enough to regulate the changes. The federal government and lawmakers have debated and proposed different consumer data protection policies to regulate how private companies collect and use consumer data. In 2012, the U.S.

government introduced the Consumer Privacy Bill of Rights Act to regulate businesses using consumer data as well as protect consumer privacy (White House, 2013). In 2018, California representative Ro Khanna proposed the Internet Bill of Rights to protect consumers from data surveillance and give them more control over their personal data (Swisher, 2018).

Companies can capture online activities when users browse information on web sites or search for information on their smartphone. Businesses can track what web sites consumers open, what products they look for, and how long they spend on a specific website. Companies can also use big data technology to store consumer data permanently in sophisticated information management systems. Furthermore, big data technology allows businesses to aggregate consumer data from multiple sources to create a comprehensive individual profile. For instance, businesses can use location information embedded in a digital picture file to find out where consumers have traveled in the past.

Information that consumers want to separate and withhold from the public may be aggregated and analyzed by businesses to reveal consumer behaviors or lifestyles. Big data technology provides companies the capability to integrate multiple layers of data so that they can unveil the private information of customers. In many cases, businesses share customer information with partners to make business processes more efficient (Bergstrom, 2015; Estrada-Jimenez et al., 2017). Each business sector has unique data sets about its customers. For example, banking institutions have information about financial activities and transactions of banking customers. Health insurance companies have customer medical records. Online retailers may have customer spending data. When

these business sectors exchange customer data to improve their products and services, analyzing and extracting insights from these data sets might have significant impacts on customers (Mantelero, 2016). Companies may also use insights from customer information to make business decisions so that they can offer products and services that meet preferences of target customers. They may develop effective marketing campaigns that target specific groups of customers.

Private companies want to expand business and generate revenues for investors. They want to establish good relationships with customers so that customers feel confident when doing business with them. Big data provides companies the capacity to better understand customers by collecting and analyzing large amounts of customer data. Companies can also exchange customer data with their partners to leverage the full capability of big data technology. The concern is that customers do not know that their personal data have been collected, analyzed, and shared among different companies. Big data ethics and privacy are main concerns in the data-driven economy (Patterson & Davis, 2012). Patterson and Davis (2012) presented four elements of the big data ethical framework: identity, privacy, ownership, and reputation.

### **Identity**

According to the U.S. Department of Labor (2019), personal identifiable information (PII) is any information that directly identifies an individual, or indirectly identifies an individual when combined with other pieces of information. Such information may include (but is not limited to) name, gender, birthdate, address, nationality, or culture. Government organizations often have the authority to issue and



validate physical identity artifacts, such as a birth certificate, driver's license, social security number, or passport, to identify an individual. These are some of the fundamental pieces of information pertaining to identity that people use to define who they are in the physical world. In the digital environment, organizations use a combination of various pieces of digital information, such as usernames, passwords, or personal identification numbers (PIN), to authenticate an individual's identity. Organizations may also combine multiple pieces of digital data, such as email, online accounts, cell phone numbers, IP addresses, social network profiles, and historical transactions, to verify individuals' identity.

With the growth of Internet-connected devices and online activities, an identity in a digital world report by World Economic Forum (2018) showed that consumers increasingly used digital identity to authenticate and accessed online services. Over a period of time, consumers generate a large volume of historical data over digital networks that describe in detail individual activities, behaviors, and profiles. Companies may use this information to make decisions on segmenting customers or offering services to particular customers. Patterson and Davis (2012) said that technologies transform individual identity in the digital environment as well as the way that businesses identify and recognize an individual. Businesses aggregate many personal data from multiple sources, verify individual identity, extract private information, and determine individual value. Patterson and Davis (2012) pointed out that ethical issues exist in this practice when consumers may not know or consent to an entire process of collection, correlation, and analysis of personal data to identify as well as define individual value.

## **Ownership**

Janecek (2018) noted that when individuals or organizations have ownership of information, this means that they have the right to make decisions on their own data, determining who can access and use their data. Today, consumers use various Internet-connected devices and applications to communicate and exchange information. Consumer online activities generate a large amount of digital data that companies collect and use for many purposes. The question is who owns these data. Patterson and Davis (2012) argued that the degree of ownership of information on the Internet environment might be broad and hard to define. The issue of data ownership is how to distinguish between personal data and non-personal data on the Internet. Online activity data, such as browsing history, search keywords, or product previews, may be non-personal information, but when companies combine it with other information, such as gender, age, and marital status, to predict the pregnancy status of an individual, this online activity data may become personal information. Nevertheless, whether organizations or consumers own the information, data owners have the right to protect their data from misuse and unauthorized access.

## **Reputation**

In the context of a data-driven economy, Patterson and Davis (2012) defined reputation as primarily being what consumers say or think about a company. Businesses build their brand reputations by offering quality products and providing good customer service. Internet, social networks, or online streaming media technologies provide companies the capability to expand their customer base in a short period of time. These

technologies also allow customers to post reviews as well as opinions about particular products, services, or businesses. Consumer review is one of the key factors that may influence the business reputation. These online capabilities are critical for businesses to build their reputations and target new potential customers.

In the Internet environment, customers can write messages, opinions, and preferences, as well as exchange their expectations about almost any business, product, and service. Technologies, such as big data, machine learning, or advanced data analytics, provide companies the ability to collect, analyze, and extract new information from these consumer data. This capability allows companies to understand customer preferences, business trends, and competitors. This useful information may help companies determine their reputation and stay competitive on the market.

### **Privacy in the Digital Age**

Trepte et al. (2015) defined online privacy as a process in which individuals manage and authorize access to personal matters when they interact with Internet-connected devices or use Internet platforms. Privacy means individuals can manage and control their own personal information. Baek, Kim, and Bae (2014) noted that online users want to balance privacy and convenience and need to make appropriate decisions on sharing personal data in each occasion. Estrada-Jimenez et al. (2017) pointed out that privacy risk is the amount of personal information that unauthorized parties can collect and use to learn about an individual. Dyke, Midha, and Nemati (2007) mentioned three essential elements in personal privacy: notice, choice, and access. The notice element refers to advance notification from business entities who collect and use consumer data.

Businesses should also need to notify customers that they are going to share or sell consumer data to other companies. The choice element refers to the user's right to decide whether to share personal data with businesses. The access element refers to online users' ability to modify their personal data.

New technologies such as the Internet, smartphones, online payment services, and social networks allow consumers to search, retrieve, create, and exchange information. Many Internet-connected platforms provide communication channels at both individual and group levels. The communication context can include many formats, such as texts, pictures, videos, and website links. Online service providers, technology firms, online retailers, and third-party companies might later use these communication contexts for many business purposes. Besides front-end online platforms that consumers use daily, sophisticated technology behind the scenes, such as big data, data mining, or cloud computing, enable companies to collect, store, and analyze large volumes of digital data at a low cost.

Baek et al. (2014) mentioned that with massive amounts of consumer data generated on the Internet environment, companies could aggregate consumer data from various sources and extract new information. Consumers may not fully understand the complexity of information management processes. As Milne (2015) discussed, privacy issues might exist in these business practices since businesses and consumers may have different expectations about personal data management. Companies recognize that consumer data can bring many potential benefits to businesses in the digital age. They gather consumer data from both internal and external sources to enrich customer profiles.

At the same time, consumers may expect businesses to protect their private information from misuse and unauthorized access. Consumers also want businesses to keep their personal information private and have more control over their own personal data (Milne, 2015). However, consumers may not know that companies exchange data with other firms to use in many business purposes, such as marketing or building customer profiles.

Collecting and using consumer data without individual consent may cause harm to businesses, consumer privacy rights, and society (Milne, 2015). Companies may lose consumer trust when they expose personal data to third parties, misuse consumer information, or use information for a purpose that is not defined in the privacy term statement (Milne, 2015). There are many factors that may influence whether customers do business with a company. An IBM and Harris Poll (2019) study showed that 71% of participants do business with a company if they protect privacy data from cybersecurity attacks, 69% look for product quality, and 65% consider whether or not a company shares their data with third-party organizations. The negative perspective of consumers about companies violating their privacy may affect business goals and customer relationships.

The process of gathering, aggregating, analyzing, using, and selling consumer data needs to be transparent, so that an individual is able to recognize the risks associated with personal data disclosure. This can help consumers make the right decisions on their own personal data. The transparency of using consumer data also helps increase the trust of customers on the companies. Prince (2018) suggested that online users should utilize privacy setting tools to manage sensitive information on the Internet. Online users need to

know which personal data they want to share so that sensitive data is not disclosed and misused by data-provider companies.

### **Big Data**

Organizations have been operating businesses over computing networks for decades. Zhang, Wang, and Pauleen (2017) said that companies digitalized their business data so that they could process, exchange, and update business information effectively and efficiently. In the past, companies could not store and process large volumes of data due to limitations of computing powers, networks, and traditional data management system capacity. In the last decade, many innovative technologies, such as cloud computing, big data, artificial intelligence, or machine learning, emerged and offered powerful IT infrastructures and information systems to manage and process data (Fulgoni, 2013). With the sophisticated capability of big data technology, Nada (2016) noted that companies could collect, aggregate, and store a massive amount of data from multiple sources. New insights may help companies better understand customer preferences and forecast business trends (Nada, 2016). Big data offers companies various potential opportunities to generate new value as well as stay competitive in a digital economy (George et al., 2014). Companies began using big data technology to collect consumer data and extract intelligence from it. Jobs, Gilfoil, and Aukers (2016) showed that big data provides companies the capability to optimize business processes, improve products, build marketing strategies, and enhance customer services based on customer data.

The United Kingdom Information Commissioner's Office (2017) discussed five characteristics of big data, which define the advanced data analytic features of this technology. The five characteristics of big data technology are as follows. The first is that big data uses computing algorithms to process and analyze data so that it can extract relationships among input variables. The second is that many computing algorithms can analyze huge amounts of data and produce results that are difficult to understand; it can be difficult to trace back how these algorithms generate particular outputs. The third is that big data analytics requires processing large amounts of data in order to generate accurate and optimal outputs. The fourth is that the collected data can be analyzed and used for different purposes. For example, location data points can be used to identify many activities of an individual, such as workplace, working hours, outdoor pattern, or travel location. The fifth is that emerging technologies, such as Internet, social networks, or IoT devices, automatically generate a large amount of personal data on the Internet environment. These new data types are essential for big data analytics.

Akhtar (2018) defined big data as large volumes of data that have structured, semi-structured, and unstructured data formats and that contain various data types, such as text, document, image, and media, as well as many other types. These data might contain insights that are valuable for businesses. Companies combine, analyze, and extract new information from big data using various sophisticated data mining techniques and tools. Big data not only refers to data size, but also has other complex elements of data that traditional information management systems might not be able to process and analyze in a short period of time (Akter, Wamba, Gunasekaran, Dubey, & Childe, 2016).

The technology community originally described big data using three major elements: volume, velocity, and variety (Akhtar, 2018; Lee, 2017; Matthias et al., 2017). As technologies and data became more complex and sophisticated, Akhtar (2018) discussed additional dimensions, including veracity, variability, and value, to define in detail the characteristics of big data technology.

**Volume.** This dimension refers to the size of data. Big data management systems store and process large amounts of data which are generated from various business operational systems (Abbasi, Sarker, & Chiang, 2016). Data can contain structured, semi-structured, and unstructured information and have various formats, such as logging files, text, documents, or images. Abbasi et al. (2016) discussed that due to large volumes of data, it might not be efficient and optimal to store and manage big data in relational database management systems. Combining massive volumes of data from multiple sources might extract insights and discover valuable information for business intelligence.

**Velocity.** This refers to the speed of data generated by business processes, sensors, Internet-connected devices, consumer online activities, and so on. Big data systems often need to capture business data in real-time and process it anonymously without human interactions. Mukhiya, Wei, and Lee (2018) discussed various approaches, such as batch, real-time, and streaming, to capture and analyze massive information using big data technology. Velocity also is associated with how fast data is created and processed in data management systems so that it might provide intelligent information for decision-making systems (Mukhiya et al., 2018).



**Variety.** This dimension refers to types of big data. Big data contains various information stored in different formats. It can be structured, semi-structured, and unstructured data. Emerging technologies, such as social networks, online media streaming, blogs, sensor data, or mobile phones, generate massive amounts of data in different formats. Structured data prefers data that fit into relational data management systems, such as Microsoft Excel files, Oracle database, or Microsoft SQL server. Structured data are stored in multiple data tables and are joined to each other using key data values. On the other hand, unstructured data are documents, images, videos, and so on. Due to the complexity of unstructured data, relational database tables might not be optimal for storing and analyzing it. For instance, companies might need big data technology to extract new information from business emails, documents, or customer product previews.

**Veracity.** This dimension refers to data quality as well as noise in data. Akhtar (2018) mentioned that uncertainty of values might exist in big data. One of the challenges in big data is how to clean and normalize large amounts of data before analyzing and extracting new information. Companies often need to capture business data in real-time and process it to update intelligent information for decision-making systems. The noise and inconsistency in data might affect the accuracy of outcomes. Veracity refers to the trust in data content when making decisions based on it (Akhtar, 2018).

**Variability.** This element refers to the inconsistency of big data (Akhtar, 2018). The meaning and understanding of big data might change over time. This might happen when analyzing natural language, such as emails, customer service call transcripts, or

product preview messages. The meanings of a word might be different depending on the context of documents. It may be difficult to capture the meaning of words precisely within unstructured data, such as documents, text, voice, or messages.

**Value.** This dimension refers to new information extracting from big data. Big data provides intelligent information to business decision-making systems. Big data offers businesses various potential benefits, such as improving product quality, forecasting business trends, and enhancing customer experiences. In order to gain the full benefits of big data, companies often need to invest in IT infrastructure and software to manage and process large volumes of data (Akhtar, 2018). New information extracted from big data provides companies with business intelligence, so that management teams can make decisions to achieve business goals. Big data contains huge volumes, various data types, uncertainty of information, and diversity of meaning. At the same time, big data might also provide advanced capabilities to businesses to stay competitive in the data-driven economy. The value of big data is in discovering new business insights from large volumes of data that have various formats and come from multiple sources.

### **Big Data Ethics**

In terms of ethics, big data technology is neutral, and has no value embedded in the platform itself (Patterson & Davis, 2012). On the other hand, companies have values, such as optimal business operations, good products, or profits, associated with their businesses. The ethical issue may arise when companies use big data technology to collect and use customer data without their consent. Patterson and Davis (2012) argued that business ethical subject is different among organizations. Companies are often

concerned about business ethics when they affect business operations, products, or profits. Due to the benefits of using big data technology to collect and use consumer data, misusing consumer data in business operations may affect the privacy of customer lives (Mai, 2016). In fact, private companies have stored various data sets related to customers in their database management systems. The question depends on how companies use big data technology to collect, store, analyze, and share consumer data. When companies use consumer data to improve current business processes or to find new market shares, they need to align business values with ownership, transparency, ethics, and consumer privacy laws. Transparency in how companies collect and use consumer data may establish trust between businesses and customers. It can also give consumers clear understandings of the benefits of using personal data to optimize products and services.

### **Information in the Digital Ecosystem**

In recent years, innovative technologies, such as big data, social networks, smartphones, or wearable devices, have been changing at the fastest pace. Internet users use various social networking platforms, such as Facebook, WhatsApp, Twitter, or Instagram, to communicate with friends and family members. Smith and Anderson (2018) conducted a survey of social media use in 2018. They interviewed 2002 adults who are 18 years of age or older and living in the United States and the District of Columbia. The study showed that YouTube and Facebook are the most popular social platforms on the Internet; 73% of respondents reported using YouTube, and 68% had Facebook accounts. The report estimated three-quarters of respondents (74%) who have Facebook accounts check the website on a daily basis, and half of them (51%) access the

website several times a day. This survey also showed that young adults often use multiple social network platforms. For instance, almost three-quarters of respondents use more than one social network platform, average respondents use three platforms, and respondents between 18 to 29 years of age often have more than four platforms.

Social network websites provide many features to online users, including sending messages, sharing pictures, and viewing friend profiles. In data-driven ecosystems, data of online activities and online personal information are valuable for businesses to understand consumer perspectives and preferences (Gupta & Schneider, 2018). Companies can use new technologies, such as big data, or data mining, to collect historical Internet browsing as well as develop rich consumer data sets. Collected data may contain many personal data points, including name, location, friends, gender, age, education, salary, marriage status, financial conditions, and other pieces of personal information (Altman et al., 2018). With large volumes of consumer data available on the Internet environment, companies can aggregate data and build individual profiles to better understand customer behaviors and predict business trends.

Innovative technologies such as big data, machine learning, or data mining, provide companies the capability to compete and gain market share in the digital economy. At the same time, the use of big data technology to collect and extract personal information without an individual's consent may be associated with privacy violations and business ethics. A data privacy survey of 525 adult consumers in the United States conducted by SAS (2018) showed that almost three-fourth of participants (73%) believed that organizations collect and use their personal data without their consent. Lawmakers

and consumers raise concern over business practices of unauthorized use of consumer data to track consumer activities and extract personal information.

SAS (2018) data privacy survey found that consumers raised concerns over their personal data being collected by companies. Overall, 73% of respondents raised concerns over data privacy, and 67% of them wanted the U.S. government to regulate businesses more in terms of protecting and handling personal data. Among participants, who want the U.S. government to do more on data privacy protection, most of them (83%) do not want organizations to sell or exchange their personal data, and 80% want to know where their personal data are sold to. SAS (2018) data privacy survey also showed that 66% of participants wanted to control over their personal data by updating privacy settings of Internet-connected applications, removing social network accounts, and declining privacy statements. Specifically, 77% of them would like to change privacy settings of Internet-connected applications, 65% of them do not want to accept the privacy policy statement, 56% want to delete mobile applications, and 36% want to remove social networking accounts.

Managing data generated from Internet activities may be complex and challenging (Milne, 2015). Technology companies may use big data technology to gather massive amounts of consumer data from various sources, such as Internet websites, social networks, or mobile devices. The data can be in various formats, such as texts, images, documents, and videos. Traditional information systems might not have capacity to store large amounts of data and process them in a short period of time. Big data technology can provide companies the capability to build sophisticated information systems to process

large data in both structured and unstructured types. The challenge is that companies might not establish adequate policies to regulate the process of collecting and using consumer data (Kshetri, 2014).

Companies may buy and aggregate external information from data providers to create complete customer profiles. Innovative technologies, such as computing machine learning algorithms, or data mining, provide companies capability to collect consumer data from multiple sources and build individual profiles. Many data mining techniques allow them to learn customer shopping behavior, Internet browsing activities, spending style, and other personal information. Businesses can generate new revenues when adopting and integrating advanced data analytic technologies into business operation processes. For instance, businesses analyze product review data from customers to understand their preferences so that companies can improve product quality and customer experiences. Consumer data, that are associated with personal characteristics, may help businesses build complete customer lifestyles as well as understand their preferences.

Besides the benefit of analyzing customer data to improve product quality, enhance customer experience, and gain new market shares, companies can create new revenues from customer data by aggregating and selling it to other companies for other purposes. These business practices may bring up questions about privacy issues as well as business ethics. The privacy issue is that customers might not know their personal information has been sold to other companies. Businesses might have the responsibility to protect customers' data and allow customers to make decisions on their personal data. A study of cybersecurity and privacy conducting by IBM and Harris Poll (2018) collected

data from 10,500 participants globally through an online survey. It showed that 78% of participants agreed that it is important that businesses have ability to protect their personal data.

An online survey of 2546 U.S. adults conducted by HarrisX (2018) reported that 83% of participants think that the government needs tougher regulate privacy data, and 62% reported that the government needs to act now as well as initiate new privacy policies within the next few years. The majority of participants do not believe that technology and social network companies protect their personal data; 84% reported companies need to take responsibility in protecting their personal data. In addition, 67% supported existing privacy policies, such as the Consent Act, or the General Data Protection Regulation. These privacy policies require companies to notify consumers when they collect, share, or sell personal information.

By default, the privacy setting features of smartphone applications or social network websites allow companies to access and collect various pieces of customer personal information, including pictures, videos, contacts, or location. Users need to find privacy settings for each application and reset these settings to prevent Internet-connected applications access to the private information. Trepte et al. (2015) argued that online users often lack the technical skills to configure privacy properties appropriately and protect their information. In some cases, online users must allow software applications to access personal information in order to use these platforms. Online users also want to protect their privacy by refraining from sharing personal information with business firms, but they must make a trade-off between giving up their privacy and having the benefits

from Internet-connected applications. Prince (2018) noted that Internet users tend to allow businesses to access private information in return for using software and services provided by these companies. Privacy settings are important mechanisms for online users to control and manage their personal data. Unfortunately, online users might not have the capacity to understand comprehensively how Internet technology works in order to protect their personal data.

Other risks associated with sharing personal data in the digital environment are that customers are not fully aware of how businesses collect and use data after granting them access to personal information. Companies might ask users' permissions one time when users install the application on their mobile devices, but the applications might collect customers' data regularly. In some cases, technology companies gather data when users do not open or interact with their applications.

According to a study of privacy data in 2014 from the Organisation for Economic Co-operation and Development (2014), there are various threads related to disclosing personal information when online users interact with Internet-connected devices or applications. Companies can view critical information such as financial records when online customers use banking applications to manage their financial transactions. Technology companies can aggregate multiple data points that belong to customers and to extract insights into customer lifestyle and spending behaviors. Companies are also selling customers' data to third-party firms with customer permission.

In digital ecosystems, businesses consider consumer data as valuable new assets that can be used for multiple purposes such as marketing, improving customer



experiences, or generating new revenues (Prince, 2018). Businesses have been adopting new innovative technologies such as big data to collect, store, and exchange customers simultaneously in the digital networking environment. They aggregate and analyze customer data to produce a comprehensive data set, which describes the characteristics and behaviors of an individual or a group of customers. By understanding customer behaviors from online activity data, businesses can customize their products and services to target a particular group of consumers. The practices of collecting, analyzing, and selling customer data without full restrictions lead to serious issues in privacy violations in which private customer data might be disclosed to business entities without awareness on the part of the customers (Marreiros et al., 2017; Zhu & Chang, 2016).

Prince (2018) argued that the information privacy of consumers has been violated due to the lack of personal data management policies at social networking sites. Existing research mentioned ethical concerns associated with collecting and using large amounts of consumer data without individual consent (Bender, Cyr, Arbuckle, & Ferris, 2017; Nunan & Domenico, 2013). Mulligan et al. (2019a) mentioned that unauthorized access, customer data breaches, and the practice of trading customer data posed serious issues in the digital economy. These data management issues might discourage consumers from sharing personal information online due to the lack of data management policies and practices. Regulations on using personal data are developing to catch up with changes in new technologies.

At the same time, there is a lack of understanding to address the nature of the issue of these business practices (Kshetri, 2014). Private companies take advantage of the

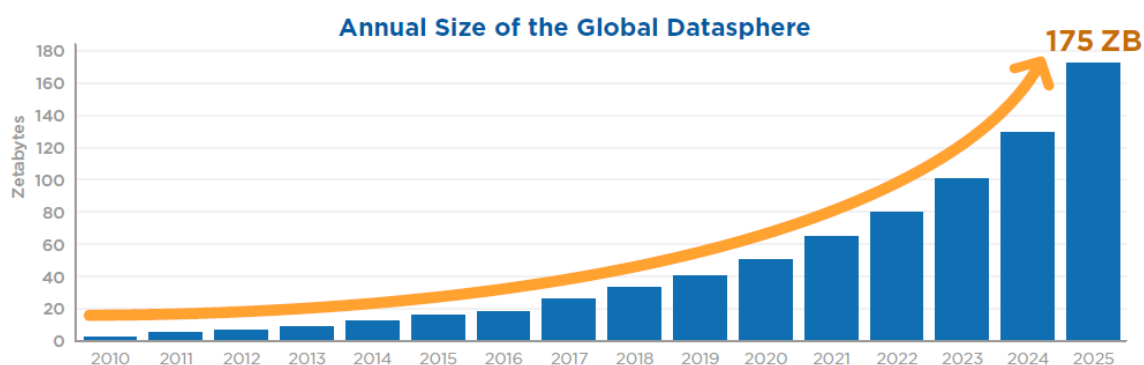
lack of data protection in the digital environment to exploit consumer data. Kshetri (2014) mentioned that companies did not adequately address the privacy issues involved in using consumer data. Customers are concerned about their data privacy when companies collect, use, and share their personal information. In this study, I explored business practices of collection and use of consumer data without individual consent. I presented various elements of these practices within the data-driven business ecosystem, including ownership, transparency, ethics, and privacy laws.

### **Data-Driven Business Model**

In the Internet environment, consumer data are generated in many ways from business reviews, product feedback, personal blogs, messages on social networks, and many other online platforms (Hartmann et al., 2016; King & Forder, 2016). Personal business transactions, such as online payment, shopping, registration, financial transactions, or health records, can create a huge amount of digital data associated with an individual (Reinsel, Shegawi, & Gantz, 2018b; Smith & Anderson, 2016). Hartmann et al. (2016) noted that any online activity could provide data to companies that offer products or services. Wamba et al. (2017) mentioned that companies collected and used these data for many business purposes, from building customer profiles, improving product quality, and marketing to customer engagement and demand prediction.

To stay competitive in the data-driven ecosystem, companies increasingly digitalized business processes and transformed their core businesses based on available data to compete with rivals (Gantz, Reinsel, & Rydning, 2019; Reinsel, Gantz, & Rydning, 2018a). Companies increasingly collect massive volumes of individual data

generated by customers' online activities. Businesses want to use consumer data so that they can personalize products to meet individual preferences. Consumers rely heavily on Internet-connected devices and applications to communicate with family, search for information, and keep track of daily lives. New technologies, such as big data, data mining, or cloud computing, allow companies to capture consumer data in real-time, and insert it into sophisticated computer algorithms to generate predictive model of what business trends and products an individual is looking for. According to a study about digital transformation conducted by Reinsel et al. (2018a), the study showed an estimated 33 Zettabytes data generated by consumer and business activities in 2018. It predicted that volumes of data continue expanding to 175 Zettabytes in 2025. The following figure 2 shows the annual size of the global data.



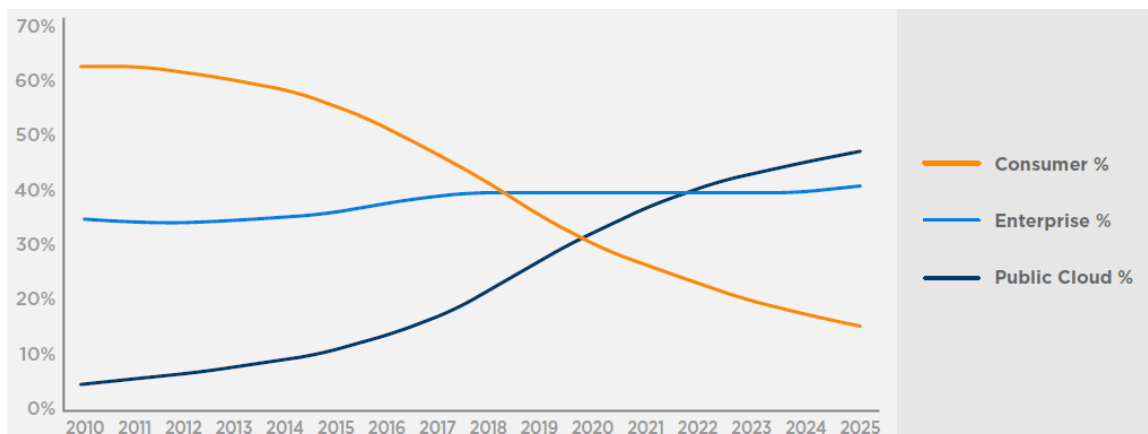
*Figure 2.* Annual size of global data. Adapted from “Data Age 2025: The Digitization of the World from Edge to Core,” by D. Reinsel, J. Gantz, and J. Rydning, 2018, IDC White Paper, sponsored by Seagate. Copyright 2018 by the IDC.

In the data-driven business model, companies use data from various sources to improve business operations, product quality, and customer services (Jobs et al., 2015; Perko & Ototsky, 2016). Companies often use computers to process business data and

exchange information with partners. Data-driven companies use emerging technologies, such as big data or advanced data analytics, to capture and analyze large volumes of data that traditional database management systems might not be able to process. A study from the Federal Trade Commission (2011) regarding consumer privacy protection presented several entities involved in data-driven business models. The data-driven business model includes (a) data sources, (b) data providers, (c) data buyers, and (d) consumers.

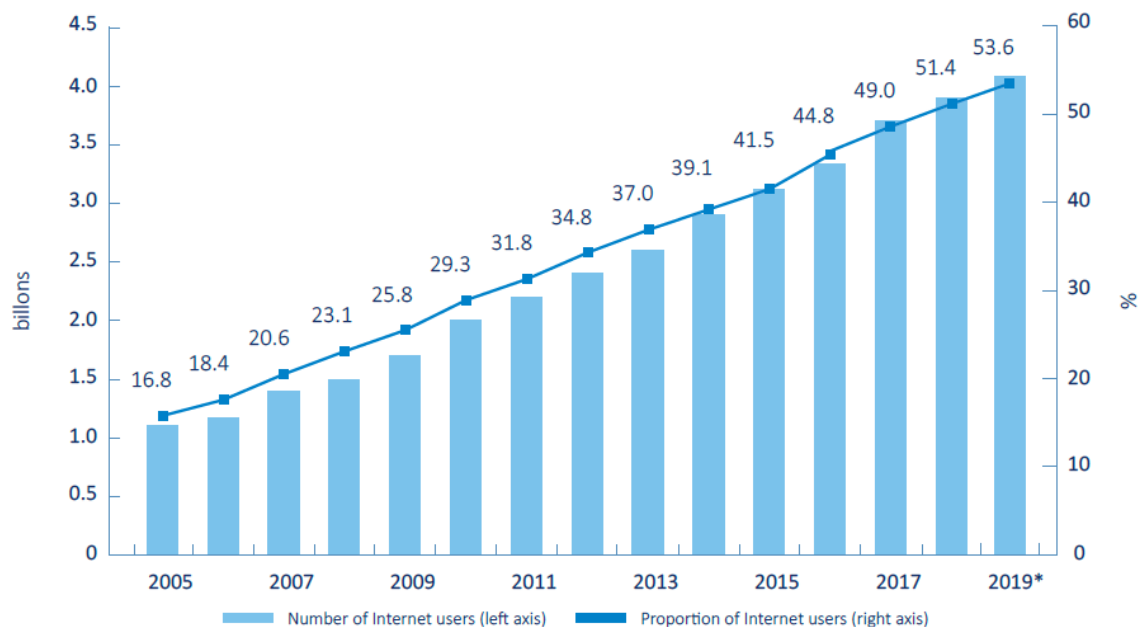
### **Data Sources**

Today, consumers use various Internet-connected devices and applications in their daily activities. Reinsel et al. (2018a) study predicted that more than 150 billion devices connect via the Internet by 2025. Smartphones, tablets, smart speakers, and many other devices are able to collect user data and store it at company or cloud computing provider data centers. This allows consumers to access and retrieve personal data anywhere through their devices. At the same time, companies can also access and analyze these data to better understand consumer behaviors, preferences, as well as to forecast business trends. Reinsel et al. (2018a) showed that more consumer data are stored at cloud-based data centers in the future. The following figure 3 shows future data store locations. With more consumer data managed by enterprises, businesses can aggregate, analyze, and extract new information that may be impossible to reveal in separated databases (Reinsel et al., 2018a).



*Figure 3.* Future data storage location. Adapted from “Data Age 2025: The Digitization of the World from Edge to Core,” by D. Reinsel, J. Gantz, and J. Rydning, 2018, IDC White Paper, sponsored by Seagate. Copyright 2018 by the IDC.

According to a digital development report conducted by the International Telecommunications Union (2019), numbers of Internet users around the world continued to increase promptly in the last decade. The report showed that an average of 10% increase per year since 2005. Specifically, 16.8% of the world population was using the Internet in 2005, and an estimated 53.6% use the Internet in 2019. The following figure 4 shows an estimated number of Internet users worldwide. This means an estimated 4.1 billion people accessed the Internet in 2019. Developed countries, Europe, and America are three regions that have the most online users.



*Figure 4.* Number of Internet users. Adapted from “Measuring Digital Development: Facts and Figures 2019,” by International Telecommunications Union, 2019. Copyright 2019 by the International Telecommunications Union.

In terms of mobile phone subscriptions, the International Telecommunications Union report (2019) showed that numbers of mobile-broadband and mobile-cellular subscriptions keep growing regularly. For instance, 83 per 100 people worldwide have active mobile phone subscriptions. The following figure 5 shows an estimated number of telephone subscriptions worldwide. The report also showed that 71.8% of household in the United States and 57% worldwide have access to the Internet at home. The figure 6 shows an estimated percentage of households accessing to the Internet at home.

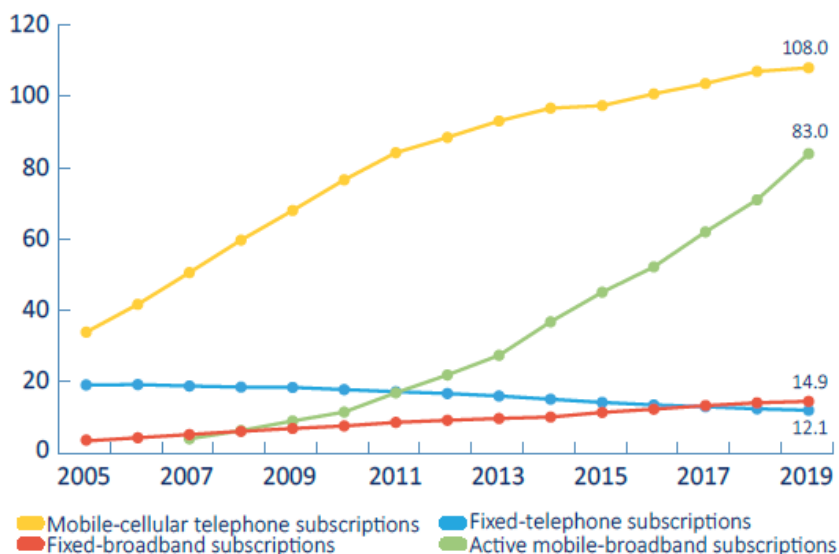


Figure 5. Number of telephone subscriptions. Adapted from “Measuring Digital Development: Facts and Figures 2019,” by International Telecommunications Union, 2019. Copyright 2019 by the International Telecommunications Union.

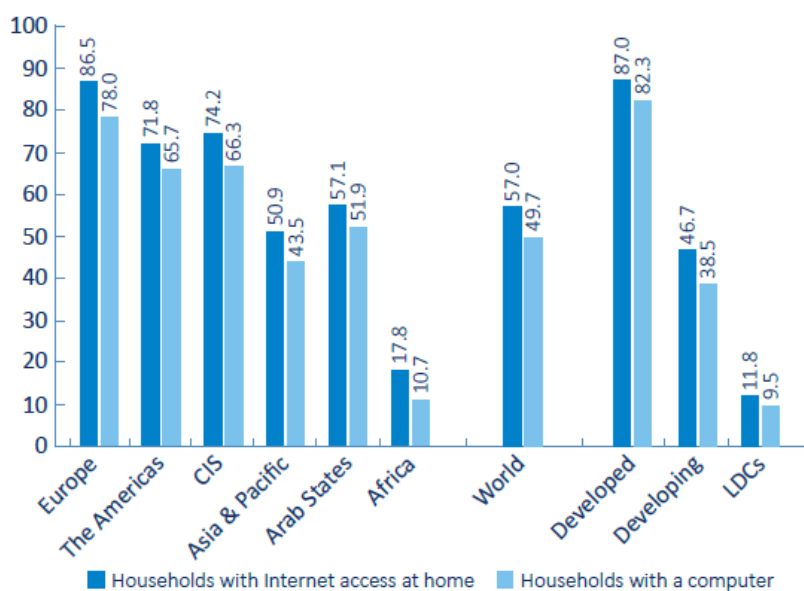


Figure 6. Percentage of households accessing the Internet at home. Adapted from “Measuring Digital Development: Facts and Figures 2019,” by International Telecommunications Union, 2019. Copyright 2019 by the International Telecommunications Union.

Data sources are also generated by organizations that provide products and services to the consumer directly. These companies exist in both private and public sectors. Businesses in the private sector, such as Internet providers, social networks, and those in healthcare, insurance, banking, retail, or telecommunication, often interact with consumers directly. This allows them to gather their customer data directly without going through third-party companies. These companies might also buy and share consumer data with other businesses to have more relevant information. For instance, life insurance companies may collect customer medical records from hospitals or doctor offices to have a general understanding of customer health conditions, so that they can issue suitable insurance policy as well as determine appropriate prices. In another case, retailers might gather information from banking and from social networks so that they can better understand customer expenditures and recommend products to target these customers. Customers may be aware of their data being collected by companies, but they may not fully understand how their personal data are transmitted among companies across business sectors.

### **Data Providers**

Besides companies that interact with customers directly, many data-provider companies operate as a middlemen within a data-driven ecosystem; consumers may not be aware of these businesses. These companies collect and aggregate data from multiple sources, organizing the data as well as extracting new information from it. Consumer data may be collected from both public and private sources. The capability of new technologies, such as big data and data mining technologies, allows data-provider



companies to process large volumes of data and unveil new information that is valuable for businesses. Finally, data providers sell processed data in useful formats to other companies. Companies can buy processed data for many business purposes, such as advertising, financial fraud detection, product quality improvement, and many other purposes.

### **Data Buyers**

Internet technology transforms traditional companies into data-driven businesses. Instead of operating businesses at physical locations, many companies leverage Internet technology to transform their businesses in the digital network environment. Internet technology provides companies the capability to present business information, capabilities, products, and services to large numbers of customers. Other emerging technologies, such as big data and data mining technologies, allow companies to collect and analyze consumer data in order to better understand business trends and customer preferences (King & Forder, 2016). Companies in different sectors, such as manufacturing, marketing, online retail, or banking, depend on consumer information to optimize their businesses, improve product quality, or enhance customer experiences. For instance, banking institutions collect financial data, such as lending information, investment records, insurance records, historical banking transactions, and many other types of information, to analyze stock markets, determine the risk of loan applications, or detect financial fraud. Grether (2016) said that marketing agencies want to gather and analyze consumer data to understand their spending behaviors, financial conditions, and

business trends. Such information helps advertising firms develop effective marketing campaigns to target potential customers.

### **Consumers in the Data-Driven Ecosystem**

In data-driven business model, consumers use social network platforms, search information, buy products, or pay for services through various Internet-connected devices and applications (Li, Ch'ng, Chong, & Bao, 2016). Li et al. (2016) argued that online user activities hold a critical role in the digital economy. King and Forder (2016) said that online customer previews created large volumes of data for businesses to collect and use, transforming their businesses into data-driven business models. Meulen (2017) estimated that consumers use 52 billion Internet-connected devices worldwide in 2017, which represents 63% of the total of Internet-connected units. Meulen (2017) also predicted that the numbers of IoT units demanded by consumers would be double in 2020, projecting 128 billion IoT devices. Consumers create streams of historical data whenever they interact with Internet-connected applications, such as social networks, purchasing online, electronic payment, and mobile phones, among others. Companies in data-driven ecosystems collect and use these data to improve their business processes, marketing, product quality, and customer experiences.

### **Benefits of Data**

Colombo and Ferrari (2015) said that business and consumer data, which are generated from business processes, social networks, Internet websites, or mobile devices, are valuable to businesses. Companies might collect analyze consumer data to understand customer behaviors so that they can improve product quality or enhance customer

experiences, or for many other purposes (King & Forder, 2016; Volker, 2016). They can use business data to optimize business operations and stay competitive in e-commerce. For instance, manufacturing companies might use product operational and maintenance data to predict the service life of particular products (Reinsel et al., 2018a). Innovative technologies, such as big data and predictive data analytics, provide businesses the capability of storing large amounts of data and aggregating multiple data sources to reveal hidden information (Gunasekaran et al., 2017). Consumer data are collected and used in many businesses, such as online retailers, health care businesses, and manufacturing industry. The HarrisX (2018) survey showed that 63% of participants believed that technology is good for society, and 70% believe technology has positive impact on their daily life. At the same time, small numbers of participants (14%) said technology is somewhat bad for society.

### **Risks of Collecting and using Consumer Data**

Business practices of collection of large amounts of consumer data may cause issues such as exposing personal information and violating individual privacy rights (Jai, Burns, & King, 2013; Zhu & Chang, 2016). For instance, businesses may predict that an individual is pregnant by combining age, gender, and historical buying products. Companies might also gather social network data to predict personal information, such as emotions, political points of view, religion, and other sensitive data. In the health care industry, medical insurance companies might collect hospital data and information about medical claims to identify health conditions of customers. Based on such information,

insurance companies might determine different insurance premiums based on health pre-conditions.

Kshetri (2014) mentioned that consumers often had little control over their data being collected and used by private companies. Consumers might lack the knowledge to understand the sophisticated operations of data processes. A lack of essential technical skills also limits consumers from configuring Internet-connected applications appropriately to protect their data. Kshetri (2014) argued that many companies established complicated data processes, which make it difficult for consumers to understand how their information is collected and used. Prince (2018) also pointed out that customers have limits on managing their personal information when using various platforms to conduct daily business over the Internet.

Privacy setting features are a common mechanism for controlling personal data sharing when using online applications. Prince (2018) found that online users do not pay attention to privacy setting tools to manage personal information. Customers often raise concerns about their privacy rights instead of about how to protect their personal information when using Internet-connected devices or applications (Prince, 2018). According to an online survey conducted by IBM and Harris Poll (2019), the study collected data from 1,000 adult participants in the United States regarding consumer perspectives on information privacy. It showed that 76% of participants want to have the ability to withdraw from sharing personal data with third-party organizations, and 73% want to have the ability to take back or retrieve their data.

In 2017, many governments in the European Union (EU) initiated new privacy policies to protect consumer data in the Internet environment (Butterworth, 2018). Big data technology provides companies the capability to collect, store, and analyze huge amounts of data that might be impossible to process manually by humans. One of the major concerns about this technology is that it can operate autonomously and make independent decisions without human interactions.

While collecting and using consumer information offer businesses many potential benefits in the digital economy, issues associated with using consumer data without authorization or consent are infrequently addressed (Zhu & Chang, 2016). Business leaders are in difficult situations when making decisions on using consumer data in digital business models. A survey of artificial intelligence ethics by Hashmi (2019) found that ethical issues were a significant concern with using consumer data in business and that many leaders were still unable to find suitable solutions to address them.

Innovative technologies, such as the Internet, social networks, or online video streaming, create new opportunities for online marketing agencies. Bleier and Eisenbeiss (2015) presented that many online marketing models focus on advertising personalized products and services to specific groups of customers. In order to deliver customized products to targeted individuals, companies need to collect and analyze large amounts of customer data behind the scenes. Although collecting and using consumer data provides companies many potential opportunities, these business practices prompt a number of issues in terms of ownership, transparency, ethics, and consumer privacy laws. Technology companies collect and classify customers into particular groups based on

their private information and behavior. Estrada-Jimenez et al. (2017) noted that collecting and analyzing customer data takes place in the background, and customers may have little control over the process. With help from emerging technologies, such as big data and Internet-connected devices, companies can collect, store, analyze, and exchange massive consumer data. Due to the large amounts of data, companies often use sophisticated computing algorithms to process data anonymously. This makes it difficult to keep track of specific customer data points. Companies also exchange information with partners or business customers. Estrada-Jimenez et al. (2017) argued that business practices, such as building customer profiles or exchanging consumer personal information, could lead to abuses in using consumer data and to discrimination against customers when they receive business services.

Internet-connected platforms, such as social networking websites or smartphones, have provided a convenient way for consumers to share memories and connect with loved ones. Online retailers have also offered convenient ways for customers to shop merchandise at any time on their Internet-connected devices. In order to enjoy the convenience that businesses offer, consumers have to disclose personal information, so that companies can collect private data and analyze it to understand more about consumers. The dilemma of privacy protection and convenience leads consumers to a situation in which they want control over their privacy, but also continue disclosing personal information to get benefits from online service providers (Marreiros et al., 2017). Trepte et al. (2015) argued that consumers want to protect their private information and have control over their data flow in the Internet environment. At the

same time, consumers might not be able to prevent businesses from gathering their data from multiple sources due to a lack of knowledge about privacy online as well as a lack of data protection mechanisms.

### **Privacy Risks in Digital Marketing**

In the traditional advertising approach, marketing agencies present advertisements to potential customers without customizing the content. Nowadays, new technologies provide companies the capability to tailor advertised content to reach a particular group or individual (Bleier & Eisenbeiss, 2015; Tran, 2017). Estrada-Jimenez et al. (2017) mentioned that there were many parties and partners involving in online advertising infrastructures. Marketing agencies depend on technology companies to collect and analyze customer data so that they may develop mathematical models to predict customer preferences. Consumer data points, such as website browsing history, location, search keywords, financial transaction records, and other data points, are valuable to companies working to engage customers effectively (Grether, 2016). Based on such personal information, online marketing organizations may distribute suitable advertisements to customers at the right time.

Smit, Noort, and Voorveld (2014) discussed that Internet-connected applications and devices generate huge amounts of personal data that allow companies to collect and track online activities of consumers in the real-world. Companies can build a complete individual profile based on how customers search for information online, as well as on shopping behaviors, age, gender, and other personal data (Smit et al., 2014). Companies can collect these personal data directly from online applications or buy them from data-

provider companies. Estrada-Jimenez et al. (2017) mentioned that privacy risks existed when personal data are collected and shared among companies without the awareness of customers. Trepte et al. (2015) noted that online users often lack knowledge of how personal data flow in the Internet environment. This leads to online users tending to allow businesses to collect personal data without knowledge of how businesses use their data. A lack of knowledge about privacy also prevents consumers from maintaining control over personal data or protecting their data from cybersecurity threats. The risks might also be associated with privacy rights, discrimination, fraud, and unauthorized access.

A mobile privacy report by the Federal Trade Commission (2013) presented that big data technology provides companies the capability to build individual profiles and categorize the profiles into different groups. By doing that, private companies can advertise to a particular group with new products or services. Butterworth (2018) argued that this business practice might result in harm to consumers due to the inaccuracy of data. Companies may also obtain consumer data from internal sources such as data-provider companies. These secondary data sets may contain inaccurate information. Analyzing low-quality data may lead to biased results.

### **Sharing Personal Information on the Internet**

A report about big data technology by the White House (2014) discussed many privacy concerns when consumers share their personal information through Internet-connected applications and devices. Private companies may use consumer data for different business purposes as well as share customer information with business partners. The IBM and Harris Poll (2019) showed that less than one third (31%) of participants



believe that their personal data remain within the original company that they shared data with. Schudy and Utikal (2017) noted that consumers are willing to share personal information in exchange for the benefits and convenience of Internet-connected applications or mobile devices. Consumers also tend to share information with unknown parties or people on the Internet. For instance, people feel comfortable sharing information on social networks. Schudy and Utikal (2017) mentioned that customers might not know that their personal data is being shared with companies that they do not have business with. This business practice may expose the personal information of consumers to unauthorized parties.

### **Consumers Managing and Controlling Their Personal Data**

The ability to have authority over one's personal data is a critical element in data privacy. In addition to managing personal data, consumers want to control how their personal data is used and shared among businesses. Prince (2018) pointed out that online users want control over their personal data when interacting with Internet-connected platforms. Prince (2018) said that there is a significant relationship between self-disclosure and privacy controls. Online users who are likely to reveal and share personal information on the Internet network tend to use privacy setting features to control their information sharing. Prince (2018) said that personal data have a variety of values and weights depending on different types of data. Some pieces of information can be more sensitive than others. For instance, online users pay more attention to health or financial records than text messages or virtual identities. Since consumer data could be valuable assets to businesses, businesses have been collecting, analyzing, and selling it to

advertising companies. Prince (2018) argued that businesses should provide effective privacy control mechanisms that allow online users to configure the privacy settings for their data. Individuals who are concerned about privacy are more likely to want to manage and control data when disclosing private data on the Internet.

### **Current Data Protection Laws**

The numbers of data breach incidents that have occurred recently at both public and private organizations have raised concerns over how data privacy legislation and policies protect the personal data of consumers in the digital environment. Consumers continue using Internet-connected devices and sharing personal information over the Internet at higher rates (Mulligan et al., 2019a). Consumer data are collected directly by business service providers as well as third-party data providers. Mulligan et al. (2019a) mentioned that business practices of collection and use of consumer data exposed personal information to unauthorized or third-party organizations. There are numbers of personal data protection laws at the federal levels in the United States.

Table 1 includes current data protection laws in the United States. However, these federal data protection laws are not comprehensive to regulate businesses where they collect and sell consumer data for business purposes (U.S. Senate Committee on Commerce, Science, and Transportation, 2013). Data protection laws differ in terms of industries, scope, law enforcement, and penalties (Mulligan, Linebaugh, & Freeman, 2019b). There is also lack of unified data protection laws at the federal level that regulate how private companies collect and use consumer data in the digital ecosystem (Mulligan et al., 2019a). Current U.S. privacy laws do not provide consumers the right to know

what data companies have collected about them and how they use it (U.S. Senate Committee on Commerce, Science, and Transportation, 2013). Consumers do not have the capability to verify or make corrections to their data being collected by companies.

Table 1

*U.S. Federal Data Protection Laws*

Data Protection Laws	Industry	Description
Gramm-Leach-Bliley Act	Finance and Banking	This law prohibits financial institutions from sharing nonpublic personal information (NPI) with third parties. It requires financial institutions to provide consumers with their policies for how they collect, use, and share NPI.
Consumer Financial Protection Act	Finance and Banking	This Act regulates financial institutions offering financial services to consumers with unfair or abusive practices.
Children’s Online Privacy Protection Act	Online Businesses	This Act regulates the business practice of collecting and using children’s information on the Internet. The law prohibits companies from collecting information about children who are under the age of thirteen without obtaining consent from their guardians.
The Communication Act	Telecommunication	This law regulates data privacy from telephone companies that provide interstate communication. These companies may not use or share personal information without customers’ consent. This Act does not cover television, radio, or Internet providers.
Video Privacy Protection Act	Video Service Providers	This law prohibits video service providers from disclosing personal identifiable information or information regarding rental and purchased video materials. This Act provides a privacy right to media consumers. It does have law enforcement as well as penalize violations.

*(Table continues)*

Data Protection Laws	Industry	Description
Fair Credit Reporting Act (FCRA)	Credit Reporting	This Act regulates entities who operate in credit reporting (CR) services, provide information for CR, or use credit reporting. This Act does not prohibit CR agencies from collecting, using, or sharing consumer data with third parties. Instead it requires CR agencies to report consumer credit information accurately. This Act requires entities to use credit information only for appropriate purposes, such as financial transactions.
Federal Securities Laws	Across Industries	These laws regulate how companies protect against, prevent, and respond to data breach incidents.
Computer Fraud and Abuse Act	Consumer and Provider	This law prohibits unauthorized access to computer systems. It does not enforce data protection or data security. This Act allows individuals to file legal actions when companies collect online activities without their authorization.
Electronic Communications Privacy Act	Across Industries	This Act prohibits unauthorized access to electronic communication channels, such as telephone, email, or digital database systems. This Act provides a general approach to protecting private conversations, which are made, transited, or stored on electronic networks.
Federal Trade Commission Act	Across Industries	The Act was signed into law in 1914, which established the Federal Trade Commission. In terms of data privacy and security, the Commission prevents companies from collecting, using, or disclosing the information of other organizations or businesses.

*(Table continues)*

Data Protection Laws	Industry	Description
Family Educational Rights and Privacy Act	Education	This Act prevents educational institutions from releasing student educational records without the consent of parents or students. Parents or students have the right to review records as well as challenge them when they are inaccurate or incorrect.
Health Insurance Portability and Accountability Act	Healthcare Providers	Under this Act, healthcare providers need to protect health records. It generally prohibits healthcare providers from using or sharing a patient's information with third parties without their consent. Healthcare providers need to notify patients when disclosing information to third parties for treatment or payment purposes. Patients also have the right to request that healthcare providers provide them with copies of their own records.

*Note.* Adapted from “Data Protection Law: An Overview,” by S. P. Mulligan, C. D. Linebaugh, and W. C. Freeman, 2019a. Retrieved from <https://crsreports.congress.gov/product/pdf/R/R45631>

Outside of the United States, other countries also have raised concerns about data privacy and introduced new comprehensive data protection frameworks to protect personal information. In 1995, the European Union introduced the Data Protection Directive laws, which regulate how companies process personal data (Ott & Zylberberg, 2016). For instance, articles 10 and 11 of the Protection Directive require organizations that do not obtain data directly from individuals to provide data owners the organizational identity, the purposes of using the data, as well as the right to access to their own data (Cradock, Stalla-Bourdillon, & Millard, 2017). In 2018, the EU established a comprehensive data protection framework, named the General Data Protection Regulation (GDPR), to regulate how businesses can collect and process personal

consumer data (Tikkinen-Piri, Rohunen, & Markkula, 2018). Businesses have the responsibility to implement appropriate information management systems to prevent the misuse of personal data as well as to protect data from unauthorized access (Tikkinen-Piri et al., 2018). With GDPR policies in place, EU citizens have the right to access, remove, or retrieve personal data stored by businesses. While U.S. data privacy laws focus on the ability of the U.S. government to access the private lives of citizens or on regulating specific business sectors, the GDPR provides individuals a general right to protect their privacy. The GDPR also regulates any organizations that collect large amounts of personal data in the EU (Mulligan et al., 2019a).

The GDPR applies to any businesses if they collect, store, and use information of individuals from the EU, regardless of the locations of these organizations (Mulligan et al., 2019a). Personal data under this law can be any information used to identify an individual, such as name, address, phone number, IP address, and many other pieces of information. Since non-EU companies can offer businesses and services to EU citizens over the Internet and collect personal information, these companies need to comply with the GDPR even they locate, manage, and process the information of individuals from the EU outside of EU countries. Mulligan et al. (2019a) noted that the GDPR introduced several key principles related to the processing of personal data, such as transparency, the purpose of using data, the accuracy and limitations of data, and data security and accountability.

## **Transparency of Processing Personal Information**

Custers, Dechesne, Sears, Tani, and Hof (2018) mentioned that the transparency condition in personal data processing is one of the key principles of the GDPR. It is not only a designable condition in the privacy protection laws, but it is also a success factor that allows consumers to check how businesses comply with privacy laws. By GDPR guidance, businesses need to develop transparent data management processes when handling consumer privacy information. For instance, businesses need to notify consumers what information they collect and how they use it. Businesses also need to provide customers mechanisms that allow them to access and manage their own data (Custers et al., 2018). The transparency in using consumer data allows not only companies to ensure compliance with regulation but also to establish trust with their customers.

Chua et al. (2017) discussed that the notification feature is critical in business practices of collection and use of consumer data. It might help customers be aware of what happened to their data. Companies have the responsibility to notify customers what personal data they collect and how they use these collected data sets (Chua et al., 2017). Customers are then able to make appropriate decisions on sharing data with online business providers.

Cradock et al. (2017) mentioned that innovative technologies in the data management domain developed and advanced rapidly. The Organisation for Economic Co-operation and Development (2014) reported that innovative technologies advanced in collecting, aggregating, and analyzing personal data without the knowledge of individuals

over the last twenty years. The big data analytic algorithms become sophisticated, which provide businesses the capacity to capture and link information together from multiple sources. Due to huge amounts of data being gathered and fed into these big data analytic processes, consumers may not know what personal data are being collected and used by private companies (Cradock et al., 2017).

Cradock et al. (2017) argued that private companies have the responsibility to inform consumers what personal data they collect and process. Otherwise, consumers have little control over the process of collecting and using their personal information. The non-transparency of using consumer data might cause issues regarding the compliance of private companies. The goal of transparency in consumer data-driven practices is to allow consumers to identify the compliant companies that they can trust to share personal information (Cradock et al., 2017).

Cradock et al. (2017) also showed that there was uncertainty over the obligations of organizations to inform individuals about the practice of using personal data. This leads to reducing the transparency for individuals to understand how their personal data are being collected and used within companies. Governments might need to develop new policies to clarify the obligations of companies to inform individuals about the use of their personal data.

### **Purpose of Using Consumer Data**

Libaque-Saenz, Chang, Kim, Park, and Rho (2016) showed that companies might collect and use consumer data for many business purposes, such as online marketing, decision-making systems, and customer engagement. These data points provide



companies efficient ways to understand customer behavior and preferences. For instance, many Internet websites allow customers to post previews of particular products or services after they use it. Such feedback data offers rich information that companies may use to improve customer experiences.

Butterworth (2018) suggested that companies should explicitly define the purposes of collecting individual data and not use these data sets for other undefined purposes. Due to the capability of big data to process large data sets, companies intend to collect many consumer data sets and use these collected data for different business purposes. Companies might also aggregate consumer data from multiple sources and generate new data sets. Some companies even sell these new data sets or the results from analyzing these data as new products and services. Butterworth (2018) noted that the key factors in using personal data for purposes that diverge from the original purpose are fairness and compliance with the laws. When companies collect personal data for a particular purpose based on customer consent, these collected data should only be used for that original purpose. Companies might need to notify customers and obtain new agreement from them in the case that consumer data is used for different purposes.

### **Big Data Analytic Quality**

Big data technology has been providing companies the capability to collect, store, and analyze a large amount of data in both structured and unstructured formats from multiple data sources. Each data source contains a piece of information related to a particular individual or a group of customers. The FTC (2013) reported that when businesses aggregated these pieces of consumer data, they were able to understand

individual behaviors and build complete profiles. Organizations develop customer profiles using many data points of personal information such as financial records, shopping behavior, online spending time, and other personal information (Federal Trade Commission, 2013). Baesens et al. (2016) argued that the information might be inaccurate due to the large volume of these data sets. Analyzing these low-quality data sets may generate inaccurate results (Baesens et al., 2016). Organizations should collect data that accurately present their customers to train computing algorithms so that they can generate accurate results (Ginosar & Ariel, 2017).

### **Consumer Consent**

Companies use individual consent to comply with federal laws when collecting and using consumer data for a particular business purpose. Consumers often need to accept a privacy statement or data policy when they want to use products or services offered by the firms (Aimeur, Lawani, & Dalkir, 2016; Marreiros et al., 2017). Altman et al. (2018) argued that relying on privacy statements to obtain an agreement from customers does not adequately protect consumer privacy data, since customers may not fully understand the broad terms of these documents. Consumers may risk exposing personal information to unknown parties when giving their consent to companies for collecting and using personal data. Butterworth (2018) noted that companies need to gather and process personal data based on individual consent. Individual consent should be meaningful and unambiguous so that consumers can understand precisely the purposes of collecting and using their personal information. One of the key advantages of big data technology is that it combines massive data from multiple sources to generate new types

of data. In this case, the original consent for collecting and using consumer data may not reflect exactly the purpose for which companies plan to obtain data.

Governments often recommend that private companies disclose business practices of collecting, storing, using, and sharing consumer data. Privacy laws such as the Consumer Privacy Bill of Rights or Fair Information Practice Principles recommend that companies establish appropriate data management processes to protect consumer data from unauthorized access or data breach incident (White House, 2014). Protecting consumer data from data breaches is one of the key factors in e-commerce industry. The cybersecurity online survey by IBM and Harris Poll (2018) reported that 75% of participants do not purchase products when they do not trust a company to protect their personal data. This cybersecurity survey also showed 85% of participants agreed that businesses should prioritize cybersecurity strategy over a focus on profit. 71% of them expected the government take appropriate actions in the case that cybersecurity is compromised. The challenge for collecting consumer data based on individual consent is that companies need to define the purpose of using personal data before starting to collect data. It is challenging to develop meaningful consent if companies do not have a clear purpose and object for collecting and using consumer data. It is also challenging to re-obtain consent when companies did not collect data directly from customers. For instance, private companies can buy consumer data from data-provider companies and aggregate them with their internal data to create rich data sets. In these cases, companies may not obtain consent from customers who own the data.

Butterworth (2018) mentioned that people might not read privacy statements when they interact with business websites. In other cases, customers have to accept privacy terms; otherwise, they cannot access Internet-connected applications or devices. Marreiros et al. (2017) noted that customers simply accepted the privacy terms in order to use particular online services. In these cases, the consent may not be provided voluntarily because customers have no other selections. Butterworth (2018) discussed that consent is necessary whenever companies need to process personal data to provide services to customers. Otherwise, companies should not present privacy statements and ask customers to agree with the terms. Companies should not bundle privacy terms with other business services that do not require access to personal data (Butterworth, 2018).

### **Current Approaches to Personal Data Protection**

Current studies propose a number of privacy protection approaches (Broeders et al., 2017; Federal Trade Commission, 2011; Ginosar & Ariel, 2017; White House, 2013). Many approaches focus on blocking technology companies from collecting data on the user-side. Online users can configure their Internet-connected devices to prevent technology companies from gathering personal information when doing business online. Trepte et al. (2015) pointed out that privacy knowledge could help consumers balance private information management by sharing needed information only in order to access digital businesses. Butterworth (2018) argued that companies should provide customers the capability to withdraw consent that they provided in the past. This would allow customers to give consent voluntarily as well as to withdrawing freely from privacy terms agreements previously agreed to.

As data breach incidents occur frequently, companies begin tightening data protection processes, developing internal data management review, and establishing restricted privacy policies. For instance, many companies allow consumers to reject company websites or mobile phone applications that collect their personal data. Some companies also use sophisticated computing algorithms such as the Randomized Aggregatable Privacy-Preserving Ordinal Response technique, to protect customer sensitive information automatically (Erlingsson, Pihur, & Korolova, 2014).

Estrada-Jimenez et al. (2017) proposed three main solutions to protect consumer data from digital advertising, which include protection parameters, academic research, and commercial solutions. The protection parameters include location, scope, and strategy. In terms of academic research, there were numbers of current research technologies that can help to secure personal data in the online environment. The academic research includes oblivAd, AdJail, and Privad, as well as many others. Technology companies also provide many commercial platforms such as Google Sharing, Brave, Ghostery, and so on.

In terms of the data types, companies might limit themselves to collecting and using personal data that are relevant to business objectives. They should restrict the scope of data-collecting processes to the business purpose described in the individual consent. Organizations should define the purpose of using consumer data and collect data only for that business goal. This data limitation principle might reduce the capability of big data technology, which requires a large amount of data to extract insights. The challenge of

this requirement is that companies collect many available data sets rather than limiting the data collected to a particular purpose.

Butterworth (2018) discussed the need for companies to establish an internal data management board to oversee the process of collecting and using consumer data in compliance with legislation. The data management team needs to consider business ethical aspects of collecting and using private data. The purpose of collecting and using personal data needs to align with the privacy statement that companies provide to customers. For example, the location data generated by the mapping application should not be sold to a travel agency to advertise a particular vacation location.

### **Gaps and Study Findings Related to the Literature**

New emerging technologies such as the Internet, big data, social media networks, and smartphones allow companies to collect large amounts of digital data associated with consumer activities, online and offline (Baesens et al., 2016). Companies collect large volumes of consumer data and sell them as data products to other firms. Companies might not obtain data directly from consumers. Instead, they might buy and gather consumer data from third party data providers (Perko & Ototsky, 2016). These business activities may not be transparent and may remain hidden from consumers (Cradock et al., 2017; U.S. Senate, Committee on Commerce, Science, and Transportation, 2013). Companies also aggregate data to create individual profiles or build consumer scoring systems. These data products might put consumer privacy at risk in terms of misuse, discrimination, and unauthorized access. There is a lack of understanding of the business practices of collecting and using consumer data without an individual's consent. The

findings of this qualitative archival research filled the gap about business practices in collecting and using consumer data without the knowledge of consumers.

**Finding 1. Potential benefits.** Business practices of collecting and using consumer data for marketing or improving products have occurred for a long time (Grether, 2016; The U.S. Senate Committee on Commerce, Science, and Transportation, 2013; Tran, 2017). Both public and private organizations have collected and used consumer data for many business purposes (Altman et al., 2018). Finding 1 showed that companies used consumer data to understand marketing trends, verify financial credits, develop marketing strategies, and improve product quality along with customer service. Personal data became valuable as a source of future profits for many businesses in the digital economy (Colombo & Ferrari, 2015; Marreiros et al., 2017; Prince, 2018).

Finding 1 confirmed that consumer data could provide a variety of potential benefits to companies operating in the data-driven industry. According to Bleier and Eisenbeiss (2015), consumer data allowed marketing companies to collect and understand consumer behaviors and preferences. It helps companies customize advertising to target potential individuals and particular groups. Nunan and Domenico (2013) studied the collection and use of personal information for marketing purposes. They discussed the privacy risks and challenges of using big data to collect customer data for marketing research. Jobs et al. (2015) studied marketing ecosystems where companies use big data technology to process consumer data for online advertising. Jobs et al. (2015) argued that marketing companies might influence potential customers if they invest big data appropriately.

**Finding 2. Data from multiple sources.** This finding filled the gap of knowledge about data sources where companies obtained consumer data. King and Forder (2016) said that online customer previews created large volumes of data for businesses to collect, use, and transform their businesses into data-driven business models. Prince (2018) explored the needs of online users to help them control and manage their personal data in the digital environment. There was a lack of knowledge about actual sources of consumer data. Finding 2 showed that companies collected consumer data from public records, local governments, private companies, third-party providers, and directly from consumers. Companies used different methods and techniques to obtain detailed information on individuals. Collected personal data elements included identification, Internet-connected devices, incomes, education, medical records, finance, lifestyles, and many other personal data elements.

**Finding 3. Identity information.** Online users were concerned about their private data being disclosed on the Internet. At the same time, they shared detailed personal information when browsing the Internet or using applications in the online environment (Trepte et al., 2015). Trepte et al. (2015) explored the difference between online users' privacy attitudes and behaviors in the Internet environment. A study by Ott and Zylberberg (2016) explored the ownership of data in the digital ecosystem. Colombo and Ferrari (2015) studied the aspect of data protection in big data technology to address the lack of data protection awareness and data access control in the digital environment. They explored the integration of data privacy and protection in big data platforms.



Finding 3 extended knowledge about data ownership and identity information in the data-driven industry. This finding showed that data companies used various personal identity information, such as name, address, email, phone number, government-issued identifiers, and credit card information to verify individual identities. Besides these identity data elements, companies also aggregated data from different sources, including public organizations, private companies, social media networks, surveys, or consumers themselves. Each data source provided bits of identifiable information about contact information, employment, device information, financial data, and online activities. Emerging technologies, such as big data, machine learning algorithms, or computer artificial intelligence, provided companies the ability to combine these data elements and identify an individual precisely.

**Finding 4. Transparency in data-driven businesses.** Cradock et al. (2017) pointed out that transparency is one of the fundamental principles in the European Union data protection laws. These laws require businesses to notify their customers when they collect and use consumer data. However, it is not clear what type of personal data companies collect that require them to inform their customers. A study by Butterworth (2018) discussed the difficulties in guidelines by the U.K. Information Commissioner's Office regarding the use of artificial intelligence to process personal data. The study focused on the fairness of using personal data in technology organizations. This study also evaluated the ethical aspect and social impact of using personal data.

Finding 4 extended the knowledge about transparency in the data-driven industry. Finding 4 showed that data collection processes needed to be transparent when

companies gathered or bought personal information from other companies. Although data provider companies revealed general purposes of using consumer data, companies often did not explicitly disclose either their data sources or client information. They also had contracts with their clients that prohibited clients from disclosing data sources. Finding 4 also confirmed with a study by Marreiros et al. (2017) which examined the effect of privacy policies on user behaviors surrounding personal information disclosure and sharing on the Internet environment. Finding 4 showed that privacy statements were often lengthy and difficult for consumers to understand, due to its technical terms and general agreements. Marreiros et al. (2017) asserted that online users were more careful about disclosing personal information whenever privacy concerns were discussed broadly.

**Finding 5. Data products and services.** A study by Li et al. (2016) examined the impact of product category, discount, and customer service on online sales performance. The study also identified the role of product reviews and ratings in improving sales performance in the data-driven business model. At the same time, Gunasekaran et al. (2017) suggested that companies might adopt big data to improve their business performance and transform supply chain management systems by examining the impact of big data on supply chain and organizational performance. There was a lack of knowledge about data products and services offered by data providers.

Finding 5 extended knowledge about the existing consumer data products offered by data companies. Companies developed data products and services based on two types of data, actual data, and processed data. Third-party companies or marketing agencies could buy individual profiles, consumer scoring systems, or predictive modeling products

to forecast common characteristics, health conditions, financial status, or lifestyles of people living in a particular city or zip code. Grether (2016) mentioned that companies could use big data to collect real-time and comprehensive consumer data to stay ahead of competitors. Third-party companies can profit from consumer data by acquiring, transferring, and selling consumer data for marketing purposes (Grether, 2016).

Regarding the challenges of using big data in market research, Volker (2016) noted that although big data offers potential benefits for marketing companies, many aspects of big data need to be considered when managing massive volumes of data for market research. The challenges include the representation of data, data accuracy, and key variables.

Finding 5 showed that companies monitored individuals' online and offline activities to understand their characteristics, behaviors, or lifestyles. These sensitive information provided companies the capability to build scoring systems that might be used to predict consumer preferences and target them at the right time.

**Finding 6. Affiliates and partner networks.** There was lack of understanding about partner networks of data providers. Hartmann et al. (2016) developed a framework to analyze and classify data-driven business models where companies use consumer data to operate their businesses. Hartmann et al. (2016) data-driven business model framework contributes to an understanding of start-up companies that use data as core business assets to establish business objectives. Finding 6 extended knowledge of data-driven networks by providing information about clients and affiliates of data companies. Data companies often did not disclose in detail their clients who bought data products for engaging potential customers, fraud detections, identity verification, or other business

purposes. According to privacy policy and data catalog documents collected from the 12 companies in this study, finding 6 showed that clients of data companies might operate in a wide range of sectors, such as insurance, healthcare, travel, restaurant, or finance. They relied on consumer profiles, predictive models, tracking data, or scoring systems to forecast business trends, consumer preferences, or improve internal business operations. Finding 6 also presented that consumers might not know which third-party companies would buy these data products and how third parties used data products to make decisions against them.

**Finding 7. Purposes of using consumer data.** A study by Wamba et al. (2017) examined the impact of big data analytic capability on company performance. Wamba et al. (2017) claimed that big data analytic capability has a significant impact on company performance when investing adequately in big data infrastructure and technical management knowledge. Jai et al. (2013) examined the effect of online users' behavior on shopping experiences and repurchase intention in e-commerce. Previous studies have investigated the performance of personalized advertising in the Internet environment. Tran (2017) discussed the lack of understanding of personalized advertising on the Facebook social media network. Tran (2017) developed a model to capture the effect of customized advertising on the attitude and behavior of Facebook users. Tran (2017) argued that personalized advertising on Facebook had a significant impact on viewers and improved customer responses. Zhu and Chang (2016) examined the impact of using personal data to personalize digital advertising on consumer privacy. Besides, they explored the relationship between personal advertising and online user perspectives on

privacy concerns. Bleier and Eisenbeiss (2015) investigated the impact of personalized advertising on customer responses. They pointed out that many online marketing models focus on using consumer data to personalize advertising on the Internet.

Besides using data to personalize advertising or improve customer experiences, finding 7 extended knowledge of using consumer data, explore two core purposes that drive companies to use consumer data. These purposes included internal business and external business purposes. For internal business purposes, companies used consumer data to improve business operations, audit business transactions, enhance customer service, or comply with current regulations. For external business purposes, companies used consumer data to build individual profiles, consumer scoring systems, or predictive models. Other businesses in different industries bought these data products to forecast business trends, manage supply chains, and use them for other business purposes, including marketing, fraud prevention, identity verification, or risk mitigation. Internal business operation data and external information could help business leaders make quick decisions to respond to current market demands.

**Finding 8. Complexity of the data-driven industry.** Private companies increasingly digitalized business processes and transformed their core businesses based on available data to compete with rivals and to stay competitive in the data-driven ecosystem (Gantz, Reinsel, & Rydning, 2019; Reinsel, Gantz, & Rydning, 2018a). Businesses want to use consumer data so that they can personalize products to meet individual preferences. Jobs et al. (2015) and Perko and Ototsky (2016) mentioned that companies use consumer data from various sources to improve business operations,

product quality, customer services, and many other business aspects. Finding 8 extended knowledge about businesses that operated in the data-driven industry. Data companies could acquire consumer data from the government, buy data from other companies, exchange data with affiliated partners, or collect data directly from consumers. Finding 8 showed that data provider companies made profits based on data that were either voluntary or involuntary collected from consumers. They collected large volumes of consumer data and sold it to other companies for various purposes. Data companies also provided and exchanged data with each other to enrich their data sets. Besides affiliates and partners operating in the United States, data providers also sold their data products to foreign companies.

**Finding 9. Ownership and consent.** Butterworth (2018) noted that companies need to gather and process personal data based on individual consent. Individual consent should be meaningful and unambiguous so that consumers can understand precisely the purposes of collecting and using their personal information. Smit et al. (2014) discussed that online users were concerned about private information when sharing personal data on the Internet. Smit et al. (2014) focused on online users' knowledge of private consent, presented to them by companies. Smit et al. (2014) showed that it is critical that online users understand the data processes where their personal data are collected and used by technology companies. A study by Schudy and Utikal (2017) addressed the lack of understanding of what factors form online users who were willing to disclose their personal information. In order to understand privacy issues when using social networks to recruit participants for health research, Bender et al. (2017) developed a privacy-by-

design framework. This privacy framework provides guidelines to mitigate privacy risks and to protect the sensitive information of health research participants. Bender et al. (2017) notified participants about privacy risks of disclosing sensitive information through social networks.

Finding 9 confirmed previous studies about ownership and consent in practices of collecting and using consumer data. Finding 9 showed that third-party companies collected and used consumer data without individuals' knowledge and consent. Companies used various technologies to collect consumers' data whenever they accessed the Internet via mobile applications, tablets, wearable devices, or computer web browsers. Finding 9 showed that consumers might have known that companies collected their online activity data when they accessed the Internet via various Internet-connected devices. However, consumers may not know the purpose behind companies' intended use of their personal data. Companies could track online user activities across Internet-connected devices and applications. They could follow an individual through multiple devices, including a laptop, tablet, or mobile phone. In many cases, companies operating in the data-driven economy did not gather data directly from consumers. They instead bought consumer data products from data provider companies without consumer knowledge.

**Finding 10. Access and control over personal information.** Kshetri (2014) and Gupta and Schneider (2018) claimed that the access control strategies used by private companies to protect consumer data were not adequate or sufficient. Marreiros et al. (2017) noted that consumers might not have a complete picture of how their personal data

are being collected and used by businesses. A study by Estrada-Jimenez et al. (2017) provided an overview of current privacy risks existing in the practice of collecting, tracking, and sharing customer data for online marketing purposes. Estrada-Jimenez et al. (2017) mentioned that online users did not have control over their personal data and lacked awareness about data collection. Online users also lacked the technical knowledge to protect their data on the Internet environment.

Finding 10 confirmed claims about consumers having limited control over their personal information in previous studies. Depending on consumers' locations, they could not access, update, or remove their data from companies. Consumers could submit opting out requests by asking companies to stop sharing their data with other companies. Although companies provided consumers with options to opt out of their processes of collecting and selling data to third-party companies, finding 10 showed that the opt-out request was only executed at the company where an individual submitted the request. Consumers needed to find and submit opt-out requests at all companies if they did not want private companies to use their data. It may be impossible for an individual to contact every company operating in the information industry. Finding 10 showed that consumers might be out of control over their personal data as soon as they decide to share information with companies. Especially in the digital environment, digital data can be transmitted and exchanged among businesses easily.

**Finding 11. Risks and ethics of using personal data.** The existing literature provides various findings in terms of data privacy, privacy protection, consumer information, and other aspects of the data-driven economy. Kshetri (2014) examined the



relationship among big data characteristics, data privacy, and security with regards to collecting, using, and sharing private data. Kshetri (2014) showed that third-party companies might misuse consumer data. The use of big data technology to collect personal data could post risks to consumers (Altman et al., 2018; Kshetri, 2014). Prince (2018) examined the impact of factors, such as privacy, data sharing, incentives, and personal information, on controlled efforts by online users who use privacy settings to manage and to protect their personal data. Hashmi (2019) addressed ethical roles in the development process of artificial intelligence applications. Hashmi (2019) argued that the adoption of AI technology to process personal data might have ethical issues regarding the legal and social aspects of communities.

Finding 11 extended knowledge about purposes of using personal data that may cause harm to consumers. Data breaches were one of the main concerns in terms of collecting and using consumer data. Finding 11 showed that data breach incidents happened frequently and exposed personal data to unauthorized parties. Identity theft may affect financial reputations of consumers by being denied from loan applications, rejected employments, and disqualified from renting apartments. Through data breach incidents, consumers might have a glimpse of their personal data being collected and used by unknown organizations. Finding 11 also showed that unauthorized access to personal data might pose serious threats to consumers. Third-party companies might misuse personal data or obtain unauthorized access to sensitive information, such as financial status or health conditions. For example, an individual might be denied opening a line of credit due to inaccurate information within credit data products. The particular

individual may not know exactly why he or she was rejected or where to correct the information.

**Finding 12. Data protection regulations.** Mulligan et al. (2019a) mentioned that there was a lack of unified data protection laws at the federal level that regulate how private companies collect and use consumer data in the digital ecosystem. Current U.S. privacy laws do not provide consumers the right to know what data companies have collected (The U.S. Senate Committee on Commerce, Science, and Transportation, 2013). Regarding the lack of privacy in big data technology, a big data study by King and Forder (2016) explored privacy concerns in using big data to collect and analyze consumer data so that companies could build individual profiles. The study developed foundational privacy principles for lawmakers to establish privacy policies. The aim of the King and Forder (2016) study was to help legislators draft appropriate privacy laws to protect consumer data.

Finding 12 confirmed previous studies about the lack of comprehensive federal data protection laws in which giving consumers the right to control and protect their personal data from being collected by private companies. There are different laws for regulating consumer data processing in different businesses, sectors, and industries. Finding 12 showed that each data company had its own privacy policies to control and manage consumer data. Businesses in the data-driven economy promote self-regulation policies to control and protect consumer data. There were no federal regulations that govern consumer data products, such as predictive models, consumer scoring systems, or

individual profiles, offered by data providers. U.S. legislations have a variety of separated consumer data protection laws.

### **Summary and Conclusions**

This chapter covers the big data ethical conceptual framework, which provides a foundation for building new knowledge about consumer data in the digital economy. Data-driven business models use consumer data to engage potential customers and provide products that meet customer preferences. While new technologies such as big data or predictive data analysis, provide benefits to data-driven businesses, these business practices lead to many concerns in terms of consumer rights and privacy protection. This chapter also includes various aspects of the data-driven ecosystem. It provides an in-depth explanation of benefits, risks, and concerns regarding the collection, use, and sharing of consumer data on the Internet. I also discussed existing data protection laws in the United States and the European Union which regulate the process of managing consumer information. In Chapter 3, I present the rationale for my research design as well as discuss the archival research method, including document selection logic, research instrumentation, data collection procedures, and data analysis plan.

### Chapter 3: Research Method

This chapter includes an overview of the archival research method chosen for this study. I present the rationale for selecting the archival method over other traditional qualitative approaches. I discuss different elements of the archival research method, including document selection logic, research instrumentation, and data analysis processes. I also discuss the role of the researcher as well as trustworthiness and ethical concerns associated with the selected research method. The purpose of this qualitative archival research is to explore private company practices of collection and use of consumer data without an individual's consent. It focuses on exploring the central research question:

*RQ:* How do private company practices of collection and use of consumer data without an individual's consent align with the ownership, transparency, ethics, and consumer privacy laws?

Existing documents from different sources are the primary data sources for answering this qualitative archival research question. Stan (2010) discussed that archival documents might provide important evidence to study past activities and reflect it to the present events. Researchers can gather pieces of historical information from multiple sources to gain an understanding of organizational businesses, public agency activities, or individual goals.

#### **Research Design and Rationale**

The research design is a blueprint for research that describes precisely how researchers conduct a study (Harindran & Chandra, 2017). Researchers use it as a

foundation to collect evidence, analyze data, and present the findings. According to Lipu, Williamson, and Lloyd (2007), when researchers establish a research design that presents each component of a study clearly, it can help them express the research problem, define the research question, collect data, and discuss findings accurately. A quality research design can help researchers answer the research problem accurately and deliver quality results (Harindran & Chandra, 2017). It also provides researchers directions to ensure the validity and reliability of research findings. Generally, researchers can choose one of three research methods, including quantitative, qualitative, and mixed methods, to conduct their studies. Researchers select one of these research approaches depending on the research problem, research question, purpose of the study, theoretical perspective, and available data.

Yilmaz (2013) said that quantitative research involves deductive approaches to confirm or reject existing theories or hypotheses. In quantitative research, researchers collect data from the sample based on predefined questionnaires. They then map collected data into numerical form so that they can use statistical calculations to analyze and draw inferences about tested hypotheses (Yilmaz, 2013). Quantitative research allows researchers to study a sample of the population and later generalize findings into entire population (Yilmaz, 2013). At the same time, predefined and structured quantitative instruments may not allow participants to express their feelings, opinions, and experiences about the phenomenon. According to Yilmaz (2013), quantitative research involves focusing on analyzing relationships between variables so that researchers may predict, confirm, or reject hypotheses. Quantitative research is

appropriate for formulating relationships between variables or understanding the impact of independent variables on dependent variables. In this study, I did not test, confirm, or reject existing hypotheses. It was also not intended to establish relationships among variables. This study focuses on exploring company practices of collection and use of consumer data without an individual's consent. I did not use the quantitative approach to conduct this study.

A publication by Mack, Woodsong, MacQueen, Guest, and Namey (2005) showed that researchers often use qualitative studies to explore and understand social phenomena. Mack et al. (2005) stated that qualitative studies do not involve statistical numbers to test theories or hypotheses. Instead, researchers observe the phenomenon that they are interested in, collect evidence, and explain the event in terms of their existing experiences, points of view, and assumptions. Researchers analyze qualitative data to develop theories using an inductive approach. Birks, Fernandez, Levina, and Nasirin (2017) noted that qualitative methods are often used to build grounded theories to explain events that may not be studied. While researchers use statistics to prove existing theories in quantitative research, they use observations, interviews, or documents to explore a phenomenon in qualitative studies (Birks et al., 2017).

According to Myers (2002), qualitative research can be divided into five models: narrative, phenomenological, grounded theory, ethnographic, and case study models. With the qualitative approach, researchers use observations, interviews, or documents to collect qualitative data. Myers (2002) highlighted that qualitative studies in IS fields focus on understanding and explaining problems involving management, computer

technology, human and machine, big data, management, machine learning, or artificial intelligence.

Harindran and Chandra (2017) noted that archival research is common in qualitative studies. In archival studies, researchers collect data from historical documents, facts, and evidence to explore a particular event (Harindran & Chandra, 2017; McIntush et al., 2019). Fischer and Parmentier (2010) mentioned that researchers increased their use of archival data as primary data to contribute to understanding in consumer studies. Archival data contain a wide range of empirical materials, including web sites, company annual reports, financial reports, and organizational documents. While interview and observation approaches may be considered main data collection methods in qualitative studies, archival data can provide researchers with rich information and valid evidence to support research findings (Tesar, 2015).

Historical documents can provide valuable information related to individuals, organizations, events, and social concerns. Heng, Wagner, Barnes, and Guarana (2018) showed that researchers could use archival methods to collect historical documents systematically, analyzing earlier events to understand their impact on modern society. Mills and Mills (2018) mentioned that researchers used archival documents not only to confirm the past events but also to contribute knowledge to the existing literature. In researching business practices, archival methods can provide insights into how companies operate and make business decisions over period of time (Mills & Mills, 2018).

### **Role of the Researcher**

Gaillet (2012) discussed the increase of using archival research methods in many study disciplines as well as the role of the researcher holding in using archives to explore new phenomena. Archival researchers can reinvestigate existing theories, answer new research questions, and develop new knowledge based on existing documents (Gaillet, 2012). I chose qualitative archival research methods to explore private company practices of collection and use of consumer data without an individual's consent. Archival documents were primary data that were collected to answer the research question as well as support findings. I served as the data collection instrument who collected and analyzed data from both industry and U.S. government agency web sites. I focused on selecting appropriate strategies so that I could carry out this study scientifically and present findings without bias.

Research ethics and integrity are important for producing trustworthy findings. Researchers need to consider research concerns such as ethics, bias, and integrity at each step of the study, from choosing a suitable research design and data collection method to drawing inferences and presenting findings (Wood, 2011). Although the data collection procedures in this study did not involve individuals or animals, I still needed to manage collected data based on Walden University research ethical guidelines, especially collected documents associated with private company operations. I did not have any personal or professional relationships with organizations where I collected data. This helped eliminate conflict of interest concerns between researcher and research entities. I searched, categorized, and interpreted archival documents truthfully to ensure their



credibility, authenticity, and representativeness. Gaillet (2012) mentioned that one of the archival researcher's roles is to spend time gathering archives and synthesizing them so that they triangulate evidence and present findings truthfully. I was responsible to verify the author, original purposes, and context of collected documents. I used triangulation techniques to reduce errors and mitigate bias existing in archival documents. Welch (2014) mentioned various triangulation techniques that can be used in the archival research method. Researchers can collect data from multiple sources, check data from archival documents, and validate conceptual frameworks from different theorists (Welch, 2014).

Besides collecting data from credible sources, I needed to select an appropriate data analytic process to categorize, code, and discover common themes within collected archives. Compiling and coding documents are critical roles of archival researchers (Gaillet, 2012). Terry, Hayfield, Clarke, and Braun (2017) noted that thematic analysis is a well-known research method for analyzing and presenting qualitative data. It provides researchers the capability to develop code lists, categorize patterns, and identify themes emerging from data (King & Brooks, 2018). I chose thematic analysis method to code, interpret, and present collected data. I used the archival document selection protocol (Appendix A) to collect archives from the U.S. government and private company websites. I present the archival material selection logic and data collection procedure in the methodology section. I documented data collection procedures and presented coding development processes. The aim is to increase the transparency of research methods and

data analytic processes, so that other researchers are able to duplicate and reproduce the results.

### **Methodology**

I used archival research methods to conduct this study. Archival research methods systematically provide instructions for knowledge inquiry that researchers can use to study historical materials about an individual, organization, or event (Barnes, Dang, Leavitt, Guarana, & Uhlmann, 2015; Ventresca & Mohr, 2017). Besides using archival research methods to study past phenomena, Ventresca and Mohr (2017) noted that researchers could use archival research methods to investigate contemporary events so that they can gain a better understanding of present. With emerging technologies such as Internet or digital data management systems, organizations can create and archive large amounts of business documents in their computer systems. These archival materials can provide insights into organizational business practices (Ventresca & Mohr, 2017). Archival research may allow researchers to develop necessary evidence and establish fundamental arguments about organizational practices over time. In this methodology section, I present several components of archival research methods, including participant selection logic, instrumentation, data collection procedures, a data analysis plan, and issues of trustworthiness.

#### **Document Selection Logic**

Today, many organizations in both public and private sectors use computer systems to create and archive business documents. Innovative technologies, including Internet, social networks, blog websites, online media streaming, or digital archival

libraries, provide large amounts of data in various formats. Lucas (2018) noted that archival materials could be documents, images, videos, texts, or voices. New technologies also allow online users to search, filter, and retrieve digital materials through various Internet-connected devices, such as computers or smartphones. Historical digital materials which are public and accessible on the Internet allow researchers to seek various pieces of information and to gain insights into organizations or social phenomena in which they are interested (Corti, 2004; Stan, 2010).

Archival research methods rely on archival materials as main sources to support the study. Historical materials may provide information that researchers can collect to understand past events as well as how it influences and forms the present (Stan, 2010). Corti, (2004) noted that archival material could be found at various sources, which may contain official and original documents. I used the archival document selection protocol (Appendix A) to gather archival documents from multiple sources at both U.S. government agencies and private companies. U.S. government agencies and private companies may document and publish large amounts of official records, reports, policies, or letters on the Internet. Many U.S. federal libraries provide a wide range of legislations, policies, hearing transcripts, hearing videos, committee activities, and other archival documents. Table 2 includes a list of data sources from websites of U.S. government agencies, U.S. government libraries, and nonprofit organizations where I collected archival documents to explore current data protection and privacy legislations.

Table 2

*Archival Library*

Archival Library	Website Uniform Resource Locator (URL)
The United States House of Representatives hearing transcripts	<a href="https://www.house.gov/">https://www.house.gov/</a>
The United States Senate hearing transcripts	<a href="https://www.senate.gov">https://www.senate.gov</a>
The U.S. Federal Legislative Information	<a href="https://www.congress.gov">https://www.congress.gov</a>
The U.S. Federal Trade Commission testimonies	<a href="https://www.ftc.gov/">https://www.ftc.gov/</a>
The U.S. National Archives	<a href="https://www.archives.gov">https://www.archives.gov</a>
Library of Congress	<a href="https://www.loc.gov/">https://www.loc.gov/</a>
Videos and transcripts of business executive officer interviews	<a href="https://www.youtube.com">https://www.youtube.com</a>
Catalog of U.S. Government Publications (CGP)	<a href="https://catalog.gpo.gov">https://catalog.gpo.gov</a>
GovInfo	<a href="https://www.govinfo.gov/">https://www.govinfo.gov/</a>
The U.S. Government Publishing Office (GPO)	<a href="https://www.gpo.gov">https://www.gpo.gov</a>
DocumentCloud	<a href="https://www.documentcloud.org/home">https://www.documentcloud.org/home</a>
American Cable Television Industry	<a href="https://www.c-span.org/">https://www.c-span.org/</a>
The World Privacy Forum	<a href="https://www.worldprivacyforum.org">https://www.worldprivacyforum.org</a>

Private companies often publish many business documents, financial reports, policies, or statements on their websites. Table 3 presents a list of private companies, where consumer data are used for many business purposes. Documents from these websites provide rich information that I could explore their business practices of collecting and using consumer data.

Table 3

*Data Provider Company*

Company Name	Website URL
Acxiom	<a href="https://www.acxiom.com">https://www.acxiom.com</a>
Corelogic	<a href="https://www.corelogic.com">https://www.corelogic.com</a>
Epsilon	<a href="https://us.epsilon.com">https://us.epsilon.com</a>
Equifax	<a href="https://www.equifax.com/business/">https://www.equifax.com/business/</a>
Experian	<a href="https://www.experian.com">https://www.experian.com</a>
Facebook	<a href="https://www.facebook.com">https://www.facebook.com</a>
GfK	<a href="https://www.gfk.com/en-us/">https://www.gfk.com/en-us/</a>
Google	<a href="https://www.google.com">https://www.google.com</a>
i360	<a href="https://www.i-360.com/">https://www.i-360.com/</a>
Mastercard	<a href="https://www.mastercardservices.com">https://www.mastercardservices.com</a>
Oracle Data Cloud Service	<a href="https://www.oracle.com/cloud/data-hotline/">https://www.oracle.com/cloud/data-hotline/</a>
Salesforce	<a href="https://www.salesforce.com">https://www.salesforce.com</a>

**Instrumentation**

In qualitative research, there are several approaches that researchers can use to collect data to address research questions. Depending on the research design, researchers can use interview methods to gain understandings of a participant's experiences or viewpoints. Researchers can observe a participant's activities and take notes in the study field. Existing documents, archives, and artifacts are also primary sources for researchers to explore contemporary phenomena (Merriam & Tisdell, 2015). Archives are often generated broadly for many purposes, such as business documents, contracts, financial records, business annual reports, public records, and so on. Especially in the digital age, organizations can produce large amounts of documents and make them accessible over the Internet. Barnes et al. (2015) argued that archival documents might provide rich information about the phenomenon of interest. They offer researchers an opportunity to discover insights in existing archives and gain a better understanding of the research

problem (Merriam & Tisdell, 2015). Since archives are produced for purposes other than that of the current study, researchers need to develop a strategy to select relevant documents to solve research problems. Researchers also need to connect evidence from the archives to their research questions as well as to interpret archival contents to support research findings.

Welch (2014) mentioned that private companies often controlled and managed business operational documents as proprietary materials. They may keep these business documents confidential and allow for internal use only. Business operational documents might not be accessible to individuals from outside the company (Welch, 2014). As the non-disclosure agreement is a common practice to protect proprietary business information in the industry, it is not feasible to gather internal information from private businesses or data-provider companies. Instead, I gathered existing documents from multiple sources to address the research problem. The data sources are presented in the previous section. I used the archival document selection protocol (Appendix A) to collect data from U.S. government agencies, private companies, and non-profit organizations.

There is no specific technique to research archival materials and to select appropriate documents for a study (Corti, 2004). Researchers can start reviewing and analyzing archival materials as soon as they locate and retrieve relevant information. With the help of emerging technologies, such as the Internet, big data, or cloud computing, organizations can generate and store large amounts of business documents in their information management systems (Gantz et al., 2019). Both industry and U.S.

government agencies might release and make their business documents accessible from outside via their Internet websites.

Williamson and Johanson (2013) discussed five archival evaluation techniques that researchers can use to explore archival materials. These evaluation techniques are detective, verifier, attributor, clarifier, and storyteller techniques (Williamson & Johanson, 2013). The detective technique involves exploring the context of archival documents and establish links among them. The next step is to verify the creditability of documents to ensure they were produced by reliable and credible sources. The attributor technique helps researchers to discover the true meaning of archival documents. Researchers need to find the original purpose and target population of archives. The clarifier technique provides researchers a tool for internal checking. It helps researchers find the use of archival documents by others as well as cross check with outside sources. The last step in the evaluation process is to evaluate original hypotheses, which were examined based on archival documents. The storyteller technique describes the accuracy and authenticity of archival documents by synthesizing and connecting evidence from the chain of events.

### **Procedures for Data Collection**

In archival research methods, researchers collect existing documents from archival repositories or libraries. Procedures for searching, locating, and selecting archival documents are critical steps in archival studies. Lucas (2018) noted that archival researchers need to focus on evaluating existing documents to ensure that they are reliable and relevant to the research problem and question. Regardless of the data

collection approaches through which researchers choose to gather data, researchers might establish a procedure for data collection based on research questions and methodology (Lucas, 2018). The principle of collecting and using existing documents is the same as other approaches, such as interview, or observation. In archival research methods, the data collection procedure is a systematic process of locating documents, selecting relevant data, discovering new insights, and tracking down the information

The first step of data collection procedure is to locate documents related to the research question. The data sources are presented in the document selection logic section. The research question of this study is how do private company practices of collection and use of consumer data without an individual's consent align with the ownership, transparency, ethics, and consumer privacy laws. In this study, I used the archival document selection protocol (Appendix A) to collect archives from the U.S. government and private company websites. Archival documents from each website offer different information, which describe business operations as well as organizational policies and activities. I used different keywords to search relevant documents from data sources, presented in Table 2 and Table 3. The search keywords included: *data broker, data provider, data-driven business, collect consumer data, consumer data in online marketing, privacy concerns, consumer privacy, privacy violation, online privacy, consumer data, personal data, privacy information, data privacy, big data, ethics of big data, data ethics, privacy law, privacy protection, regulation control, and data breach*. I considered any possibility that can lead to relevant documents. I managed archival materials based on numbers of metadata, which provide detailed information related to



particular documents. Table 4 provides a list of metadata, which is used to select and manage collected documents.

Table 4

*Archival Material Metadata*

Archival material metadata	Description
Title	The title of archival document
Author	The archival authors
Date	Date created or released
Location	The URL links to documents
Type	Describes document type (e.g., text, video, voice, or image)
Original organization	Organization that documents belong to
Note	Notes on or descriptions of documents
Search keyword	Contains keyword, which is used to search documents.
Content	The full text of the document.

After locating documents to support the study, researchers need to evaluate and understand archives to ensure that they are coming from reputable and credible sources (Merriam & Tisdell, 2015). I used the evaluation criteria developed by Wood (2011) to select and determine which documents should be used in this study. Wood (2011) presented five archival evaluated criteria, which researchers can use to collect and manage archival documents. Archival evaluated criteria include provenance, purpose, context, veracity, and usefulness (Wood, 2011). Table 5 includes the purpose of each criterion.

Table 5

*Reliability Evaluated Criteria*

Evaluation criteria	Purposes
Provenance	This is used to determine the original information of archives. The author, original reasons, and document preservation need to be identified to prove the authenticity of documents.
Purpose	The purpose of document needs to be defined. This information provides objectives of the document that the author wants to send to target audiences.
Context	This information provides context for the documents' creation, such as original professional purpose, the location where document was created, or the time created.
Veracity	This criterion is used to determine the creditability of documents. Documents need to be written by the credited author and based on reliable evidence.
Usefulness	This criterion is used to ensure that the selected documents are valuable and useful for current research. They provide evidence and information that researchers could not find from other resources.

*Note.* Adapted from “Understanding and Evaluating Historical Sources in Nursing History Research,” by J. P. Wood, 2011, *Nursing Praxis in New Zealand*, 27, p. 28.

As mentioned, there are many approaches that researchers can choose to collect relevant documents for the study. Using research questions and the research method as a foundation, researchers can explore massive archives that are available over the Internet environment. Each document can potentially provide useful evidence that, when pieced together, may lead to the answer to the research question. Analyzing and interpreting credible documents may lead to a better understanding of the phenomenon of interest as well as present research findings truthfully (Francis & Taylor, 2013). The next step of the archival research method is data analysis. The data analysis process allows researchers to systematically analyze collected data, discover common themes, and interpret findings.

## **Data Analysis Plan**

After collecting data to support the study, researchers need to process and analyze collected data, so that they may discover insights, draw inferences, and answer the research question. Depending on the selected research design and method, researchers choose an appropriate data analysis process to make sense of collected data (Willig, 2017). Different theorists have different assumptions, point of views, and approaches to interpretation. There is no specific standard guideline instructing researchers on how to analyze and code the contents of qualitative data (Riemer, Quartaroly, & Lapan, 2011). Researchers compile, code, and interpret the contents of archives based on their best judgment, assumptions, and experiences (Wood, 2011). Through a systematic data analysis process, researchers may discover common patterns from collected data and present findings in another scholar report. In other words, an appropriate data analysis process allows researchers to transform unstructured data into trustworthy evidence that explain the phenomenon of interest (Williamson & Johanson, 2013).

Williamson and Johanson (2013) showed that there are many data analysis approaches that researchers can choose to organize and analyze qualitative data. They include thematic analysis, content analysis, and discourse analysis, as well as many other types of analysis methods. The focus is to analyze collected data systematically and extract key facts from data, so that researchers represent the phenomenon of interest truthfully based on the evidence of the data. Williamson and Johanson (2013) mentioned that researchers also hold essential roles in the data analysis process. They interpret the differences and similarities of patterns emerging from collected data based on their points

of view and experience. All factors involved in the data analysis process need to be considered in order to ensure the trustworthiness of study results.

In this study, I used thematic analysis to organize, categorize, and code collected archival documents. King and Brooks (2018) defined thematic analysis as a method for identifying, categorizing, discovering patterns, and interpreting themes within data.

Thematic analysis is a well-known method for analyzing qualitative data. It provides researchers the capability to develop code lists, categorize patterns, and identify themes emerging from data (King & Brooks, 2018). Terry et al. (2017) mentioned that thematic analysis might be used to analyze many types of qualitative data, including interviews, surveys, documents, texts from online discussion, and other qualitative sources.

Researchers can use thematic analysis to analyze different sizes of qualitative data as well as apply it to answer a wide range of research topics (King & Brooks, 2018).

There is no single approach for conducting a thematic analysis of qualitative data (Terry et al., 2017). Instead, there is a wide range of methods that researchers can select to explore and interpret collected data. The thematic analysis focuses on identifying and interpreting common themes that carry essential information relevant to the research problem. These key features help target audiences better understand the phenomenon of interest. The thematic analysis also refers to the terms of coding, category, and theme (King & Brooks, 2018).

Code, category, and theme are essential elements of thematic analysis. Elliott (2018) defined that code is a unit of data that summaries the essential content of a document. A code may represent a word, one sentence, or a single paragraph of text

within a particular document. It identifies and captures ideas, concepts, or meanings within data contents (Elliott, 2018). The research problem, research question, collected data, and the researcher hold critical roles in driving the coding process (King & Brooks, 2018). Coding is an iterative and recursive process, which might repeatedly progress from one cycle to the next until common patterns emerge from collections of codes. Williamson and Johanson (2013) discussed themes as the concept of common patterns emerging from code lists. Themes hold common meanings and general insights among data items. They summarize the topic of a common code list. Researchers develop themes to present essential information for explaining research problems and answering research questions.

This study adapted the Terry et al. (2017) thematic analysis guideline to analyze collected documents. This thematic analysis template contains six phases, including familiarization, coding, theme development, theme reviewing, theme defining, and report. Researchers have the flexibility to move back and forth among phases within this process to develop a code list and identify potential themes (Terry et al., 2017). These analysis phases provide a foundation for researchers to develop a data analysis plan that can help them achieve their research objectives (King & Brooks, 2018).

Familiarization is the first step in the data analysis process. It provides a foundation for researchers to get familiar with collected data. The existing digital archives might have a large volume of materials stored in various formats, such as documents, images, transcripts, or websites. Terry et al. (2017) mentioned that researchers might be overwhelmed by the mass of unstructured data of existing archives.

Familiarization provides researchers an opportunity for engaging and exploring archival materials (Terry et al., 2017). Researchers can start reading documents or watching media files so that they can gain insights as well as recognize patterns within data. During this step, I read over the entire set of collected archival materials to get general ideas and to start asking questions related to the business practices of collection and use of consumer information. I took notes while reviewing collected documents for referent purposes. These notes gave general ideas and highlights that I needed to focus on during the coding process.

The next step in the data analysis process is coding. Coding provides researchers the capability to capture meaningful information from single sentences or paragraphs and to highlight them with meaning labels. During the coding process, I reviewed collected documents and started identifying information related to the research question. Terry et al. (2017) noted that the coding process in thematic analysis is open and flexible. There is no standard that requires researchers to code collected data according to a particular approach. I opened to any lead that may provide insights to gain better understandings of the research problem. I focused on building a code list, which may provide critical concepts and summarize contents of collected archives without referring to the actual documents. The result of this phase was a comprehensive code list that may present meanings and patterns within the collected documents.

Theme development is the next phase, where researchers can begin categorizing codes, identifying common patterns, and developing themes. Researchers might gain an in-depth understanding of collected data after developing a code list that captures the

contents of the data (Terry et al., 2017). The theme development phase allows researchers to review collections of codes and categorize them into groups, which present general themes. Archival documents might have a large volume and rich information that might distract researchers during the data analysis process. Terry et al. (2017) mentioned that researchers could use the research question as a guideline to recognize potential patterns as well as to construct themes. During the theme development phase, I focused on identifying codes that had common meanings across the data segment. The effort was to cluster common ideas, concepts, and codes into groups so that I could develop potential themes. Theme development is an iterative process that provides an ability to review, explore, and reconstruct themes repeatedly (Terry et al., 2017).

Although the theme development phase produces a collection of potential themes existing in data, researchers need to review and verify them against collected data and the research question. During the theme reviewing phase, Terry et al. (2017) noted that researchers need to verify potential themes to ensure they represent the meaning of coded data. Each theme holds a specific meaning that represents a particular data segment. The theme also has a key feature related to the research problem. During this phase, I reviewed the collection of candidate themes and verified them with coded data to ensure that they represented coded data and related to the research problem.

After checking potential themes to ensure they capture key information from collected data, I began interpreting and analyzing by writing extracted data from collections of themes. Terry et al. (2017) noted that the theme needs to be described by a short summary. A definition of each theme might help clarify the scope, meaning, and

content of the theme. Each theme might provide rich information to researchers for interpreting and writing about collected data. In the theme defining phase, I interpreted common patterns of coded data as well as explore the meanings of the collections of themes.

The final step of the thematic data analysis is to write an analytic report. Terry et al. (2017) represented two approaches to develop a report based on evidence extracted from collected data. Two styles of reporting were illustrative and analytic. During the illustrative approach, researchers use insights from collected data as evidence to support findings. During the analytic style, researchers might analyze essential features from data and discuss how these data elements construct the foundation of claims. After getting familiar with the data, developing a code list, and identifying themes, I was able to write a final analysis report based on what I find in the data analysis process. I used NVivo software to store, manage, and organize archival materials that I collected to support this study. NVivo software provided me a comprehensive tool to conduct coding as well as analyze collected data.

### **Issues of Trustworthiness**

#### **Credibility**

During the archival research method, researchers can use different ways to evaluate the credibility of archival materials. Myers (2002) said that archival materials need to be credible, meaningful, and authentic. One of the main steps in the archival research method is that researchers need to evaluate the data source to ensure its credibility (Corti, 2004). To increase the credibility of this study, I collected documents



from credible sources. Collected documents may represent original evidence to support findings. I verified the background information of collected documents when searching and collecting them through archival libraries. For instance, I collected and analyzed archival materials, collecting from public websites of U.S. government agencies and private companies. These sources provide official documents, which describe activities and business operations of both federal agencies and private companies. The triangulation technique is an essential technique to establish the validity of the study (Mills & Mills, 2018). Gathering data from different sources might reduce errors existing with archival materials, such as missing information and conflicting arguments (Mills & Mills, 2018). I collected data from multiple sources to ensure the accuracy and authenticity of study findings.

Existing documents, newspapers, records, memos, emails, reports, policies, and other materials are primary data in the archival research method. Researchers might have assumptions about bias factor in archival documents since archives might be documented for purposed other than current studies. Archives could carry opinions or interpretations of other researchers. For that reason, interpreting and drawing inferences based on archival collections may have bias since it contains different viewpoints. However, archival documents provide researchers with a complete picture and multiple views about an event. The advantage of archival documents is that it allows researchers to check information from different sources and observe the phenomenon of interest from multiple views. This helps archival researchers present findings truthfully based on different types of evidence compiling from multiple sources.

**Transferability**

Transferability criteria refer to the degree to which research findings can be transferred or generalized to other contexts (Myers, 2002). The result of this study might not be generalizable across industries, but it can be transferred to other businesses with the same study contexts and settings. For instance, the findings might be transferred from businesses, where they collect personal information directly from consumers to businesses where they aggregate consumer data from third-party companies. Anney (2014) mentioned that researchers might have the responsibility to validate the transferability of the findings. Nevertheless, I used the data saturation approach to ensure that key information was captured adequately to support findings.

**Dependability**

The archival research method requires researchers search, select, compile, and interpret existing documents based on their own judgment (Fischer & Parmentier, 2010). Digital archives exist in various types and forms, including documents, transcripts, images, videos, and so on. In archival research, researchers do not focus on selecting research participants or choosing a sample size. Instead they are concerned about how to locate, search, and select relevant documents that already exist (Wood, 2011). They aim to compile, synthesize, and interpret existing archives (Hill, 2011). I used the thematic analysis method to analyze collected documents. In order to increase the dependability of this study, the selected research design, including research instrument, data collection procedure, and data analysis process is described in detail and presented in a systematic manner. The focus is to ensure that other researchers are able to reproduce research

results by repeating the same knowledge inquiry method. I described and presented in detail data source locations, where existing archives were stored and accessible for inquiry. I documented both data collection procedure and data evaluation criteria, which were used to select relevant archives. I used a well-defined data analysis process, developed by Terry et al. (2017). Each step in the process was presented to help researchers develop code lists and identify common themes systematically.

### **Confirmability**

I gathered archival documents from multiple sources from both public and private organizations. Mills and Mills (2018) mentioned that data coming from different sources help triangulate information and provide a holistic understanding of the research problem. Elder, Pavalko, and Clipp (1993) pointed out that multiple data sources can bring different perspectives and viewpoints related to the research topic. It allows mitigating potential biases of a single archival repository. To increase the confirmability of the findings, I collected archival documents in an open-minded way and follow any leads that guide me to credible documents. I used the thematic analysis method to analyze and map collected documents into collections of codes and common themes. I used a reflexive journal to record any issues that might potentially affect the data analysis process. I took notes on coding development process, reasons behind coding categorization, and justifications of theme changing. These notes allow me to track and validate data analysis processes, ensuring that potential biases are mitigated.

## **Ethical Procedures**

Ethical procedures are an important aspect of any scholarly study (McKee & Porter, 2012). Most academic institutions require researchers to obtain approval from the Institutional Review Board (IRB) when the study involves accessing humans or animals. I received permission and approval from the Walden University IRB (# 05-19-20-0105615). Myers (2002) mentioned that ethical procedures not only protect study participants, but also protect researchers from ethical issues when handling participant's data. Ethical procedures also help in maintaining the integrity of the study in terms of its research design, data collection, and data analysis. Myers (2002) noted four ethical research principles: truthfulness, thoroughness, objectivity, and relevance. Researchers need to conduct research with positive purposes and study it thoroughly. They also need to report findings truthfully based on collected data without bias (Myers, 2002).

In archival research, ethical procedures might include issues related to protection of copyright, ownership, and confidentiality of collected data (Young & Brooker, 2006). Especially in the digital environment, it is easy for researchers to search and access massive amounts of information, which are published and distributed over Internet websites. Both public and private organizations release documents related to their operational activities and products, including terms of use, privacy policy statements, and product solutions. Innovative technologies, such as the Internet, mobile devices, and digital data software, allow researchers easily to search, locate, and store these digital materials for researching purposes. This capability leads researchers to easily ignore the ownership, privacy, or intellectual property of digital documents on web sites (Merriam

& Tisdell, 2015). Especially on social network web sites or online video sharing platforms, individuals might exchange personal information or videos that may lead researchers to disclose private data of online users.

McKee and Porter (2012) mentioned that protecting the confidentiality and intellectual property of documents is critical in archival research. I managed and stored collected documents in two private locations, including my personal laptop and private Microsoft OneDrive. This Microsoft OneDrive is digital online storage, which was created by Walden University. It allows students to access Walden University email as well as OneDrive digital storage. All collected documents were used to support this study. I did not share any documents with other entities or use collected documents outside the scope of this study. Although I collected documents from the public domain, I obtained permission to use any documents which required authorization from authors.

### **Summary**

This chapter includes an overview of the qualitative archival research design and research methodology. Discussion of the rationale for selecting the archival research method highlights motivations and reasons why the archival method was chosen over other traditional qualitative approaches. I also discuss the role of the researcher in this study. I explain the logic of document selection, research instrumentation, and data analysis process. I discuss six phases of the thematic data analysis process that provide a foundation for coding, analyzing, and making sense of existing archival documents. Chapter 4 includes study results and findings extracted from collected archives. I address

the research question as well as discuss the trustworthiness of evidence based on collected data.

## Chapter 4: Results

The purpose of this qualitative archival research is to explore private company practices of collecting and using consumer data without an individual's consent.

Consumer data have become critical assets that allow private companies to improve their products and customer experiences. Private companies also use new technologies such as big data, data mining, and computer artificial intelligence to produce data products that they sell to other companies. I conducted this qualitative archival research to gain an understanding of these business practices in the data-driven economy. I used archival documents collected from multiple sources to answer the following research question: How do private company practices of collection and use of consumer data without an individual's consent align with ownership, transparency, ethics, and consumer privacy laws? There are large volumes of digital archives that are available on the Internet and accessible for archival research. I collected archival documents from both public and private sectors, and used a thematic data analysis method to code and extract supporting evidence.

This chapter includes a description of the research setting, data analysis, evidence of trustworthiness, and research findings. In the section on the research setting, I discuss challenges that might affect searching for and finding archives on the Internet. Characteristics of archival documents such as locations, types of documents, number of collected documents, and document formats are presented in the demographic and data collection sections. Next, I discuss the data analysis method that was used to code collected documents, categorize patterns that emerged, and develop common themes.

Finally, I present research findings based on themes as well as supporting evidence extracted from collected data.

### **Research Setting**

The ability to access archival materials was a primary consideration during the document selection process. I used various predefined search keywords and was open to any possibility that might lead to relevant materials. Not all archives were accessible. For instance, I was not always able to retrieve all documents referenced within a particular document. Beyond availability and accessibility of archives, there were additional considerations during my search for documents. Selected documents needed to be credible, authentic, representative, and meaningful. Findings from selected documents needed to be trustworthy so that they could be used as evidence to answer the research question. I had to verify authors, original purposes, and content of selected documents to ensure they were valid and reliable. Another challenge in terms of studying archival documents involved reviewing and analyzing a large volume of available archives.

### **Demographics**

I collected archival materials from nine public websites from the U.S. Congress as well as various U.S. government agencies and one digital record repository of U.S. nonprofit organizations. Table 2 in the section on the logic of document selection in Chapter 3 shows the list of these public websites. The U.S. House of Representatives and Senate released records of hearings, testimonies, legislations, and other legal records on their web sites. These documents were available to the public to access and review. Other U.S. federal agencies, including the FTC, Catalog of U.S. Government Publications, and



U.S. National Archives also published various data privacy policies, guidelines, and documents regarding best practices that regulate the use of consumer data within private companies. Besides archival documents from the U.S. government, I also collected videos of business executive interviews and data protection expert panel discussions from two online streaming media websites.

The business operations of data provider companies that sell information might be less known to consumers. They hold a critical role in the data-driven economy by collecting, buying, analyzing, and selling consumer data to other companies to use for business purposes. In addition to documents collected from U.S. government legislation, I collected documents from 12 data provider companies. Table 3 in the section on the logic of document selection in Chapter 3 shows the list of these companies. These companies have their main headquarters or regional headquarters in the United States. They offer data solutions to other companies operating in many industries such as the automotive, insurance, baking, marketing, and healthcare industries.

### **Data Collection**

Although I collected archival documents that are available and accessible from public web sites, I ensured that the data collection procedures complied with Walden University's ethical standards. I received permission and approval from the Walden University IRB. The IRB approval number is 05-19-20-0105615. Archival materials were the primary data sources in this study. I used the archival document selection protocol (Appendix A) as a guideline for collecting archival materials. Table 6 shows the number of documents collected from archival libraries. Documents collected from these web sites

included transcripts from 75 U.S. House of Representatives and Senate hearings, 15 data privacy protection panel discussions, and other federal agency reports. Collected documents also included transcripts of interviews with data privacy experts, court records, and witness statements.

Table 6

*Number of Documents Collected from Archival Libraries*

Archival Library	Website Uniform Resource Locator (URL)	Number of Documents
The United States House of Representatives hearing transcripts	<a href="https://www.house.gov/">https://www.house.gov/</a>	42
The United States Senate hearing transcripts	<a href="https://www.senate.gov">https://www.senate.gov</a>	12
The U.S. Federal Legislative Information	<a href="https://www.congress.gov">https://www.congress.gov</a>	8
The U.S. Federal Trade Commission testimonies	<a href="https://www.ftc.gov/">https://www.ftc.gov/</a>	13
The U.S. National Archives Library of Congress	<a href="https://www.archives.gov">https://www.archives.gov</a>	4
Videos and transcripts of subject matter expert interviews	<a href="https://www.loc.gov/">https://www.loc.gov/</a>	4
Catalog of U.S. Government Publications (CGP)	<a href="https://www.youtube.com">https://www.youtube.com</a>	6
GovInfo	<a href="https://catalog.gpo.gov">https://catalog.gpo.gov</a>	1
The U.S. Government Publishing Office (GPO)	<a href="https://www.govinfo.gov/">https://www.govinfo.gov/</a>	20
DocumentCloud	<a href="https://www.gpo.gov">https://www.gpo.gov</a>	1
American Cable Television Industry	<a href="https://www.documentcloud.org/home">https://www.documentcloud.org/home</a>	3
The World Privacy Forum	<a href="https://www.c-span.org/">https://www.c-span.org/</a>	3
	<a href="https://www.worldprivacyforum.org">https://www.worldprivacyforum.org</a>	1

The U.S. government has been digitizing historical Congressional hearings and publishing on many websites of U.S. government organizations, such as the Discover U.S. Government Information, the U.S. Government Publishing Office, the United States

House of Representatives, or the United States Senate. Transcripts of Congressional hearings, testimonies, statements, and reports are in portable document format (PDF) files. A congressional hearing is a formal meeting where House, Senate, or special committees conduct investigations, gather information, and explore current issues or events. I collected these archival documents using search features provided by U.S. government agency websites. I used predefined search keywords to retrieve documents that were relevant to the research problem and question. I also documented metadata of each selected document for the record. The archived document selection protocol (Appendix: A) describes in detail steps involved with searching for and selecting archival documents.

Table 7 shows the total number of documents collected from each of 12 data provider companies. Documents from these companies included privacy statements, privacy policies, terms of use agreements, reports on product solutions, data catalogs, data directories, and business annual reports. These data provider companies provided data solutions for businesses in a wide range of industries, such as financial services marketing, retail, healthcare, mortgage, telecommunication, insurance, entertainment, restaurants, education, business services, and travel.

Table 7

*Number of Documents Collected from Each Data Provider Company*

Company Name	Website URL	Number of Documents
Acxiom	<a href="https://www.acxiom.com">https://www.acxiom.com</a>	7
Corelogic	<a href="https://www.corelogic.com">https://www.corelogic.com</a>	2
Epsilon	<a href="https://us.epsilon.com">https://us.epsilon.com</a>	7
Equifax	<a href="https://www.equifax.com">https://www.equifax.com</a>	8
Experian	<a href="https://www.experian.com">https://www.experian.com</a>	23
Facebook	<a href="https://www.facebook.com">https://www.facebook.com</a>	5
GfK	<a href="https://www.gfk.com/en-us/">https://www.gfk.com/en-us/</a>	2
Google	<a href="https://www.google.com">https://www.google.com</a>	9
i360	<a href="https://www.i-360.com/">https://www.i-360.com/</a>	7
Mastercard	<a href="https://www.mastercardservices.com">https://www.mastercardservices.com</a>	10
Oracle Data Cloud Service	<a href="https://www.oracle.com/cloud/data-hotline/">https://www.oracle.com/cloud/data-hotline/</a>	9
Salesforce	<a href="https://www.salesforce.com">https://www.salesforce.com</a>	7

### **Data Analysis**

The thematic analysis method is a scientific approach to explore emerging patterns and themes from qualitative data (Braun & Clarke, 2006). It allows researchers to code words, phrases, or sentences in collected documents that might contain rich information related to the research problem. As Braun and Clarke (2006) discussed, the thematic data analysis method can be considered a foundational approach to analyzing qualitative data. Although the thematic analysis method provides a general way to conduct analysis in a qualitative study, Braun and Clarke (2006) claimed that thematic data analysis is independent of theory and flexible enough for conducting analysis broadly in qualitative studies. I used the Terry et al. (2017) thematic analysis guidelines to analyze collected documents. These guidelines have six steps, which include

familiarization, coding, theme development, theme reviewing, theme defining, and reporting. I describe in detail these six steps in the data analysis plan section.

In this study, I used Nvivo software to code, categorize, and develop common themes from archival documents. First, I imported archival materials into Nvivo software and organized the materials in different groups, including hearings, reports, privacy statements, data catalogs, panel discussions, and interviews. This allowed me to categorize the collected archival materials and find relevant documents to answer the research question. Next, I reviewed all collected documents to familiarize myself with and explore the content of each document. Due to the large volumes of archival documents, this step provided opportunities to understand the content and remove any unrelated materials from the collected data set (Terry et al., 2017). The next step was manually coding each document using the Patterson and Davis (2012) conceptual framework as a foundation as well as focusing on the research problem and question. The result of this step was a code list representing relevant data extracted from the collected documents. I then reviewed and categorized the codes into different groups. The next step was to develop candidate themes based on frequent patterns unveiled from the categorized codes. I then reviewed the candidate themes and verified them with the coded data to ensure that they represented the meaning of the collected data.

These steps of the thematic data analysis process were repeated so that I was able to revise the codes, categories, and themes. After multiple revisions of the candidate themes, I developed short descriptions for each theme to clarify the scope, meaning, and content of the theme. Following are common themes that emerged from the collected

documents, and summaries of each theme. A complete list of findings and supporting evidence are discussed in the study results section.

**Theme 1: Benefits of using consumer data.** Using consumer data could provide benefits to both consumers and businesses in the data-driven economy. Consumers could get a wide range of products and services at no cost over the Internet environment. Consumers did not have to pay for products such as Internet search engines, social media networks, or online streaming. Data-driven companies could take advantage of consumer data to expand their businesses and generate new revenues. Consumer data allowed new small start-up companies to compete with existing businesses by lowering overhead costs when introducing new products to markets. Marketing agencies could use data to understand consumer preferences so that they could target consumers precisely. Finding 1 is that there are various potential benefits of using consumer data in the data-driven economy.

**Theme 2: Data sources for consumer data.** Companies collected data from multiple sources, including both public and private sectors. They obtained data from local government records, bought data products from data providers, or collected data directly from consumers. They used various methods, such as web crawler technologies, surveys, questionnaires, or contracting with data providers, to aggregate consumer data. Companies collected a wide range of consumer data points depending on their industries and business objectives. Finding 2 is that private companies collect consumer data from multiple sources in both public and private sectors.

**Theme 3: Identity data.** Companies collected various personal data elements that allow them to identify and verify information about individuals. Companies claimed that they protected private data by gathering anonymous data. However, the objectives of digital advertising or recommendation systems were to offer products that would meet individual preferences or to tailor services to individual requirements. Such objectives would be meaningless if companies did not connect data together and target potential customers at the individual level. With the advance in new technologies, such as big data, artificial intelligence, or big data analytics, companies could combine different data sets and build relationships among data elements to obtain a complete picture of an individual. Finding 3 is that companies aggregate consumer data that can identify individuals precisely.

**Theme 4: Transparency in data-driven businesses.** Privacy documents were often lengthy with technical terms and conditions. It was difficult for consumers to understand their rights in term of data privacy protections and options. Consumers needed to have clear information, so that they could make decisions about their personal data. Consumers might have known that companies collected their data but may not have been aware of the business practices of collecting data by third-party or data provider companies, since these companies often did not interact directly with consumers. Third-party companies also did not disclose exactly which data sources they obtained data from or which affiliates and buyers they sold data products to. Finding 4 is that business practices of collecting and using consumer data are not transparent.

**Theme 5: Data products and services.** Companies aggregated a variety of consumer information depending on their business objectives or on the industries they wanted to target. They collected, analyzed, and categorized consumers into groups and profiles based on shared common characteristics or behaviors. Data catalog documents from the 12 companies in this study revealed two types of data products: actual data and processed data. Actual data contained information related to individuals, including names, age, gender, phone, date of birth, and other personal data elements. Processed data were derived from aggregating actual data. Companies used emerging technologies, such as data mining and big data analytics, to transform actual data into analytic forms, scoring systems, or segments. Individual profiles were another form of processed data, developed by combining associated actual data from multiple sources. Finding 5 is that companies offer a wide range of consumer data products and services.

**Theme 6: Affiliated companies and partners of data providers.** Businesses operating in different industries bought consumer data from data providers and third party companies to learn more about their potential customers. Data provider clients were in a number of industries, including hotel chains, retailers, banking institutions, telecommunications, technology companies, real estate, manufactures, media agencies, and nonprofit organizations. Archival documents also mentioned companies which bought consumer data products and resold them to other third-party companies. Finding 6 is that affiliated companies and partners of data provider companies have businesses in different industries.



**Theme 7: Purposes of using consumer data.** Companies collected and used consumer data for many purposes. They also bought consumer data products and analytic information from data providers to enrich information about potential consumers. They used data for both internal and external businesses. For internal purposes, companies might have used data to improve internal business operations, auditing, or quality control. For external purposes, consumer data might have been used for marketing, fraud detection, or for improving customer engagement. Finding 7 is that companies collect and use consumer data for many business purposes.

**Theme 8: Complexity of the data life cycle in data-driven industries.** Data-driven companies collected consumer data from multiple sources, including federal and local governments, private industry, and directly from consumers. Companies also bought, exchanged, and sold a wide range of data products and solutions to their affiliated partners, third-party companies, and customers. It may have been impossible for consumers to keep track of their data's circulation or to find out how private companies had obtained their data. Finding 8 is that the data-driven industry is complex and data provider networks hold critical roles in the data-driven economy.

**Theme 9: Ownership and consent.** Consumers might have known that companies collected their personal data when they used online products and services. Documents collected from the 12 companies in this study show that they collected data from different sources. In many cases, they collected data not only directly from consumers, but they also obtained or bought from third parties. Technology provided companies the capability to aggregate various personal data elements to build completed

individual profiles, generate scoring systems, and sell a variety of data products to third parties. Consumers might not have known which companies collected their data, what personal data elements companies obtained, or which companies bought their data. Finding 9 is that third party companies collect and use consumer data without individuals' knowledge and consent.

**Theme 10: Limited access and control over personal information.** Privacy policy documents from the 12 companies in this study showed that they provided consumers options to opt out of data collection processes depending on where consumers lived. Five of these companies stated explicitly that they allowed consumers to correct their own personal data. Under the California Consumer Privacy Act (CCPA), companies allowed California residents to access and remove their data from company database systems. The opting out or deletion option was only fulfilled at the company where the individual submitted the request. Consumers needed to find and submit requests at all of companies where they wanted companies to delete data. It was impossible for an individual to completely remove their personal data from private companies. Finding 10 is that consumers had limited access and control over their personal information.

**Theme 11: Risks, issues, and ethics of using personal data.** There were many issues and risks associated with collecting and using consumer data in data-driven businesses. These could be related to data breaches, misuse of data, unauthorized access, discrimination, and consumer consent. Data breaches were one of the main risks, happening frequently in both public and private sectors. Data breach incidents might have exposed personal information to public or third-party organizations. Other concerns were

that companies developed individual profiles, generated scoring systems, and segmented consumers to target them with different products and prices. Bias might occur in computer algorithms that were used to predict consumer preferences or behaviors. Data provider companies also sold personal data to their affiliates and third parties without consumer knowledge. Finding 11 is that many purposes of using personal data may cause harm to consumers.

**Theme 12: Data protection policies and regulations.** Collected archival documents revealed that there were limited federal data protection regulations that give consumers the right to access, control, or remove their personal data being collected by third-party companies. Companies provided consumers options to opt out, view, and delete their personal data depending on where they live. Under CCPA law in California, companies allowed California residents to submit requests for viewing or removing their data from company database systems. Finding 12 is that there is no comprehensive federal data protection regulation in the United States that gives consumers the right to control and protect their personal data from being collected by private companies.

### **Evidence of Trustworthiness**

#### **Credibility**

Credibility is one of the critical pieces of evidence through which qualitative researchers seek to ensure the validity of a study. According to Guba (1981), qualitative researchers are most concerned about testing the credibility of collected data interpretations and research findings. Shenton (2004) discussed several techniques that a researcher can use to increase credibility. These might include research methods such as

random sampling or triangulation. In this study, I selected a qualitative archival research method to explore business practices of collecting and using consumer data without an individual's consent. Archival materials were the primary data sources in this study. In term of random sampling, I developed the archival document selection protocol (Appendix A) to collect documents from U.S. Congressional digital libraries, federal agencies, and private companies. I collected relevant data based on the research question, the Patterson and Davis (2012) conceptual framework of big data ethics, and existing concepts from the literature view. I used different search keywords in the protocol (Appendix A) to find archival documents. I was also open to any leads that might lead to a better understanding of the research problem. For example, I often found new hints from reference materials in a particular document or report. Archival documents obtained from multiple sources pertained to different aspects of the research problem. I could compare and verify information from various opinions and points of view. This also allowed me to triangulate information to ensure the validity of the findings.

### **Transferability**

As Myers (2002), and Shenton (2004) discussed, transferability is the extent to which the research findings can be generalized to different populations. Shenton (2004) suggested that qualitative findings often draw from the specific research setting of a particular environment. It might not be feasible to apply the findings to other situations that have different contexts. The purpose of this qualitative archival research is to explore private company practices of collecting and using consumer data without an individual's consent. I conducted this study using archival documents relevant to business practices of

data-driven companies which have products and services based on consumer data.

Although the findings of this study are limited to this setting, other researchers may view this study as an example in a broader environment. In that case, researchers might associate or extend this study's findings to their new studies.

### **Dependability**

This evidence of trustworthiness refers to the extent to which research findings can be replicated by other researchers with the same research setting, context, and method. In this study, I used an archival research method to conduct this study. Archival documents were the primary data sources. I developed an archival document selection protocol (Appendix A) that allows researchers to collect relevant archives systematically. The protocol provided necessary steps to identify data source locations, search keywords, document metadata of archives, and review documents. I present in detail the six phases of the thematic data analysis method that other researchers can follow to analyze collected documents. I also describe the challenges of the document selection procedure in the digital environment due to changes in technologies.

### **Confirmability**

To reduce bias in the research findings, I applied the triangulation method to data sources by collecting archival documents from multiple sources at U.S. Congress digital libraries, federal agencies, and private companies. Collecting archives from different sources allows researchers to cross-check information and verify the reliability of sources. This method also provides the researcher with different points of view about the practices of using consumer data in the digital age. I used a thematic data analysis method

to review and analyze different aspects of this phenomenon. To reduce bias in the findings, I gathered archives from organizations that I was not associated with and did not have any relationships with. The findings were based on evidence extracted from collected documents.

### **Study Results**

The following common themes were emerged from the collected archival documents. The themes were (a) benefits of using consumer data, (b) data sources for consumer data, (c) identity data, (d) transparency in data-driven businesses, (e) data products and services, (f) affiliate companies and partners of data providers, (g) purposes of using consumer data, (h) complexity of the data life cycle in data-driven industries, (i) ownership and consent, (j) limited access and control over personal information, (k) risks, issues, and ethics of using personal data, and (l) data protection policies and regulations. The following findings were derived from the 12 common themes. The findings provide answers with supporting evidence to the research question for this qualitative archival research: How do private company practices of collecting and using consumer data without an individual's consent align with ownership, transparency, ethics, and consumer privacy laws?

#### **Finding 1: Potential Benefits**

There are various potential benefits of using consumer data in the data-driven economy. In collected transcripts of a Congressional hearing about data privacy protection, a U.S. senator stated that “the benefits of online tracking and data collection are very clear. Facebook is free. Gmail is free. Google Maps is free. There are thousands

of mobile device applications that are free” (U.S. Senate, Committee on Commerce, Science, and Transportation, 2012, p. 3). In recent years, consumers have used various Internet-connected devices and applications to socialize, entertain, communicate, and shop on the Internet. In collected documents of a Congressional hearing on fair market in the digital economy, a U.S. representative stated that “the Internet and digital platforms are so deeply woven into American lives” (U.S. House of Representatives, Committee on Small Business, 2019, p. 3). Online activities generate massive data over the Internet that private companies can collect to learn more about consumer online activities or behaviors. For example, consumers could use smartphones or tablets to search information, the news, or home merchandise. They could also use these Internet-connected devices to connect with friends and relatives anywhere at any time. In a collected opening statement within a joint hearing about how companies made decisions based on data, a U.S. representative stated that “these services are convenient, efficient, and provide valuable and tangible benefits to American consumers” (U.S. House of Representatives, Committee on Energy and Commerce, 2017, p. 9).

Privacy policies from the 12 companies showed that they used consumer data to understand marketing trends, verify financial credits, develop marketing strategies, and improve product quality as well as customer service. Consumers might get recommendations for products that they were searching. Consumers gained benefits from reaching a wide range of products and services offered by various companies. This allowed small companies to engage potential customers and compete with well-known companies. Aggregating banking transactions from multiple sources allows companies to

prevent identity fraud. Banking institutions could analyze credit transactions or spending activities to detect abnormal transactions. For example, based on residential address and historical transactions, banking institutions might put a lock on an individual's account if they saw a certain transaction occurring far away from the account address.

In the healthcare and insurance industries, companies could use health data such as weight, drinking behavior, diabetes, or depression to introduce appropriate healthcare prevention programs. Collecting and using consumer data for marketing purposes was not a new activity for private companies. The difference was that companies had the opportunity to use advanced technologies, such as big data, artificial intelligence, or computer machine learning algorithms, to collect and analyze massive amounts of digital data over the Internet environment. This allowed companies to forecast business trends and better understand consumer preferences.

In addition to the advantages of using consumer data for marketing, customer engagement, or supply chains, data-driven companies also create jobs in the U.S. economy. In a collected opening statement within a joint hearing about data-driven companies, a U.S. Congressman stated that “the companies behind the services have created thousands and thousands of jobs and brought the U.S. into the forefront of technology and innovation” (U.S. House of Representatives, Committee on Energy and Commerce, 2017, p. 9).

## **Finding 2: Data from Multiple Sources**

Private companies collect consumer data from multiple sources in both public and private sectors. In a hearing about fair market in the digital economy, a U.S.



representative mentioned that “our data is an economic asset which garners more value as one collects more of it” (U.S. House of Representatives, Committee on Small Business, 2019, p. 2). Private companies use different techniques and technologies to collect data from Internet-connected devices. For example, healthcare providers could use IoT devices to monitor patients’ condition in real-time and send urgent alerts to doctors in case of emergency. In collected transcripts of a Congressional hearing about the IoT technology, a U.S. senator mentioned that “IoT devices can collect sensitive consumer and business data” (U.S. Senate, Committee on Commerce, Science, and Transportation, 2015, p. 2). A witness testified that “IoT enables the collection of an unprecedented quantity and quality of data through sensors and devices” (U.S. Senate, Committee on Commerce, Science, and Transportation, 2015, p. 8).

Archival government agency reports showed that web crawler technology was another common method that companies used to capture website contents on the Internet environment. Web crawler software allowed companies to search, download, and archive website contents in their repositories. Companies also acquired data from local government records, bought processed data sets from other companies, and obtained data directly from consumers. Archival documents showed that companies often obtained or exchanged data in large volumes. They might collect more data than what they wanted for their businesses. In the hearing on data privacy protection, a U.S. senator stated that “they are collecting more than simply the information that you type in. And a lot of Americans aren’t necessarily aware of that” (U.S. Senate, Committee on Commerce, Science, and Transportation, 2012, p. 4). Companies used the additional data for purposes

other than their original intent for using data. Archival documents disclosed that companies collected broad ranges of personal data elements. Table 8 shows examples of common personal data elements that were being collected and used by companies.

Table 8

*Examples of Personal Data Elements*

Data Type	Data Element
Identification	Name, date of birth, age, gender, phone number, ethnicity, social security number, passport number, resident address, email address, driver's license, or state identification card number.
Internet-connected device	IP address, type of browser, time zone, location, or type of device (smartphone, tablet, personal computer, or smart television)
Geographic	Global positioning system (GPS) location, sensor data, and information from Bluetooth-enabled devices or cellular phone towers.
Income	Personal income and household income.
Finance	Banking accounts, payments, or credit transactions.
Education	Colleges and universities attended, professional certificates, degrees, or education level.
Medical record	Health insurance, types of insurance, cholesterol level, blood pressure, depressive disorder, Bipolar disorder, or Parkinson's disease.
Employment	Company name or occupations.
Online and offline activity	Web browser history, website types, time spent online, visited web pages, or search keywords.
Behavior	Shopping preferences or household purchases.
Public record	Census information, voter registrations, motor vehicle registration, or court files.
Property	Car information or real estate information.

U.S. Congressional reports and privacy policies from the 12 companies in this study revealed that companies collected public records of consumers from federal and state agencies. Public agencies provided various records, such as census information, voter registration, court filings, driver's licenses, or professional certificates. Private

companies also purchased consumer data from other private companies operating in different sectors. These private sectors were retail, banking, entertainment, and other third-party data providers.

**Retail sector.** Third-party companies bought various pieces of information on purchases from companies operating in the retail industry. Purchased data might contain names, addresses, or purchased items. Retailers collected such information from purchase transactions occurring both online and in-store.

**Banking sector.** Banking institutions offered financial data as well as data on mutual funds, bank withdrawals, bank deposits, or financial assets to third-party companies. Third-party companies could buy and develop financial scoring products.

**Entertainment sector.** Companies operating in the entertainment industry could provide data, including popular movies, content, or shows. They also offered data on historical activities such as searches, watched content, reviews, and ratings.

**Third-party data providers.** Besides collecting or buying data from individual private companies in different industries, companies also shared and exchanged data with other data providers. The practice of exchanging data among data provider companies allowed them to enrich their data sources and enhance their capacity to analyze consumer data.

Data provided directly from consumers was another data source. Companies used different ways to collect data directly from customers. They conducted marketing surveys, offered sweepstakes, or promote discount programs. Through these programs, private companies could obtain a wide range of information from consumers, including

both online and offline data. Collected data contain information about household demographics, income, or purchased items. Besides data on shopping behavior, companies also conduct surveys to collect data on health, such as data on blood pressure, cholesterol levels, depression, Parkinson's disease, and other health conditions. To encourage consumers to respond to marketing surveys, companies promoted discount programs or sweepstakes and initiated incentives so that consumers could gain benefits from answering questionnaires. Through these marketing questionnaires, consumers disclosed various pieces of personal information to businesses as well as to third-party companies. Archival documents showed that companies often collected data directly from consumers and shared it with other companies.

Companies also collected historical records of online activities directly from consumers. Documents from one company stated that it collected visited websites, information about average time spent online, or paths leading to visited webpages. One of the 12 companies in this study indicated that it did not collect data directly from consumers. This company also did not gather sensory data such as video or voice from consumers.

Social media networks were one of the main data sources for consumer data. Online users used social media networks to communicate and share personal information with friends or family members. Data provider companies gathered data that consumers posted on social networks such as Facebook, LinkedIn, or Twitter. Social media network data might contain names, hobbies, location, education levels, occupations, or friends. Types of data collected from social networks varied from text to pictures to media files.

**Finding 3: Identity Information**

Companies aggregate consumer data that can identify individuals precisely. Privacy policy documents collected from the 12 companies in this study revealed that they collected various pieces of personal identifiable information. These data elements might include a name, address, email, phone number, government-issued identifiers, credit card information, and other identifiable data points. Even when companies might not collect personal information directly from consumers, advanced technologies such as big data, artificial intelligence, big data analytics, and computer machine learning algorithms allowed companies to aggregate data from multiple data sources and build complete individual profiles.

Companies aggregated data from different sources, including public organizations, private companies, social media networks, surveys, or consumers themselves. Each data source provided bits of identifiable information about contact information, employment, device information, financial data, and online activities. Using technologies to aggregate and link these data elements together provided companies the ability to verify and identify an individual precisely. Companies relied on identifiable data elements to perform identity verification. At the same time, these elements could put consumers at risk when these personal data were exposed to unauthorized access or misuse. In a hearing about identify verification before the Subcommittee on Oversight and Investigation, a U.S. representative stated that “while these breaches themselves are troubling enough” (U.S. House of Representatives, Subcommittee on Oversight and Investigations of the Committee on Energy and Commerce, 2017, p. 2). The U.S.

representative emphasized that “they also raise a subtle more complicated series of questions and issues around the ways in which organizations including government agencies, banks, health care organizations, and retail companies perform identity verification of their citizens and their customers” (U.S. House of Representatives, Subcommittee on Oversight and Investigations of the Committee on Energy and Commerce, 2017, p. 2). Companies might increase data protection by separating identified information from sensitive data, but data management technologies have the capability to combine massive records from multiple sources and build data relationships to identify individuals.

#### **Finding 4: Transparency in Data-driven Businesses**

Business practices of collecting and using consumer data are not transparent. In collected transcripts of a joint hearing about social media privacy, a U.S. senate stated that “consumers must have the transparency necessary to make an informed decision about whether to share their data and how it can be used” (U.S. Senate, Committee on Commerce, Science, and Transportation, & Committee on The Judiciary, 2018, p. 6). Privacy statements or terms of use documents were often lengthy and difficult for consumers to understand, with technical terms and general agreements. In collected transcripts of another Congressional hearing about how companies made decisions based on data that impact consumers, a witness, who was a law professor, stated that “if consumers tried to read the disclosures, they would of course not understand them and would not be able to put them to profitable use. To use complex information, one needs

experience and expertise which people simply do not have” (U.S. House of Representatives, Committee on Energy and Commerce, 2017, p. 16).

Archival documents revealed that data provider companies often did not specifically disclose either their data sources or their clients. They also had contracts with their clients that prohibited the clients from disclosing data sources. In a joint hearing about how companies made decisions based on data that impact consumers, a U.S. congressman stated that “although there are legitimate reasons and benefits to the collecting and using information online, we want to ensure that Americans understand how their information is being used” (U.S. House of Representatives, Committee on Energy and Commerce, 2017, p. 4).

Companies needed to provide consumers flexible choices about their data when they used any free-of-charge products. A U.S. representative stated in a joint hearing about social media privacy that “it is consumer choice. Do users understand what they are agreeing to when they access the website or agree to terms of service?” (U.S. Senate, Committee on Commerce, Science, and Transportation, & Committee on The Judiciary, 2018, p. 21). The U.S. representative continued highlighting that “are websites upfront about how they extract value from users, or do they hide the ball? Do consumers have the information they need to make an informed choice regarding whether or not to visit a particular website?” (U.S. Senate, Committee on Commerce, Science, and Transportation, & Committee on The Judiciary, 2018, p. 21).

Policy documents from one company in this study stated that it collected data from social media networks, to verify identity or enhance individual profiles. However,

this company did not specify which social networks it collected data from or what data elements it obtained. In collected transcripts of a Federal Trade Commission workshop about consumer protection issues in online advertising, a panel member stated that “consumers don’t have knowledge or transparency about the kind of ways that their privacy is being invaded upon and how companies are using their information” (U.S. Federal Trade Commission, 2018, p. 171). The argument was that companies gathered consumer data from multiple sources. They then generated predictive models and scoring systems of consumer financial status, behaviors, preferences, or health conditions. These predictive models or scoring systems represented groups of people, not specific individuals. Thus, companies could not provide information on individuals so that consumers would be able to make corrections. Consumers needed transparency in data collection processes when companies gathered or bought personal information from other companies. A witness statement in a joint hearing about social media privacy stated that “companies collecting data must provide transparency regarding their data processes” (U.S. House of Representatives, Committee on Energy and Commerce, 2018, p. 123).

#### **Finding 5: Data Products and Services**

Companies offer a wide range of consumer data products and services.

Companies developed data products and services based on two types of data, actual data and processed data. Actual data contain various personal data points, such names, age, gender, phone, date of birth, and other personal data elements. Data provider companies built individual profiles based on various actual data elements, segmenting consumers into different categories which shared common characteristics. Policy documents from



the 12 companies in this study revealed household incomes, financial assets, financial credit data, fraud detection, and other data products, providing many characteristics of individuals. These financial data products provided businesses insights into the financial status of both existing and potential customers, so that they could target them with new products at the right time. In a hearing about digitalization, data, and technology in the financial industry, a U.S. representative mentioned that “many products and services in the FinTech sector revolve around big data analytics, data aggregation, and other technologies that make use of consumer data” (U.S. Senate, Committee on Banking, Housing, and Urban Affairs, 2018, p. 1).

Processed data included information which was analyzed and processed into scoring systems or predictive models. Third-party companies or marketing agencies could buy these predictive modeling products to forecast common characteristics, health conditions, or lifestyles of people living in a particular city or zip code. For instance, archival documents showed that companies might be able to predict an individual’s health condition based upon purchase of medication; or businesses might predict the hobbies of an individual based on purchases of sportswear and sports equipment. Companies bought scoring measures to assess certain consumer spending behaviors or financial conditions. Data providers also collected offline data to know more about consumers when they are not on the Internet. In addition to collecting online data, companies also gathered offline data to better understand consumer behaviors, so that they could deliver accurate advertisements to potential customers when they got online. Offline data might help businesses better understand consumer behaviors to optimize

their advertising campaigns in the online environment. A U.S. representative commented in a Congressional hearing about digital advertising ecosystem that “information collected about our online activity is increasingly being merged with our offline identity to create extremely detailed profiles” (U.S. House of Representatives, Subcommittee on Digital Commerce and Consumer Protection of the Committee on Energy and Commerce, 2018, p. 7).

Archival documents showed that some companies removed certain types of data points from data products and services. For example, companies removed information about children 12 or under to comply with the Children’s Online Privacy Protection Act (COPPA). Collected documents revealed some data provider companies checking information, such as dates of birth or age, to ensure that children’s data and other associated information were removed from their products. The privacy policy of one of the 12 companies in the study stated that the company did not include detailed personal information, such as medical records, financial transactions, or social security numbers, in their marketing products.

There were different ways that companies stored consumer data. They might have stored records of completed individual profiles, which contained all of the associated information, such as contact information, career, shopping style, or favored sports. On the other hand, companies might have managed a list of events associated with an individual’s record. When needing to build a complete profile, they developed queries to retrieve these events and link them to the individual’s record. Archival documents showed that companies managed and stored consumer data in multiple databases

depending on types of data, products, or services. For example, they separated real estate databases from court record databases. Companies also stored data used for marketing separately from product data. In term of the length of time for keeping data, companies often archived data permanently in their data management systems. They even kept out-of-date information, such as phone numbers or residential addresses, to build historical records as well as verify the identity of individuals.

### **Finding 6: Affiliates and Partner Networks**

Affiliated companies and partners of data provider companies have businesses in different industries. Documents of data products and services collected from the 12 private companies showed that they had data solutions for various business sectors. Their affiliates could use data products to engage customers, fraud detections, identity verification, and other business purposes. Table 9 shows examples of the industries for which the 12 companies in this study provided data.

Table 9

#### *Examples of Industry Data*

Industry	Solution
Financial service	Data providers offered a number of financial solutions such as credit reports, fraud detection, fraud prevention, credit monitoring, information verification, or risk mitigation.
Automotive	Data providers had various automotive data, including vehicle history, number of operational cars, maintenance schedule, automotive credit, where consumers bought cars, or what are popular car models.
Insurance	Data providers helped the insurance industry to identify existing customers who were highly profitable, find new high-value customers, or predict customer preferences.

*(Table continues)*

Industry	Solution
Healthcare	Data providers had data solutions for hospitals, medical groups, or healthcare providers. Data solutions included digitalizing medical records, insurance claims, or patient identity verification. Data also helped medical providers better understand the health condition of patients.
Mortgage	Data providers had data that could help with monitoring debt, identifying qualified borrowers or renters, verifying employment information, or auditing loan applications.
Retail	Data providers had data solutions that could help retailers identify influential customers, retain loyal customers, target potential customers, or improve customer experience.
Travel	Data providers had data that could help identify rich customers or frequent travelers, and attract them by providing offers at the right time.
Restaurants	Data providers used data to help the restaurant industry identify potential customers various data elements, such as household income, household spending scoring systems, or payment affordability.

One company in this study stated that it did not sell data to individual customers. Its clients, both for-profit and nonprofit organizations in different sectors, included cloud computing technology firms and security consultants. Clients bought data products for advertising, marketing, fraud detection, and customer services. Other companies in this study did not state explicitly their client information in privacy documents. Archival documents revealed that data provider companies and their affiliated partners often had contract agreements that specified that they would not disclose data sources or the companies which bought data products from them.

### **Finding 7: Purposes of Using Consumer Data**

Companies collect and use consumer data for many business purposes. In collected documents of a Congressional hearing about protecting consumer information, a U.S. representative stated that “companies across the globe are changing the way they

collect and use consumer data, and we are seeing more sophisticated practices, which obviously results in more challenges to Americans' privacy" (U.S. House of Representatives, Subcommittee on Communications and Technology of the Committee on Energy and Commerce, 2018, p. 75). Policy documents from the 12 companies in this study revealed that companies used consumer data for two main areas, including internal business purposes and external commercial data products. Table 10 shows examples of the purposes of using consumer data.

Table 10

*Purposes of Using Consumer Data*

Purpose	Description
Internal business	Companies used consumer data for business operation improvement, customer engagement, auditing, quality control, decision making systems, regulation compliance, and other areas of internal business management.
External business	Companies used consumer data to generate data products for business statistical models, to forecast business trends, to manage supply chains, and for other business purposes including marketing, fraud prevention, identity verification, or risk mitigation.

For internal business purposes, companies used consumer data to improve business operations, audit business transactions, enhance customer service, or comply with current regulations. For example, documents from one of the 12 companies in this study stated that it collected online user activities, such as visited websites, average spending time online, or browser types, to improve the content of websites.

For external data products and services, archival documents showed that companies used innovative technologies, such as big data, data mining, or advanced data

analytics, to develop data products for different businesses, sectors, and industries. For example, there were various ways of using consumer data for marketing purposes in data-driven business models. Data provider companies offered various marketing products and services to other companies. They clustered consumers into groups which shared common demographics, behaviors, characteristics, or lifestyles. Segregating consumers into different groups allowed businesses to better understand consumer preferences so that businesses could target them effectively with specific products and prices.

In addition to classifying consumers into groups, private companies used new technologies such as big data and advanced data analytic algorithms to build predictive mathematic models and scoring measures about consumer behaviors. Scoring systems assigned a number or range to an individual as a score which represents the performance of that person in specific areas, such as finance, insurance, healthcare, good driving, politics, or social influencing. In a collected documentary about technology companies collected, monitored, and control online user data, a business professor discussed that companies “collect data and some of it is used to improve the service to you, but even more of it is analyzed to train what they call models patterns of human behavior” (VPRO Documentary, 2019, p. 3).

### **Finding 8: Complexity of the Data-Driven Industry**

The data-driven industry is complex and data provider networks hold critical roles in the data-driven economy. Data provider companies collected consumer information from multiple sources. They could acquire data from the government, buy data from other companies, exchange data with affiliated partners, or collect data directly from

consumers. Companies collected large volumes of individual data points as well as business transactions. Each data source provided a few data points for consumers. Companies could use new technologies, such as big data or computer machine learning algorithms, to combine data together and build complete individual profiles.

Data provider companies collected large volumes of consumer data and sold it to other companies to use for various business purposes. These companies also provided and exchanged data with each other to enrich their data sets. Besides affiliates and partners operating in the United States, data providers also sold their data products to foreign companies. Privacy documents from the 12 companies in this study stated that they had contracts with offshore firms that allowed them to access U.S. data products. It may have been impossible for consumers to keep track of their data's circulation or to find out how companies obtained their information.

Data-driven companies made profits based on data that were either voluntary or involuntary collected from consumers. In a joint hearing before the subcommittee on communications and technology and the subcommittee on digital commerce and consumer protection, a U.S. representative stated that “tech companies and online platforms make their money because they know who you are, where you are, what you like, what photos and videos you take and watch, and what news you read” (U.S. House of Representatives, Committee on Energy and Commerce, 2017, p. 9). In exchange, consumers received numerous convenient and efficient services free of charge over Internet platforms. U.S. representative noted that “consumers are willing to share

personal details about themselves - names, locations, interests, and more” (U.S. House of Representatives, Committee on Energy and Commerce, 2017, p. 9).

Although consumers may not have been aware of business practices of collecting and using consumer data by data provider companies, data provider companies held key roles in the data-driven economy. Data providers facilitated the growth of businesses by using data to forecast business trends, improve product quality, and better understand customer experiences. Data provider practices of collecting and using consumer data were often hidden from public view since the companies did not have direct contact with consumers. Privacy documents from the 12 companies in this study did not disclose the sources from which they collected data or the customers who bought their data products. These documents mentioned in general terms that their data products provided solutions for different business sectors, and industries. Affiliated companies and partners needed to sign contracts agreeing not to disclose the sources from which data providers obtained consumer data.

### **Finding 9: Ownership and Consent**

Third-party companies collect and use consumer data without individuals’ knowledge and consent. In a testimony on data security before the Committee on Financial Services, a witness presented that “data brokers obtain and share vast amounts of consumer information, typically behind the scenes, without consumer knowledge. Data brokers sell this information for marketing campaigns and fraud prevention, among other purposes” (U.S. House of Representatives, Subcommittee on Financial Institutions and Consumer Credit of the Committee on Financial Services, 2017, p. 65). In collected



documents of another Congressional hearing on the IoT technology, a U.S. senator remarked that “some companies may transmit the information they collect to third parties without consumer consent” (U.S. Senate, Committee on Commerce, Science, and Transportation, 2015, p. 3). Consumers might have known that companies collected their online activity data when they accessed the Internet via various Internet-connected devices. However, consumers may not know the purpose that companies intended to use their personal data. In collected transcripts of a Congressional hearing on identity verification in data breach incidents, a witness testified that “individuals' personal data is also frequently collected without their informed consent, that is it's obtained without them consciously opting in to the service and the purpose for which it's being used” (U.S. House of Representatives, Subcommittee on Oversight and Investigations of the Committee on Energy and Commerce, 2017, p. 19).

It was also not clear that consumers understood the capability of advanced technologies that allowed companies to collect data even when consumers were offline. Combining offline data with online data could enrich personal profiles and allow companies to target customers more accurate and effective. For example, companies might have put cookie information in web browsers, so that marketing agencies could keep track of the movements of consumers and target them when they were online again. In collected transcripts of a documentary about technology companies collected, monitored, and control online user data, a business professor at a well-known university in the United States discussed that “misconception of what's really going on? We think that the only personal information they have about us is what we've given them and we

think we can exert some control over what we give them” (VPRO Documentary, 2019, p. 3). However, in the reality, the professor explained that “the information that we provide is the least important part of the information that they collect about us” (VPRO Documentary, 2019, p. 3). Behind the scenes, companies “retrieve a lot of information from the digital traces we leave behind unwittingly” (VPRO Documentary, 2019, p. 3).

Companies used various technologies to collect consumers’ data whenever they accessed the Internet via mobile applications or computer web browsers. Computers stored historical online activities through web browser tracking cookies, historical online activity sniffing, or device fingerprinting. The collected archival materials disclosed that companies could track online user activities across Internet-connected devices and applications. Companies could follow an individual through multiple devices, including a laptop, tablet, or mobile phone. A witness in a Congressional hearing about identity verification presented that “numerous digital threats to consumers, from data breaches to data brokers running amok to the very architecture of the digital ecosystem, where nearly every company, known and unknown, is tracking consumers, building a dossier on them” (U.S. House of Representatives, Subcommittee on Financial Institutions and Consumer Credit of the Committee on Financial Services, 2017, p. 76). Companies also collected location data from smartphones or mobile devices. A witness in a Congressional hearing about the effect of privacy regulation on consumers and competition testified that “a large number of popular mobile apps gather detailed location information. Third parties sell profiles based on location information” (U.S. Senate, Committee on Commerce, Science, and Transportation, 2016, p. 44). In another hearing on protecting consumer proprietary

information, a witness confirmed that “devices often have much more detail location information than what carrier location provides” (U.S. House of Representatives, Subcommittee on Communications and Technology of the Committee on Energy and Commerce, 2018, p. 47). Another witness agreed that “most consumers don’t anticipate or know the extent to which somebody could be tracking them” (U.S. House of Representatives, Subcommittee on Communications and Technology of the Committee on Energy and Commerce, 2018, p. 79).

Digital technologies allowed companies to collect both online and offline data about consumer preferences and needs. They stored what websites online users visited, what web pages users viewed, or when online users access to certain web pages. These data allowed businesses to understand what information online users searched when they were online. Although new technologies allowed online users to block cookies that were collecting data, companies found new ways to track online users through other Internet-connected devices. A U.S. representative stated in a hearing on data privacy protection that “these companies watch your behavior, and they measure your behavior - how long you linger on a site, your specific searches. A lot of people think they’re just going in and searching privately. Somebody’s watching you. Somebody’s tracking you” (U.S. Senate, Committee on Commerce, Science, and Transportation, 2012, p. 5).

In many cases, companies operating in the data-driven economy did not gather data directly from consumers. Instead, they bought consumer data products from data provider companies without consumer knowledge. Data provider companies aggregated consumer information from multiple sources, including public and private organizations.

In addition to buying data products or services from data providers, companies also exchanged data with affiliated companies and partners so that they could enrich their data sets and build complete consumer profiles. Consumers had limited understanding of these business practices.

### **Finding 10: Access and Control Over Personal Information**

Consumers had limited access and control over their personal information. In collected documents of a Congressional hearing on protecting consumers from data breaches, a U.S. senator remarked that “the consumer does not have a choice on the data that you’re collecting. That’s what I hear from my consumers. That’s what I hear all the time” (U.S. Senate, Committee on Commerce, Science, and Transportation, 2017, p. 37). Companies published terms of use or privacy policy documents on their websites that provided instructions for consumers on how to request to opt out of data collection processes or how to access and view their own personal information. The collected archival materials showed that consumers could submit requests to access and view their own actual data. To fulfill options for opting out or for accessing data, companies required that consumers provide personal information, including name, address, phone number, utility bills, or credit information, so that they could verify the identity of requested person. In collected policy documents, companies stated that they only used these pieces of personal information for verification purpose. The opting-out or accessing options did not guarantee that consumers would see all of data related to their profile or the processed data that companies developed based on the actual data.

To comply with the California Consumer Privacy Act, the 12 companies in this study provided options to consumers for opting out, accessing, and deleting their data. California residents could submit requests to access, view, and delete their personal data from company database systems. They could submit requests via the online form, mail, and telephone. Documents from the 12 companies in this study provided consumers options to opt out, or to access or delete personal data, over the company website. The access and delete options were only available to California residents under CCPA law. There were companies which did not accept requests for accessing or deleting from residents living outside the State of California. On the opt-out request website, companies discussed the impact of opting out of data collection processes at the top of the web page. Consumers needed to scroll all the way down to see the request form at the bottom of the page. California residents needed to provide personal information, including name, address, date of birth, phone number, and email, to view or remove their personal data from company databases.

According to privacy documents from the 12 companies in this study, consumers could hire authorized agents to submit opting-out, access, or deletion requests on their behalf. Authorized agents must be registered with local government to represent consumers. In the same manner as requests directly from consumers, authorized agents must submit appropriate information to data providers, so that they could verify identity and execute the request. The resulting reports from executed requests were sent to consumers directly.

Although companies provided consumers with options to opt out of their operation of collecting and selling data to third-party companies, the opt-out request was only executed at the company where an individual submitted the request. Consumers needed to find and submit opt-out requests at all of companies if they do not want private companies to use their data. When an individual selected the opting-out option, companies would not use that individual's data for marketing purposes in the future, but they did not delete the data from their database systems. Archival documents showed that companies kept opting-out records for identification purposes. Archival documents from the 12 companies in this study stated that they still stored and maintained certain pieces of personal information after performing the deletion option. It allowed companies to verify new collected data, comply with regulations, and detect data breaches. One company in this study also mentioned that they kept data even when customers did not have any business relationships with them. They might use opting-out data in processed data or anonymous products. Records of individuals who had opted out may still appear or link to other individual profiles, such as profiles of a spouse or friend. Companies maintained and archived consumer data permanently in the database systems unless there were restrictions in the contract conditions. Keeping historical personal data might benefit companies, by allowing them to have complete records for an individual. At the same time, the practice put consumers at risk when unauthorized access occurred, or when data was misused by third-party entities.

**Finding 11: Risks and Ethics of Using Personal Data**

Many purposes of using personal data may cause harm to consumers. In collected documents of a Congressional hearing before the subcommittee on communications and technology and the subcommittee on digital commerce and consumer protection, a witness stated that “the data-driven economy delivers enormous convenience and benefits too by offering personalized experience to consumers, but concerns about discrimination, manipulation, data security, and market power and the potential harms they might cause ought to be taken seriously” (U.S. House of Representatives, Committee on Energy and Commerce, 2017, p. 16). In collected transcripts of another Congressional hearing on using computer machine learning before the Subcommittee on Communications Technology, Innovation, a witness presented that “anytime you’re dealing with data, is going to be an issue of concern” (U.S. Senate, Subcommittee on Communications, Technology, Innovation, and the Internet of the Committee on Commerce, Science, and Transportation, 2017, p. 47).

Data breach is one of the main concerns in collecting and using consumer data. In collected transcripts of a Congressional hearing about a data breach incident occurring at a U.S. financial institution, a U.S. representative mentioned that “the criminals got basically everything they need to steal your identity, open credit card accounts in your name, and cause you untold frustration and financial calamity” (U.S. House of Representatives, Committee on Financial Services, 2017, p. 1). The U.S. representative emphasized that “this may be the most harmful failure to protect private consumer information the world has ever seen” (U.S. House of Representatives, Committee on

Financial Services, 2017, p. 1). This company failed to notify its customers about the incident immediately, so that customers could take appropriate steps to protect personal information. This company also did not disclose what personal information had been exposed to unauthorized parties. In collected documents of a Congressional hearing on data security and data breaches, a U.S. senator mentioned that “individuals expect companies collecting their sensitive personal information to do everything in their power to protect their data and their security and privacy, notify them promptly when there is a breach that endangers those consumers” (U.S. Senate, Subcommittee on Consumer Protection, Product Safety, Insurance, and Data Security of the Committee on Commerce, Science, and Transportation, 2018, p. 8). A witness also testified that “I would like to echo statements made by new leadership, and state publicly that it was wrong not to disclose the breach earlier. The breach should have been disclosed in a timely manner” (U.S. Senate, Subcommittee on Consumer Protection, Product Safety, Insurance, and Data Security of the Committee on Commerce, Science, and Transportation, 2018, p. 15).

Due to the sophistication and complexity of new technologies and computer networks, data breaches happened frequently at organizations in both public and private sectors. In a data breach hearing, a U.S. representative stated that “large-scale security breaches unfortunately are becoming all too common. By the increasing frequency and sophistication of cyber attacks, this clearly demands heightened vigilance and enhanced efforts to safeguard consumers” (U.S. House of Representatives, Committee on Financial Services, 2017, p. 2). Criminals stole consumer PII information “to buy a boat in their name, a house in their name, people are going to commit crimes in their name” (U.S.



Senate, Committee on Commerce, Science, and Transportation, 2017, p. 37), a U.S. senate remarked in a hearing about protecting consumers from data breaches. The U.S. senate added “people whose identities were stolen, they are going to have to clear their record for the rest of their lives” (U.S. Senate, Committee on Commerce, Science, and Transportation, 2017, p. 37).

According to policy documents from the 12 companies in this study, they implemented multiple security layers to protect consumer data from unauthorized access and data breaches. Companies also perform both internal and external audits to maintain the integrity of database systems. In cases of data breach occurring, company documents stated that they put alerts on company websites and notified individuals who were affected in the data breach incidents.

Private companies might misuse personal data or obtain unauthorized access to sensitive information, such as financial status or health conditions. For example, an individual might be denied for opening a line of credit due to inaccurate information within credit data products. That individual may not know exactly why he or she was rejected or where to correct the information. That individual may not be able to prevent the same thing from happening again in the future. Unauthorized access to personal data might pose serious threats to consumers. In collected transcripts of a Congressional hearing about improving data security, a witness presented that “threats to consumers can include fraud on existing accounts, new account identity theft, medical identity theft, tax refund identity theft, and imposters committing crimes using your identity” (U.S. House of Representatives, Subcommittee on Oversight and Investigations of the Committee on

Energy and Commerce, 2017, p. 69). In addition, companies might use consumer data for purposes that consumers did not consent for collecting and using. Evidential documents from a Congressional hearing about the transparency of using consumer data revealed a popular social media network company compromised their “users throughout the United States by invading their privacy, allowing for disclosure and use of information in ways and for purposes other than those consented to or relied upon by such users” (U.S. House of Representatives, Committee on Energy and Commerce, 2018, p. 187). This social media network company also misled their users that “they have full control over the use of their information, and undermining the ability of users to avail themselves of the privacy protections promised by the company” (U.S. House of Representatives, Committee on Energy and Commerce, 2018, p. 187).

In the same manner, unequal opportunities or discrimination may occur when businesses develop computer algorithms to generate scoring systems. Marketing agencies then use these scoring systems to target financially vulnerable consumers. In collected documents of a Congressional hearing on digitalization and data in the financial technology industry, a witness testified that “there are concerns that the enhanced use of algorithms may lead to more discrimination, a lack of transparency, or diminished access to essential services like credit” (U.S. Senate, Committee on Banking, Housing, and Urban Affairs, 2018, p. 41). Marketing agencies might offer different prices for the same product based on scores and characteristics associated with individual profiles. In a Federal Trade Commission workshop about consumer protection issues in online advertising, a panel discussion member presented that travel consumers might have

different ticket prices depending on their online activities. The panel member remarked that “there’s no transparency around how these fair prices are reached or what kind of information they’re using in order to serve you with that price” (U.S. Federal Trade Commission, 2018, p. 197).

Another concern of using machine learning algorithms to generate predictive models was that it was difficult to explain outputs of machine learning algorithms. In collected transcripts of a Congressional hearing on advanced computing solutions, a professor in computer science testified that “machine-learning results are often lacking in explanations, interpretations, or error bars, a frustration for scientists. And scientific data is complicated and often incomplete. The algorithms are known to be biased by the data that they see” (U.S. House of Representatives, Committee on Science, Space, and Technology, 2018, p. 32). There were several factors that might contribute to the bias in predictive data models. It could be using incomplete data, developing inaccurate algorithms, or incorrect predictive models. Unfortunately, the processes of collecting data, analyzing it, and generating scoring systems were not transparent. Consumers did not have control or options to intervene in the data processing of private companies. Collected transcripts of a data privacy and security panel discussion at the Indian Health Industry Forum, a panel member, who was an information privacy expert, discussed that an individual might access to the Internet and search for information related to cancer symptoms. These online activities of this person might trigger data analytic algorithms at health insurance companies, such that they might put this person into a high-risk category.

Another issue in data-driven businesses was discrimination in term of products, services, and prices. Private companies used consumer data to build individual profiles. These individual profiles allowed companies to target individuals differently and equally with the same products or services. In collected documents of a Congressional hearing about the digital advertising ecosystem, a U.S. representative stated that “targeted ads can also be tools for discrimination” (U.S. House of Representatives, Subcommittee on Digital Commerce and Consumer Protection of the Committee on Energy and Commerce, 2018, p. 4). The U.S. representative continued presenting that “we have also seen ads for junk financial products that are directed to communities of color” (U.S. House of Representatives, Subcommittee on Digital Commerce and Consumer Protection of the Committee on Energy and Commerce, 2018, p. 4). Another privacy issue was the concern about the practice of sharing private personal information in data-driven businesses. In collected documents of a Congressional hearing about digital advertising ecosystem, a representative stated that “when they’re combined together, they create a vivid mosaic of both our online and offline who we are, and we don’t know who that’s being shared with” (U.S. House of Representatives, Subcommittee on Digital Commerce and Consumer Protection of the Committee on Energy and Commerce, 2018, p. 76).

When companies used data to build individual profiles and offer products based on these profiles, consumers might have received different information, services, or prices. This could lead to price discrimination and unequal access to information. For instance, data provider companies group consumers based on their financial status. In the same hearing about digital advertising ecosystem, a U.S. representative argued that

“targeted ads can result in blatant discrimination. It’s been well documented than targeted advertising systems have allowed housing ads to exclude people of color and job ads to exclude older workers” (U.S. House of Representatives, Subcommittee on Digital Commerce and Consumer Protection of the Committee on Energy and Commerce, 2018, p. 6). Understanding financial status might help businesses engage consumers effectively, providing products and services which meet consumer preferences. At the same time, ethical issues might occur when businesses use financial data to take advantage of consumers who are in difficult financial circumstances. For instance, credit institutions could use financial data to find financially vulnerable consumers and offer high interest rate credit cards to them. These business practices need to be explored and scrutinized.

Other issues related to individual profiles and scoring systems were that companies sent advertising products about financial services to vulnerable consumers who currently had financial difficulties. Individuals in this segment might have to accept unfair terms compared to other segments who had stable financial status. Similarly, food manufacturers might offer low level of sugar products to individual segments who had symptoms of diabetes based on their profiles. At the same time, health insurance companies might classify these groups of individuals into high risk categories.

Archival documents revealed that companies use data to build scoring products and sell them to other companies to assess potential customers or to evaluate financial applications. Under the Fair Credit Reporting Act, companies cannot use consumer credit information or Fair Isaac Corporation (FICO) scores for marketing purposes. However, data provider companies gathered consumer credit information and credit scoring systems

to develop summarized credit scoring products similar to FICO scores. They then sold these credit scores to other companies for marketing purposes.

Privacy rights are one of the major issues in the business practices of collecting and using consumer data. Collecting and using consumer data for marketing purposes were not new strategies for private companies. The difference is that companies might use advanced technologies to collect, aggregate, and analyze massive amounts of digital data over the Internet environment. Companies aggregated data from multiple sources. Each data source provided bits of personal information associated with an individual. Advanced technologies such as big data, data mining, and advanced data analytic algorithms provided private companies the capability to connect associated data together and built complete individual profiles. Aggregating data from multiple sources might expose private and sensitive information of consumers that was not supposed to be disclosed in public. Archived documents from one company in this study stated that the company kept consumer data in their database systems for business purposes, even if consumers did have any business relationship with it.

Privacy documents of a company in this study stated that they did not reveal identifiable information of minors, who were under 18 years old, to their affiliates for advertising. However, they used the minor's data for non-marketing purposes, including identity verification, opt-out requests, or fraud detection. They also linked minor data with adult data, such as parent records, to create rich adult profiles.

**Finding 12: Data Protection Regulations**

There is no comprehensive federal data protection regulation in the United States that gives consumers the right to control and protect their personal data from being collected by private companies. In collected documents of a Congressional hearing on data security before the Committee on Financial Services, a witness testified that “there are thousands of underregulated and unregulated data brokers out there” (U.S. House of Representatives, Subcommittee on Financial Institutions and Consumer Credit of the Committee on Financial Services, 2017, p. 7). The witness explained that “they sell numerous consumer profiles to businesses. Consumers have no rights to know about these files, to examine these files, to correct these files or to limit their use” (U.S. House of Representatives, Subcommittee on Financial Institutions and Consumer Credit of the Committee on Financial Services, 2017, p. 65). There are different laws for regulating consumer data processing in different businesses, sectors, and industries. For example, the privacy policy of one company in this study states that when the company processed financial data or medical records for their clients, it needed to comply with either the Fair Credit Reporting Act or the Health Insurance Portability and Accountability Act.

The policy documents of the 12 companies in this study revealed that it established business practices of processing data compliance with current data protection laws and regulations. Each data company had its own privacy policies to control and manage consumer data. Businesses in the data-driven economy promote self-regulation policies to control and protect consumer data. Companies might adopt general guidelines for ethical business best practice, developed by Direct Marketing Association, to handle

sensitive information. In collected transcripts of a hearing on data security, a witness testified that “with the lack of control, it is very difficult for consumers to do anything about misuse of their information” (U.S. House of Representatives, Subcommittee on Financial Institutions and Consumer Credit of the Committee on Financial Services, 2017, p. 7). The witness emphasized that “we have very little authority to vote, to determine that companies can’t use our information, very limited under Gramm-Leach-Bliley. In most cases, companies simply collect information about us and sell it” (U.S. House of Representatives, Subcommittee on Financial Institutions and Consumer Credit of the Committee on Financial Services, 2017, p. 7).

Under the California Consumer Privacy Act, companies need to provide opt-out option to consumers who do not want companies to sell or exchange their data. The 12 companies in this study provided consumers the option to opt out of data collection and data use for business purposes. Companies offered these opt-out options via their privacy statements or policies. However, when consumers chose opt-out options, companies might not remove consumer information completely from their databases. According to privacy policy documents, they just marked individual records to indicate that the records should not be shared with other companies. Companies continued to maintain certain personal information in database systems for identity verification and other purposes. The issue is that consumers may not know that companies still had their personal data. Consumers may not know how much of their personal data or what types of information companies remove when they choose opt-out options. Although privacy policy documents of the 12 companies in this study revealed that they did not sell



consumer data to certain types of businesses, there are no comprehensive privacy regulations that control and enforce how private companies use consumer data in their businesses. Practices of collecting, using, and selling consumer data within data-driven businesses remained unclear to public views.

### **Summary**

In this chapter, I presented several key components of the data analysis process through which I drew the findings and answered the research question. I discussed the research setting, demographics, and characteristics of the collected archival materials. I used a thematic data analysis method to code, categorize, and develop themes from collected archival materials. The following findings were derived from common themes that emerged from the collected archival documents. The key findings included (a) potential benefits, (b) data sources from multiple sources, (c) identity information, (d) the transparency in data-driven businesses, (e) data products and services, (f) affiliates and partner networks, (g) the purposes of using consumer data, (h) the complexity of the data-driven industry, (i) ownership and consent, (j) access and control over personal information, (k) risks and ethics of using personal data, and (l) data protection regulations. The findings revealed many aspects of the business practices of collecting and using consumer data in the data-driven economy. In the next chapter, I discuss and interpret the findings of this study that contribute new knowledge to the existing literature. I also provide recommendations for future research in broader study domain. Finally, I discuss research limitations and present research implications that may affect positive social change.

## Chapter 5: Discussion, Conclusions, and Recommendations

The collection and use of consumer data for business purposes was not new to private companies. Issues involving collecting and using consumer data emerged when new technologies such as big data, IoT devices, Internet, and social media networks allowed companies to collect large volumes of personal data to target or influence individuals in different ways. Consumer data have become valuable assets in the data-driven economy. Data companies have transformed consumer data into strategic products and services that can be sold and exchanged for benefits. Using consumer data has many potential benefits for businesses. Aggregating data to create individual profiles or segregate consumers into groups might put consumer privacy at risk in terms of misuse, abuse, discrimination, and unauthorized access.

The purpose of this qualitative archival research is to explore private company practices of collection and use of consumer data without an individual's consent. The results of analyzing archival documents gathered from different sources provided insights into various aspects of this emerging phenomenon. The key findings included (a) potential benefits, (b) data sources from multiple sources, (c) identity information, (d) the transparency in data-driven businesses, (e) data products and services, (f) affiliates and partner networks, (g) purposes of using consumer data, (h) complexity of the data-driven industry, (i) ownership and consent, (j) access and control over personal information, (k) risks and ethics of using personal data, and (l) data protection regulations.

### **Interpretation of Findings**

Innovative technologies such as social media networks, smartphones, wearable devices, and video streaming services offer individuals different ways to communicate with friends, share pictures with family members, post product reviews, purchase goods, and search for information (Zaki, 2019). These technologies also create new business models that rely on consumer data, allowing both businesses and consumers to access information quickly. Finding 1 of this study shows that companies collected consumer data in order to understand consumer preferences. Consumer data allowed companies to improve the quality of existing products and recommend new products to potential customers. Data-driven companies also applied sophisticated computer algorithms to generate predictive models and scoring systems to forecast business trends or predict consumer behaviors. Marketers, retailers, car dealers, and real estate agencies could buy these data products to use for many business purposes.

Marketing was one of the focused purposes for using predictive data products (Tran, 2017). Finding 2 and finding 5 of this study show that companies monitored online and offline activities of consumers to understand their characteristics, behaviors, and lifestyles. This vital information provided companies with the capability to predict consumer preferences and target them at the right time. Understanding consumer needs and recommending relevant products to consumers might improve marketing strategies and reduce advertising costs. Using historical data to understand business trends also helped companies develop new products and expand businesses into new markets.

Finding 2 shows that companies obtained the personal data of consumers from public records, local governments, private companies, online activities, and social media networks. They used these data elements to build individual profiles, segregate consumers into groups, or generate predictive scoring systems. These data products offered businesses a complete understanding of consumers so that they could adapt to the current market and stay competitive. Finding 7 shows that having customer profiles allowed companies to customize products and target customers with products that met personal requirements. Internal business operation data and external information can help business leaders make quick decisions to respond to current market demands. Essential information, such as inventory, sales data, customer reviews, and product ratings, help businesses address the need for products and markets at specific locations.

Although companies collected different types of personal data from multiple sources, finding 11 shows that businesses might miss data from certain groups of consumers that had limited access to the Internet or lacked technology. The collected data sets that companies had in their databases might not represent the entire population in particular areas. Under certain privacy laws, companies also needed to remove some data elements from their data products, such as information on underage individuals, gender identity, or race. Finding 11 shows that when third-party companies relied exclusively on incomplete data sets to segregate consumers and predict their behaviors, missing information might lead to biases in predictive models or scoring systems. Companies used data analyzing algorithms to generate predictive data products automatically, and

did not always consider the quality and accuracy of the collected data in their database systems.

Since third party companies such as data providers, advertisers, marketing agencies, and financial institutions do not do business directly with consumers, consumers might not know these third parties' practices involving developing individual profiles, segregating consumers into groups, and building scoring systems. Finding 4 shows that data companies often did not disclose data sources or partner networks that bought data products from them. Their practices involving analyzing, using, and selling were hidden from public view. Privacy documents from the 12 companies in this study showed they collected data from third parties, but did not reveal the names of these companies. The complexity of exchanging data among data-driven companies made it difficult for consumers to trace what companies had their data.

According to Altman et al. (2018), consumers often needed to accept the terms of use provided by companies when consumers visited business websites, downloaded mobile applications, watched videos online, or opened social media network accounts. Through terms of use notification, consumers might know that companies collected and used their data. However, finding 11 shows that companies often did not reveal in detail how they were going to use data in the process of generating predictive models or scoring systems. Companies also did not disclose their clients and affiliates who bought these data products.

Privacy documents collected from the 12 companies in this study showed generally how those companies were going to use consumer data. For example, they were

going to use consumer data to improve their products, customer engagement, personalized advertising, business analytic measurements, identity verification, and fraud detection. These general statements left consumers out of data processes in which companies used advanced computing technologies to develop statistical scoring systems or predictive models. Consumers might not be aware of how their personal information is fed into complex computing algorithms or how their personal data contributes to scoring systems. Finding 6 shows that consumers might not know which third party companies buy these data products or how third parties might use data products to make decisions on them. As presented in finding 8, consumer data life cycle in the data-driven economy was complex and abstracted at a high level. Finding 4 and finding 9 in this study show that companies did not provide adequate information to consumers. Companies also did not involve consumers in the process of handling personal data. General statements in privacy policy documents might mislead consumers about their privacy being protected. Companies regularly monitor consumer activities and exploit their privacy for business benefit. Finding 10 shows that consumers might not be able to control their personal data from being collected and shared by data companies once consumers accept terms of use for free-of-charge products or services.

New technologies, including the Internet, smartphones, wearable devices, and IoT devices provide massive personal data to private companies. With large volumes of data from these devices, companies can easily track down individual identities and locations. All 12 companies in this study could verify individual identities based on collected PII data in their databases. Finding 3 shows that companies could use different technologies

and personal data, such as cookies technology, IP address, GPS data, device information, search keywords, and visited webpages to track consumers' activities regardless of where they are. Especially Internet-connected devices such as smartphones and tablets provide rich location information that companies can use to track individuals. Companies could aggregate location data from mobile devices such as GPS, cellular tower trilateration, or wireless network routers to determine the precise locations of online users. Finding 9 shows that companies could collect location data through mobile applications and sell individual profiles or predictive models at particular cities or zip codes. Once consumers allow applications to access location data on their mobile devices, companies can use it for many business purposes or sell it to third parties. Finding 9 presents concerns that consumers might not be aware of companies constantly tracking their location in background processes. Purposes of using location information were also not transparent.

Consumers have used different Internet-connected devices and applications to communicate with individuals and businesses in the Internet environment. With the help of new technologies, companies can track movements of individuals across devices and applications. For instance, companies could provide consumers one email account to access multiple services, such as email, social networks, online storages, streaming videos, online shopping, and many other online services. Consumers could also use one email account to access applications on different Internet-connected devices, including smartphones, wearable devices, or smart televisions. Furthermore, consumers could use an online account access service, offered by various companies. The cross-authentication technology allowed companies to link consumer online activities to one personal profile.

As finding 9 presents, the ability to track individuals across devices, applications, or business services provided companies rich personal information to target potential consumers precisely.

Vast data sources, data elements, and sophisticated data analysis algorithms allow companies to quickly identify an individual and query associated personal activities, behaviors, or lifestyles. Privacy documents from the 12 companies in this study revealed that they aggregated identity data elements, such as name, residence address, phone number, social security number, date of birth, driver's license number, and many other personal data elements, to verify an individual's identity. They also combined data from different sources to build individual profiles across Internet-connected platforms and devices. With the advance of modern computer technologies and database systems, companies could easily update large volumes of digital data and exchange it with their affiliates and partners.

The business operations of data companies are complex and hidden from consumer view. Finding 6 shows that data companies have businesses in different industries, including finance, retail, technology, marketing, and more. They collected, sold, and exchanged data from third parties and other data providers. In addition to data provider companies focusing on collecting and using consumer data, other companies in different sectors also relied on consumer profiles, predictive models, tracking data, or scoring systems. They used these consumer data products to forecast business trends, learn about consumer preferences, or improve internal business operations. The practices of collecting and using consumer data have been expanding rapidly in multiple layers of



the data-driven economy. Various layers of the data-driven economy may include data sources, data providers, data collection technologies, data analytics algorithms, data productions, and data production buyers. The new CCPA regulation requires data providers that operate in the State of California to register with the State Attorney General office. There were an estimated 140 data provider companies registered to the State Attorney General office (State of California Attorney General, 2020). These companies continuously collected and tracked consumer activities.

In the data-driven economy, marketing was one of the primary commercial purposes driving companies in various sectors to collect consumer data. Companies operating in technology, social media networks, finance, retail, mortgage, or business consulting incorporate different ways to collect consumer data to advance their businesses. Finding 6 shows that affiliates and clients bought data products from data provider companies to target potential consumers and expand to new markets. Clients of data companies also bought scoring systems and predictive models to better understand business trends and customer preferences. Individual profiles, scoring systems, or predictive models allowed companies to customize advertising, recommend relevant products, offer promotions, or improve the customer experience.

Collected documents from the 12 companies revealed a wide range of data products and predictive scoring systems. Data companies offered data solutions for businesses in different sectors, including finance, retail, real estate, travel, healthcare, and many other business sectors. Finding 5 shows that data companies segregated consumers into groups based on shared characteristics, such as financial status, ethnicity, income,

religion, behavior, education, or lifestyle. They also built sophisticated scoring systems to evaluate consumers and developed predictive models to forecast different aspects of consumer personality. Scoring systems provided predictive measurements concerning consumers' sensitive information. Predictive measurements pertained to health conditions, household income, spending styles, payment history, marital status, and many other sensitive information. Besides scoring systems, data companies also offer various predictive analysis products in different industries. Based on massive amounts of digital data and sophisticated computer algorithms, companies could track and monitor consumer movements from both online and offline activities, and from mobile devices to laptops or smart televisions. Using modern technologies to capture and analyze personal data elements allowed companies to understand consumers better and effectively target them with relevant products. Simultaneously, the quantity and accuracy of predictive scoring systems might expose sensitive information that an individual did not release to the public. Organizational leaders, government authorities, and lawmakers should consider and address privacy concerns regarding these data products.

Finding 11 shows that potential biases existed in the processes of defining target customers, aggregating the personal data of consumers, and using computer algorithms to generate predictive scores about them. Data catalogs of the 12 companies in this study show that they have predictive data products for a wide range of business sectors. Other businesses could buy these predictive models to forecast business trends, consumer preferences, individual health conditions, or the financial status of a household. Stewart (2019) noted that a potential issue of predictive models was that companies might fail to

implement error-free computer algorithms to analyze and generate predictive models. Companies also did not disclose these computer algorithms to outsiders for evaluation and assessment. New technologies, such as big data and IoT devices, provided companies the ability to gather large volumes of digital data from multiple sources. According to Stewart (2019), adding more data elements to computer algorithms might not increase the accuracy of predictive models. The large volumes of input data sets and complex algorithms made it difficult to explain the accuracy of predictive data products or scoring systems. According to West (2019), discrimination may happen when companies use predictive scores to target specific consumer groups. Finding 11 show that predictive model products and scoring systems might affect consumers differently in certain circumstances, since marketing agencies purposely used background information to target them.

The benefits of using consumer data to predict consumer experiences should not be omitted. For instance, using financial data, such as credit transactions, incomes, or payment history, might help banking institutes introduce new financial products to customers. In healthcare prevention, using wearable devices to monitor heart rate, blood pressure, cholesterol level, or blood sugar level might help consumers to avoid deadly diseases (Sudtasan & Mitomo, 2018). Although these implications may benefit consumers, collecting sensitive data elements to develop individual profiles and using these profiles for unclear purposes might outweigh the benefits. Finding 5 and finding 11 show that consumers might not know exactly how their personal data contribute to individual profiles or how third parties handle and use predictive scoring systems.

Collected data catalog documents from the 12 companies showed that they developed a wide range of scoring systems related to consumer behaviors, financial status, or health conditions. Companies assigned numbers or scores to individuals so that they could prioritize, rank, or categorize individuals into groups that had common characteristics or behaviors. Consumer scoring systems easily allowed businesses to find potential customers and target them based on behaviors, habits, financial status, or health conditions. For example, current FICO credit scores are common consumer scoring systems that financial institutions use to evaluate consumers' financial status. Based on FICO credit scores, lenders were able to assess the risks of loan applications and make appropriate approval decisions. Similarly, data companies used proprietary computer algorithms to generate scoring products based on various personal data elements.

Collected data catalog documents from the 12 companies showed that they had a broad range of consumer scoring products pertaining to finance, spending, income, credit, fraud detection, real estate, or investment style. These consumer scores were generated based on various data elements, containing PII, including information about former employers, income, social media networks, banks, insurance, loans, properties, and many other personal data elements.

Finding 5 and finding 6 show that data providers offered advertising solutions for businesses to target potential customers based on various personal data elements, categories, and characteristics. Marketing agencies and retailers use consumer data to display advertising products on e-commerce websites or mobile applications. In addition, they also use consumer data to automatically adjust website contents so that

recommended products could meet consumer preferences and needs. For example, a political campaign used individual personalities to send personalized advertising messages to a particular group of voters. Finding 11 shows that discrimination might happen in consumer scores or predictive models. For instance, online retailers or travel agencies might offer different price or promotion for a particular product to an individual based on predictive models about behaviors and the financial score of that person. Personalized advertising based on consumer profiles, predictive models, and real-time online user activities offered companies the ability to influence potential customers through dynamic web content, recommended products, and simple purchasing processes.

Finding 11 shows that data breach was one of the main concerns in the information industry. For example, in a Congressional hearing about consumer data security, a witness presented on a data breach incident that happened at a popular financial institution in the United States. This company exposed millions of personal records of consumers in multiple countries, including the United States, Canada, and the United Kingdom. In this data breach incident, compromised personal records contained accurate personal information, such as names, addresses, driver's license, date of birth, or social security number. Leaked data also contained other information that consumers used to open bank accounts, apply for loans, or buy cars. This data breach incident created huge damages for consumers. It may take years for victims to resolve identity theft. The financial institution in this data leak incident collected consumer data without consumer knowledge. This company failed to protect sensitive data by not implementing up-to-date security technologies in their IT systems. The company had known security

issues in their IT systems, but they did not fix it until outsiders accessed and retrieved millions of personal records. This data breach incident demonstrated the risk of identity theft in practices of collecting and using consumer data.

Data-driven companies use consumer data for many business purposes. Privacy documents of the 12 companies in this study showed that they used consumer data for marketing, advertising, optimizing business operations, and quality control. The company also incorporated consumer data for credit reporting, fraud detection, identity verification, sharing data with third parties, and many other commercial purposes. The privacy concern was that companies might use personal data for purposes different from what they intended. For example, companies built financial scoring systems based on records, such as names, date of birth, social security number, payment transactions, purchasing records, numbers of credit accounts, or income. They then used these financial scoring systems not only for fraud detection, loan processing, or identity verification, but also for advertising, product recommendation, targeting potential markets, or customer engagement. Companies combined various data and inserted them into complex computer algorithms to generate new types of data products and solutions. There was no restriction on which data were used to develop new data products. For example, one of the 12 companies in this study is a financial data company. It collected consumer data from employers, banking institutions, consumer credit report agencies, government agencies, public records, and third parties, as well as directly from consumers. This company aggregated these data sources and created a broad range of new data solutions for marketing, identify verification, credit reporting, or fraud

detection. A number of its predictive products offered critical information about total assets of consumers, household spending, income, investment style, available credit limits, or current account balance.

Companies should consider issues regarding bias, discrimination, transparency, and privacy that might exist in business practices of collecting and using consumer data. In the digital environment, innovative technologies, such as big data, advanced data analytics, machine learning, and artificial intelligence, provided companies powerful tools to aggregate, store, and analyze large volumes of consumer data. Companies should balance the value that data products provide with business ethics in considering individual privacy and fairness. For instance, consumer predictive models might offer potential benefits to companies. Business leaders can make decisions based on real-time data, so that companies can quickly respond to business trends and consumer preferences. Finding 11 shows that individuals might lose opportunities, such as gaining employment, receiving discounts for health insurance premiums, accessing financial credit, or getting low prices for certain products, when companies evaluate consumers based on incomplete and inaccurate data. The concern is that consumers might not be aware of the complexity of the data life cycle happening behind the scenes, so that they can take appropriate actions to fix errors.

The collected policy documents from the 12 companies in this study showed that the companies provide opting-out and deletion options to California residents. However, consumers must submit requests to each company that they do not want to collect data. Finding 4 showed that consumers might not know which companies collected and used

their personal data. It may be impossible for an individual to contact every company operating in the information industry.

Technologies in data management and analytics have been changing rapidly in recent years. Big data, computer artificial intelligence, predictive analytics, and machine learning were a few examples of emerging technologies that provided companies the capability to collect and analyze large volumes of digital data. In the United States, data protection laws applied to a specific industry, such as banking, education, healthcare, or communication. Finding 12 showed that there is no comprehensive federal legislation regulating personal data being collected and used by private companies. There were also no federal regulations that govern data products, such as predictive models, scoring systems, or consumer profiles, offered by data providers. U.S. legislation includes a variety of separate laws that govern consumer data differently depending on the business industry. For example, the Gramm-Leach-Bliley Act and Fair Credit Reporting Act regulate personal data in the banking and financial sector. The Children's Online Privacy Protection Act protects children's information from being collected and used by companies. The Health Insurance Portability and Accountability Act protects the health records of patients in healthcare providers. These regulations for practices of collecting and using consumer data are applied differently in different business industries. There were no unified regulations at the federal level that cover data privacy across industries.

Finding 10 showed that consumers did not have the right to control their personal data under current federal laws. Depending on where consumers live, consumers could not access, update, or remove their data from companies. Consumers could submit



opting-out requests to ask companies to stop sharing their data with other companies. The opting-out request might be limited by current laws depending on the industry in which companies have businesses. California residents have more control over their personal data under new CCPA statutes. Under this new data protection regulation, California residents can submit requests to view their personal data being collected by companies. They can also request to delete their data from company databases.

Individuals might expect that their personal data are protected if they do not share them with any entities. For example, if they do not use PII to open a bank account, apply for loan applications, make reservations for vacation, or create email or social media network accounts. However, this might not be practical in modern society. Finding 10 showed that consumers might lose control over their personal data as soon as they decide to share information with companies. Especially in the digital environment, digital data can be transmitted and exchanged among businesses easily. U.S. data privacy legislation needs to be updated to catch up with the complexity of processing data within data-driven companies. Consumers need strong data privacy legislation to secure data privacy and increase trust in the digital economy.

### **Limitations of the Study**

I used archival documents as primary data sources in this study. Collected documents were available on public websites of both government agencies and private companies. I could not control the accessibility, volume, and content of archival materials. I also could not define in advance the availability of archival collections on digital repositories. Although I could find document titles while performing a search, I

could not always retrieve the contents of the documents in the search result. This study was limited to 12 companies to demonstrate the current business practices of collecting and using consumer data among data companies. Although these 12 companies were in different sectors, such as technology, social media networks, finance, and data services, they might not represent the entire data-driven network. The data-driven industry is complex and dynamic, involving multiple layers from data sources and data providers to clients, partner networks, and other third-party data providers. The findings show that companies operating in the data-driven industry offered various consumer data solutions for different industries. Examining the impact of consumer scores and predictive statistic models on certain populations was beyond the scope of this study.

Due to the comprehensive nature and complexity of legal systems in the United States, this study did not include a thorough analysis of current data privacy laws in the United States. Analyzing the coverage of every data protection law in different industries was outside the scope of this study. This study reviewed several data protection statutes that regulate business operations in different sectors, to demonstrate the approach of protecting consumer privacy in the United States. The purpose of this study is to explore private company practices of collecting and using consumer data without an individual's consent. I could not promote data protection solutions to a particular business or offer new data privacy policies for the data-driven industry. Instead, I provided a list of recommendations as a foundation for future research to expand knowledge in the data protection and information management field.

## Recommendations

This section summarizes recommendations for future studies that extends knowledge gaps in the information management field. These recommendations are drawn based on fundamental concepts from the literature review and the findings sections of this study. Researchers may find these recommendations helpful as foundations for conducting further work in data protection, privacy policies, and consumer rights.

**Finding 1: Potential benefits.** Consumer data provided massive volumes of information to companies, so that companies could understand in detail individual behaviors, preferences, or lifestyles (Yeh, 2018). Consumers could use a wide range of applications, platforms, or services on their smartphones, tablets, or laptops at no cost. Consumers could search for information, share pictures, stream videos, buy goods, make payments, and use other products and services over the Internet. Companies could offer merchandise to customers, recommend new movies based on viewing history, or introduce new credit lines based on financial conditions. IoT devices could generate a vast amount of sensor data on consumer health conditions and daily activities. Aggregating these data could provide companies with complete and sensitive personal profiles for an individual. By using consumer data to provide convenient products and improve business operations, these sensitive personal data might cause harm consumers. The harm may include cases of misuse, unauthorized access, or discrimination. Future studies should focus on developing secured data management frameworks to protect consumer data from third parties' unauthorized access. Companies should also establish

simple terms of use statements that allow consumers to understand the terms and conditions of online products that require consumers to share their personal data.

**Finding 2: Data from multiple sources.** The collected archival documents showed that consumers used various Internet-connected devices, such as mobile phones, computers, or tablets, to search for information, make purchases, communicate, stream movies, or make billing payments on the Internet environment. Social media networks, such as Facebook, Twitter, Instagram, or WhatsApp, offer various features that allow online users to share personal data with friends and relatives. Internet search websites, including Google and Bing, enable users to search and access massive amounts of data in the Internet environment. Online retailers provide their customers with the ability to post product reviews and information about their experiences. These examples of online activities generate huge volumes of digital data that might be collected, stored, used, and sold by private companies.

Consumers may need to know that companies collect their online activities and build individual profiles, predictive models, or scoring systems that present their characteristics, lifestyles, and behaviors. With the advance and innovation of technologies in data management and analytic fields, companies were able to collect large volumes of digital data at low overhead costs. Simultaneously, the availability of consumer data has been increasing rapidly through Internet-connected devices, creating massive digital data collections about consumers. Companies should provide data protection settings and support material associated with any Internet-connected platform and device. These privacy features should allow consumers to set up data privacy settings

in order to control sharing data. Studying communication and education about technology and data protection is necessary in future work. Companies should provide appropriate information to communities about how they collect and use personal data, such that consumers have the basic knowledge to protect their data.

**Finding 3: Identity information.** Aggregating various personal data elements to build profiles provides companies with the capability to identify an individual.

Technologies, such as big data, data mining, and computer artificial intelligence, provide tools to verify identity information and retrieve information related to that person.

Exposing identity information to unauthorized access may put individuals at risk of misuse by third parties. For example, when criminals access PII data, such as name, address, phone number, date of birth, or social security number, they can use these PII to open bank accounts, request new credit cards, or apply for loans. Because the criminal activities may create severe damage to the financial reputations of the identity theft victims, companies have increased data protection by separating identity data from individual profiles within database systems.

Companies can use encrypted technologies to prevent re-identification or the re-association of identity data with other personal records. Although separating identity information from personal records could reduce the risk of exposing individual identity to unauthorized access, data-driven companies eventually find new ways to link data together. The key is that companies should use any technologies to implement secured database systems and prevent unauthorized access to identity information. They also should not sell identity information to their affiliates or third-party companies. Finally,

data-driven companies should publicly commit to not exchanging identity information and to protecting PII data. Future studies should find new ways to de-identify personal data from individual profiles and make personal profiles anonymous to third parties.

**Finding 4: Transparency in data-driven businesses.** Data catalog and privacy policy documents from the 12 companies showed that they collected hundreds of consumer data elements from local governments, public records, businesses, data providers, or consumers. Companies had information about various personal behaviors and characteristics of U.S. consumers. In the data-driven ecosystem, business practices of obtaining, using, or selling consumers happened without consumer knowledge. Consumers might know that businesses collected their data; meanwhile, collection processes and the life cycle of data exchange existing in the data-driven ecosystem might be concealed from public views.

Under new privacy laws, such as the GDPR in the European Union and the CCPA in California, companies were required to disclose where they collected data, what information they had, how they stored and used it, and whom they sold it to. While developing transparent data processing is necessary, companies can take advantage of data while ensuring that consumers understand what happens to their data. Increasing transparency in how companies use consumer data could help enhance the relationship between consumers and businesses. Companies could also gain consumers' trust when they have better understanding and control over their personal data. Future studies should examine the data collection process to ensure that it gives consumers options and choices

to control their data. Companies should also notify consumers before selling their data to third-party companies.

**Finding 5: Data products and services.** With the increased availability of digital data in the digital environment and the advance of computer algorithms, companies could aggregate consumer data from multiple sources and generate various scoring systems and predictive models about consumer characteristics and behaviors. Bias and discrimination might exist in these predictive models. It might result in consumers losing opportunities to access services. Incomplete and inaccurate information might create errors in the process of analyzing and generating consumer data products. Companies should disclose processes of building individual profiles, including computer algorithms that were used to produce scoring systems and predictive models. Future work should examine the impact of inaccuracy, bias, and discrimination in consumer data products. Future studies should focus on assessing privacy, bias, and discrimination concerns in consumer data products, ensuring that sensitive information, such as information on health conditions or financial status, are used appropriately in consumer data products.

**Finding 6: Affiliates and partner networks.** Collected data catalog documents from the 12 companies showed that they provided consumer data products and solutions for a variety of companies operating in different sectors. Their affiliates and partner networks included private companies, non-profit organizations, marketers, political groups, charities, and fundraising organizations. In addition, these data-driven networks also had offshore organizations operating in foreign countries. These entities operated in hotel chains, telecommunications, banking, healthcare, insurance, and advertising

agencies, as well as with other data providers. Data solutions offered by data providers provided strategic business information that allowed other companies to retain existing customers, forecast new potential markets, detect financial fraud, verify individual identity, and perform many other business activities. Although collected documents from the 12 companies showed partnering network industries, they did not disclose specific information about which companies bought their data solutions or which third parties they exchanged data with. Introducing new unified data protection policies is necessary for future work, in order to provide companies guidelines for releasing their clients and affiliates to the public. This information might help consumers know which organizations have their personal data and increase the transparency of the data life cycle.

**Finding 7: Purposes of using consumer data.** Collected privacy policy and data catalog documents from the 12 companies in this study showed that companies used consumer data for a broad range of business purposes. For internal purposes, companies used consumer data to improve business operations, quality control, or customer experience. For external purposes, companies used consumer data to build individual profiles, scoring systems, or predictive models. These consumer data products were sold to other businesses for marketing, personalized advertising, retaining loyal customers, fraud detection, business research, and many other business purposes. Future studies should focus on ensuring that companies notify consumers in advance of purposes for which they will use their personal data and ensuring that they only use the data for these intended purposes.



**Finding 8: Complexity of the data-driven industry.** Consumers might know their personal data are being collected by companies when consumers used certain online applications or services. Due to the complexity of the practices of collecting, analyzing, and selling among companies, consumers may not know precisely how companies used their personal data in the data life cycle ecosystem. The data-driven industry built multiple complex and dynamic layers of data processing networks from data sources and proprietary management systems for data products and affiliate networks. The data-driven industry also involved an array of organizations in both public and private sectors.

Documents collected from the State of California General Attorney office show that there were more than a hundred data providers operating in the State of California alone. Consumer activities, online and offline, were monitored continuously and aggregated by networks of data providers operating in multiple industries. Data companies constructed data life cycle networks that were hidden from public view. Future studies should focus on initiating the transparency of the data-driven industry, promoting open data exchange policies, and giving consumers more information about both data sources and data buyers of data-driven companies.

**Finding 9. Ownership and consent.** Even before U.S. lawmakers introduce new data privacy legislation, companies can establish privacy policies that involve consumers in the process of collecting and use personal data. Companies should allow consumers to have the ability to control their data and to decide how much data they want to share, what companies can obtain or buy their data, and for what purposes companies can use their data. New data privacy laws, such as the GDPR in the European Union or CCPA in

the State of California, give consumers the right to access, opt out, or delete their personal data from data companies. Companies should notify consumers and obtain consent from them whenever companies collect their data or share it with third-party entities. Data-provider companies are not only required to obtain individual consent, but any third-party entities involving in consumer data life cycle should get permission from consumers. Future work should focus on developing consent management frameworks and policies that help companies comply with newly initiated data privacy laws. New consent management frameworks should provide tools and fundamental guidelines for companies to access, use, and share personal data while maintaining consumer trust in businesses for protecting their data. This protocol can create transparency in notifying consumers when their personal data is being collected and used by companies.

**Finding 10. Access and control over personal information.** Regardless of where consumers live, companies should provide consumers the ability to access, view, delete, and make corrections when their own data are not accurate. With more transparency in data management processes, consumers could better understand the processes and make appropriate decisions to their personal data. Companies should include consumers in data collection and management processes. Consumers need to have options to decide what information they want to share and what companies they want to share their data with. Consumers also need to have opportunities to remove their personal data from collected data sets.

Under current U.S. regulations, companies provided opting-out, accessing, and deletion options to consumers based on their residency. California residents can request

that companies remove their personal data from company databases. However, it is not practical for consumers to contact every company that had their personal data. U.S. lawmakers should introduce new data protection regulations, giving consumers more choices and options for managing their private data. The CCPA in the State of California or the GDPR in the European Union are examples of new data protection regulations. These new data protection laws regulate the practices of collecting and using consumer data, giving consumers the right to control and manage their own data in the data-driven ecosystem.

Documents collected from the 12 companies showed that consumers need to submit opting-out or access requests to each company across multiple industries to demand that they stop using their data. This may be impractical in reality. The U.S. government should provide one centralized website that allows individuals to submit requests for accessing, updating, or deleting their personal data from the database systems of data provider companies. In addition, companies in finance, retail, technology, or social media networks should notify consumers and give them options for opting out of sharing or selling their data to third parties. Future research should focus on implementing new mechanisms or tools that allow consumers to view, update, or remove their personal data at one centralized portal. The data portal should be a one-stop-shop or user-friendly website, such that consumers can control and manage data across company data repositories.

**Finding 11. Risks and ethics of using personal data.** Consumer data become critical assets for data-driven businesses to understand consumer preferences. Companies

continuously find new ways to collect online and offline data of consumers so that they can enrich their data collections. At the same time, data breach incidents frequently happen that expose personal data to unauthorized parties. Identity theft may affect consumers' financial reputations and lead to loan applications being denied, to rejection from employment, or to disqualification from renting apartments. Through data breach incidents, consumers might have a glimpse of their personal data being collected and used by unknown organizations.

Each data-driven company should establish a data product ethical review committee to govern business practices of collecting, using, and selling consumer data among data companies. This ethical board is responsible for monitoring the ethics and compliance of any data products and services based on consumer data. Data companies should enforce privacy regulations in order to ensure that their affiliates, who buy data products, follow them. Companies should encourage software engineers to design and develop ethical algorithms that are used to analyze and generate predictive scoring products at the employee level. Any bias existing in these scoring systems might affect consumers when companies use these scoring systems to evaluate or predict them. Future research should focus on implementing private and fair predictive scoring systems. Future work should also explore the role of ethical review committees in governing ethics and privacy of data products.

**Finding 12: Data protection regulations.** Promoting new data privacy laws at the federal level to protect U.S. consumers in the digital age is necessary for future research. Current separate data privacy policies need to be revised to cover and govern

practices of collecting and using consumer data in the data-driven industry at the federal level. The GDPR from the European Union and CCPA in the State of California are two significant recent changes in data privacy regulations. U.S. companies need to comply with the GDPR law when they collect and process personal data of EU citizens. The State of California introduced the CCPA in 2018, and it took effect in 2020. Under this new data privacy regulation, California residents have the right to request access to, view, and delete their personal information from data provider companies. They also have the right to opt out of data providers that sell their personal data.

Companies have obligations to provide consumers two or more ways to submit the request. This can be a toll-free phone number, email address, or other contact information. Future research should focus on proposing new federal data privacy legislation to allow consumers to control, access, correct, and delete their personal data, regardless of where they live. Adopting principle elements of the GDPR and CCPA legislation would fill the gaps in consumer data privacy protections. Such legislation should allow consumers to make decisions on their personal data, whether they allow companies to collect or sell their data for business purposes.

### **Implications**

The findings of this study have several implications that might contribute to positive social changes. The study revealed different aspects of business practices of collecting and using personal data in the data-driven economy. Consumers should have the right to choose whether they want to share their personal data with companies. They might be aware of their personal data being collected and used by companies. They have

the right to expect that companies protect their personal data from unauthorized access or misuse. Data-driven networks engaged in more complicated activities than just collecting and using data. Data-driven companies not only collected data directly from consumers but also obtained consumer data from local governments, businesses, social media networks, surveys, or other data provider companies. Data companies used proprietary computer algorithms to analyze various data elements and generate scoring systems or predictive models. Businesses in different sectors could buy these consumer scores to predict consumer behaviors or to forecast business trends. Data companies also built personal profiles and clustered consumers into groups of individuals who shared common characteristics.

At the individual level, the study results showed the various personal data elements that companies collected and used in their data products. The findings show that companies also collected other personal information from multiple sources, including information on finance, educational degrees, employment, income, properties, and health records, as well as numerous other pieces of personal information. This finding offers essential information to individuals that companies might collect more personal data than just what consumers share with them.

Individuals should have the right to know how data companies use their personal information, and for what purposes, as well as which companies buy their data. The research findings show many aspects of these business practices and data-driven networks. The awareness of collecting, using, sharing, and selling personal data in the data-driven industry might help individuals make appropriate decisions about their

personal data. Individuals might not lose control over their data when they decide to share it with companies. Particularly, in the digital environment, data can be stored, analyzed, and transmitted over sophisticated computer systems that individuals might not be able to understand completely.

At the societal level, the findings addressed several issues existing in the data-driven ecosystem that may help society be aware of the risks of sharing personal data through Internet-connected applications and devices. Transparency, privacy, data breaches, bias, and discrimination were serious issues that existed in the consumer data life cycle, scoring systems, and predictive models. For instance, the risk of data breach incidents and identity theft not only exposes PII data to unauthorized parties, but might also create damage to consumers for a long period of time.

Companies should provide simple terms of use and privacy policy documents to ensure that consumers can fully understand what happens to their personal data. The findings addressed concerns about privacy policy statements that describe what personal data elements were used in consumer scoring systems and how companies mitigated biases in computer algorithms that were used to generate consumer scores.

Misunderstanding privacy policy documents might lead consumers to share their data with unofficial entities, causing them to mistrust businesses. Providing understandable terms and agreements also minimizes confusion about the purposes of collecting and using personal data among data companies.

At the organizational level, the findings of this study showed concerns regarding consumers' ability to control and manage their personal data within database systems of

companies. Companies may consult the findings of this study to regain consumer trust by establishing new data management strategies that allow consumers to control their personal data regardless of where they live. For instance, companies may find regulations of the CCPA in the State of California as the foundation to make changes in company policy, giving consumers the ability to opt out, review, update, and remove their personal information from collected data sets. Through the discussion of the research findings, company leaders might recognize concerns regarding bias and discrimination that exist in consumer scoring systems and predictive models. Companies may develop transparent guidelines and standard data management processes to demonstrate how personal data were used, and what data security strategies were applied to protect sensitive information. Companies should involve consumers in data management processes, giving consumers choices about sharing their personal data. In addition, this also increases the transparency of business practices of collecting and using consumer data.

This study also contributed to research methodologies in terms of using digital archives that exist in the Internet environment. According to Das, Jain, and Mishra (2018), there were various advantages of using archival materials to conduct management research, social networks, or strategic business management. Especially in the digital age, public and private organizations archived large volumes of business operation data in computer databases. This qualitative archival research demonstrated the advantages of rich digital archives that were available from public websites of government agencies and private companies. The high-quality raw data that I collected to support this study's findings included transcripts of U.S. Congressional hearing, witness statements, federal



agency reports, privacy policy statements, and panel discussions from data privacy professionals. This study might benefit researchers who are interested in using digital archives to conduct data management research, business management strategies, or social studies in the future. Each archival data source offers information through which researchers can study topics of interest from different points of view.

### **Conclusions**

Consumer data have become valuable assets in the data-driven economy. Companies operating in the technology sector, social media networks, retailers, and financial institutions collected and used consumer data to advance their businesses. They established sophisticated data life cycle networks with multiple entities and constantly monitored consumer movements in the digital environment. Data companies combined consumer data from different sources to build individual profiles, categorize people, develop consumer scoring systems, and generate predictive models. These consumer data products and services allowed other businesses in finance, travel, retail, insurance, and other business sectors to take advantage of consumer information to improve business operations, engage potential customers, advertise new products, or detect financial fraud.

The findings show that consumers have limited input into business practices of collecting and using their personal data. Their rights to control their own data are restricted and depend on a wide range of privacy statements provided by companies. Companies did not explicitly disclose the sources from which they collected or bought data. Data provider companies also did not disclose the affiliate companies that buy their data products or services. These business practices of collecting and using consumer data

posed concerns regarding ownership, transparency, and ethics in the data-driven economy.

This study shows various practices of using consumer data products, such as consumer scoring systems, to target vulnerable groups of consumers, particularly when companies made decisions purely based on individual profiles or predictive statistic models. The findings show that errors and biases might exist in these data products. Such errors and biases might lead to discrimination against certain groups of consumers. Companies might leverage rich information from individual profiles to exploit personal behaviors, characteristics, and lifestyles so that businesses could offer them biased and unfair services.

With the advancement of technologies and expanded capacity of data management systems, companies increasingly processed massive volumes of consumer data on the Internet environment. In the United States, there is a lack of federal data protection regulations that give consumers the right to control and protect their personal data from being collected by private companies. Government lawmakers should consider the benefits and potential privacy harms of collecting and using consumer data in the data-driven economy. They should promote comprehensive data privacy laws, such as the GDPR or the California Consumer Privacy Act, to protect consumer data in the digital age.

The findings of this study were derived from archival documents that were collected from both public and private organizations. It was limited by the context of what digital documents were available to search and access from public websites.

Additionally, technologies continue changing and advancing rapidly, which may lead to changes in data-driven business models. Future works may extend the findings and recommendations of this study to gain a better understanding of the dynamic information management field.

## References

- Abbasi, A., Sarker, S., & Chiang, H. L. R. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems, 17*(2), 1-32. doi:10.17705/1jais.00423
- Aimeur, E., Lawani, O., & Dalkir, K. (2016). When changing the look of privacy policies affects user trust: An experimental study. *Computers in Human Behavior, 58*, 368-379. doi:10.1016/j.chb.2015.11.014
- Akhtar, F. M. S. (2018). *Big data architect's handbook*. Birmingham, UK: Packt.
- Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment. *International Journal of Production Economics, 182*, 113-131. doi:10.1016/j.ijpe.2016.08.018
- Altman, M., Wood, A., O'Brien, R. D., & Gasser, U. (2018). Practical approaches to big data privacy over time. *International Data Privacy Law, 8*(1), 29-51. doi:10.1093/idpl/ipx027
- Anney, N. V. (2014). Ensuring the quality of the findings of qualitative research: Looking at trustworthiness criteria. *Journal of Emerging Trends in Educational Research and Policy Studies, 5*(2), 272-281.
- Anthes, G. (2015). Data brokers are watching you. *Communication of the ACM, 58*(1), 28-30. doi:10.1145/2686740
- Atlam, F. H., & Wills, B. G. (2020). IoT security, privacy, safety and ethics. *Internet of Things, 123-149*. doi:10.1007/978-3-030-18732-3\_8

- Baek, M. Y., Kim, E., & Bae, Y. (2014). My privacy is okay, but theirs is endangered: Why comparative optimism matters in online privacy concerns. *Computers in Human Behavior*, *31*, 48-56. doi:10.1016/j.chb.2013.10.010
- Baesens, B., Bapna, R., Marsden, R. J., Vanthienen, J., & Zhao, L. J. (2016). Transformational issues of big data and analytics in networked business. *MIS Quarterly*, *40*(4), 807-818. doi:10.25300/MISQ/2016/40:4.03
- Barnes, M. C., Dang, T. C., Leavitt, K., Guarana, L. C., & Uhlmann, L. E. (2015). Archival data in micro-organizational research: A toolkit for moving to a broader set of topics. *Journal of Management*, *44*(4), 1453-1478. doi:10.1177/0149206315604188
- Bender, L. J., Cyr, B. A., Arbuckle, L., & Ferris, E. L. (2017). Ethics and privacy implications of using the Internet and social media to recruit participants for health research: A privacy-by-design framework for online recruitment. *Journal of Medical Internet Research*, *19*(4), e104. doi:10.2196/jmir.7029
- Bergstrom, A. (2015). Online privacy concerns: A broad approach to understanding the concerns of different groups for different uses. *Computers in Human Behavior*, *53*, 419-426. doi:10.1016/j.chb.2015.07.025
- Birks, F. D., Fernandez, W., Levina, N., & Nasirin, S. (2017). Grounded theory method in information systems research: its nature, diversity and opportunities. *European Journal of Information Systems*, *22*(1), 1-8. doi:10.1057/ejis.2012.48
- Bleier, A., & Eisenbeiss, M. (2015). The importance of trust for personalized online advertising. *Journal of Retailing*, *91*(3), 390-409. doi:10.1016/j.jretai.2015.04.001

- Borgesius, J. Z. F. (2016). Singling out people without knowing their names: Behavioural targeting, pseudonymous data, and the new data protection regulation. *Computer Law & Security Review*, 32(2), 256-271. doi:10.1016/j.clsr.2015.12.013
- Bradlow, E. T., Gangwar, M., Kopalle, P., & Voleti, S. (2017). The role of big data and predictive analytics in retailing. *Journal of Retailing*, 93(1), 79-95. doi:10.1016/j.jretai.2016.12.004
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. doi:10.1191/1478088706qp063oa
- Broeders, D., Schrijvers, E., Sloot, B., Brakel, R., Hoog, J., & Ballin, H. E. (2017). Big data and security policies: Towards a framework for regulating the phases of analytics and use of big data. *Computer Law & Security Review*, 33(3), 309-323. doi:10.1016/j.clsr.2017.03.002
- Butterworth, M. (2018). The ICO and artificial intelligence: The role of fairness in the GDPR framework. *Computer Law & Security Review*, 34, 257-268. doi:10.1016/j.clsr.2018.01.004
- Chua, N. H., Herbland, A., Wong, F. S., & Chang, Y. (2017). Compliance to personal data protection principles: A study of how organizations frame privacy policy notices. *Telematics and Informatics*, 34, 157-170. doi:10.1016/j.tele.2017.01.008
- Colombo, P., & Ferrari, E. (2015). Privacy aware access control for big data: A research roadmap. *Big Data Research*, 2(4), 145-154. doi:10.1016/j.bdr.2015.08.001

- Corti, L. (2004). Archival research. In M. S. Lewis-Beck, A. Bryman, & T. F. Liao (Eds.), *The SAGE Encyclopedia of Social Science Research Methods*. Thousand Oaks, CA: Sage. doi:10.4135/9781412950589
- Cradock, E., Stalla-Bourdillon, S., & Millard, D. (2017). Nobody puts data in a corner: Why a new approach to categorizing personal data is required for the obligation to inform. *Computer Law & Security Review*, 33, 142-158. doi:10.1016/j.clsr.2016.11.005
- Crain, M. (2018). The limits of transparency: Data brokers and commodification. *New Media & Society*, 20(1), 88-104. doi:10.1177/1461444816657096
- Custers, B., Dechesne, F., Sears, M. A., Tani, T., & Hof, S. (2018). A comparison of data protection legislation and policies across the EU. *Computer Law & Security Review*, 34, 234-243. doi:10.2139/ssrn.3091040
- Das, R., Jain, K. K., & Mishra, K. S. (2018). Archival research: A neglected method in organization studies. *Benchmarking: An International Journal*, 25(1), 138-155. doi:10.1108/bij-08-2016-0123
- Department of Commerce Internet Policy Task Force (2010). *Commercial data privacy and innovation in the internet economy: A dynamic policy framework*. Retrieved from [https://www.ntia.doc.gov/files/ntia/publications/iptf\\_privacy\\_greenpaper\\_12162010.pdf](https://www.ntia.doc.gov/files/ntia/publications/iptf_privacy_greenpaper_12162010.pdf)

- Dyke, V. T., Midha, V., & Nemati, H. (2007). The effect of consumer privacy empowerment on trust and privacy concerns in e-commerce. *Electronic Markets*, 17(1), 68-81. doi:10.1080/10196780601136997
- Elder, G. H., Pavalko, E. K., & Clipp, E. C. (1993). *Quantitative Applications in the Social Sciences: Working with archival data*. Thousand Oaks, CA: Sage Publications. doi:10.4135/9781412986519
- Elliott, V. (2018). Thinking about the coding process in qualitative data analysis. *The Qualitative Report*, 23(11), 2850-2861. Retrieved from <https://nsuworks.nova.edu/tqr/vol23/iss11/14>
- Erlingsson, U., Pihur, V., & Korolova, A. (2014). RAPPOR: Randomized aggregatable privacy-preserving ordinal response. *ACM*, 1054-1067. doi:10.1145/2660267.2660348
- Estrada-Jimenez, J., Parra-Arnau, J., Rodriguez-Hoyos, A., & Forne, J. (2017). Online advertising: Analysis of privacy threats and protection approaches. *Computer Communications*, 100, 32-51. doi:10.1016/j.comcom.2016.12.016
- Falkenberg, L. (2010). Case study research in business ethics. In A. J. Mills, G. Durepos, & E. Wiebe (Eds.), *Encyclopedia of Case Study Research* (pp. 97-99). Thousand Oaks, CA: Sage. doi:10.4135/9781412957397.n35
- Federal Trade Commission (2011). Protecting consumer privacy in an era of rapid change: A proposed framework for businesses and policymakers. *Journal of Privacy and Confidentiality*. doi:10.29012/jpc.v3i1.596



- Federal Trade Commission (2013). Mobile privacy disclosures: Building trust through transparency. Retrieved from <https://www.ftc.gov/sites/default/files/documents/reports/mobile-privacy-disclosures-building-trust-through-transparency-federal-trade-commission-staff-report/130201mobileprivacyreport.pdf>
- Feri, F., Giannetti, C., & Jentzsch, N. (2016). Disclosure of personal information under risk of privacy shocks. *Journal of Economic Behavior & Organization*, *123*, 138–148. doi:10.1016/j.jebo.2015.12.001
- Fischer, E., & Parmentier, M. (2010). Doing qualitative research with archival data: Making secondary data a primary resource. In M. C. Campbell, J. Inman, & R. Pieters (Eds.), *NA - Advances in Consumer Research* (pp. 798-799). Duluth, MN: Association for Consumer Research.
- Francis, K., & Taylor, B. (2013). *Qualitative research in the health science*. Abingdon-on-Thames, UK: Routledge.
- Fulgoni, G. (2013). Big data: Friend or foe of digital advertising. *Journal of Advertising Research*, *53*(4), 372-376. doi:10.2501/jar-53-4-372-376
- Fuster, G. G., & Gutwirth, S. (2013). Opening up personal data protection: A conceptual controversy. *Computer Law & Security Review*, *29*, 531-539. doi:10.1016/j.clsr.2013.07.008
- Gaillet, L. L. (2012). (Per)forming archival research methodologies. *College Composition and Communication*, *64*(1), 35-58.

- Gantz, F. J., Reinsel, D., & Rydning, J. (2019). The U.S. datasphere: Consumers flocking to cloud. Retrieved from <https://www.seagate.com/files/www-content/our-story/trends/files/data-age-us-idc.pdf>
- George, G., Haas, R. M., & Pentland, A. (2014). Big data and management. *Academy of Management Journal*, 57(2), 321-326. doi:10.5465/amj.2014.4002
- Ginosar, A., & Ariel, Y. (2017). An analytical framework for online privacy research: What is missing. *Information & Management*, 54, 948–957. doi:10.1016/j.im.2017.02.004
- Grether, M. (2016). Using big data for online advertising without wastage: Wishful dream, nightmare or reality. *GfK Marketing Intelligence Review*, 8(2), 38-43. doi:10.1515/gfkmir-2016-0014
- Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology Journal*, 29, 75–91. doi:10.1007/BF02766777
- Gunasekaran, A., Papadopoulos, T., Dubey, R., Wamba, S. F., Childe, S. J., Hazen, B., & Akter, S. (2017). Big data and predictive analytics for supply chain and organizational performance. *Journal of Business Research*, 70, 308-317. doi:10.1016/j.jbusres.2016.08.004
- Gupta, S., & Schneider, M. (2018). Protecting customers' privacy requires more than anonymizing their data. *Harvard Business Review*. Retrieved from <https://hbr.org/2018/06/protecting-customers-privacy-requires-more-than-anonymizing-their-data>

- Halaweh, M., & Massry, E. A. (2015). Conceptual model for successful implementation of big data in organizations. *Journal of International Technology and Information Management, 24*(2). Retrieved from <https://scholarworks.lib.csusb.edu/jitim/vol24/iss2/2>
- Harindran, A., & Chandra, V. (2017). *Research method*. Andhra Pradesh, IN: Pearson Education India.
- HarrisX (2018). Inaugural tech media telecom pulse survey 2018. Retrieved from [http://harrisx.com/wp-content/uploads/2018/04/Inaugural-TMT-Pulse-Survey\\_-20-Apr-Final.pdf](http://harrisx.com/wp-content/uploads/2018/04/Inaugural-TMT-Pulse-Survey_-20-Apr-Final.pdf)
- Harting, R., Reichstein, C., & Schad, M. (2018). Potentials of digital business models: Empirical investigation of data driven impacts in industry. *Procedia Computer Science, 126*, 1495-1506. doi:10.1016/j.procs.2018.08.121
- Hartmann, P. M., Zaki, M., Feldmann, N., & Neely, A. (2016). Capturing value from big data: A taxonomy of data-driven business models used by start-up firms. *International Journal of Operations & Production Management, 36*(10), 1382-1406. doi:10.1108/IJOPM-02-2014-0098
- Hashmi, A. (2019). AI ethics: The next big thing in government. Retrieved from [https://www.worldgovernmentsummit.org/docs/default-source/default-document-library/deloitte-wgs-report-en-lq.pdf?sfvrsn=1acfc90b\\_0](https://www.worldgovernmentsummit.org/docs/default-source/default-document-library/deloitte-wgs-report-en-lq.pdf?sfvrsn=1acfc90b_0)
- Heng, T. Y., Wagner, T. D., Barnes, M. C., & Guarana, L. C. (2018). Archival research: Expanding the methodological toolkit in social psychology. *Journal of Experimental Social Psychology, 78*, 14-22. doi:10.1016/j.jesp.2018.04.012

- Hill, R. M. (2011). *Archival Strategies and Techniques*. Thousand Oaks, CA: Sage.
- IBM, & Harris Poll (2018). IBM cybersecurity and privacy research. Retrieved from <https://newsroom.ibm.com/download/IBM+Cybersecurity+PR+Research+-+Final.pdf>
- IBM, & Harris Poll (2019). Consumer attitudes towards data privacy. Retrieved from <https://newsroom.ibm.com/download/IBM+Data+Privacy.pdf>
- International Telecommunications Union (2019). Measuring digital development: Facts and figures 2019. Retrieved from <https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2019.pdf>
- Jai, T., Burns, D. L., & King, J. N. (2013). The effect of behavioral tracking practices on consumers' shopping evaluations and repurchase intention toward trusted online retailers. *Computers in Human Behavior*, 29(3), 901-909. doi:10.1016/j.chb.2012.12.021
- Janecek, V. (2018). Ownership of personal data in the Internet of Things. *Computer Law & Security Review*, 34(5), 1039-1052. doi:10.1016/j.clsr.2018.04.007
- Jobs, C. G., Aukers, S. M., & Gilfoil, D. M. (2015). The impact of big data on your firms marketing communications: A framework for understanding the emerging marketing analytics industry. *Academy of Marketing Studies Journal*, 19(2), 81-92.
- Jobs, C. G., Gilfoil, D. M., & Aukers, S. M. (2016). How marketing organizations can benefit from big data advertising analytics. *Academy of Marketing Studies Journal*, 20(1), 18-35.

- King, J. N., & Forder, J. (2016). Data analytics and consumer profiling: Finding appropriate privacy principles for discovered data. *Computer Law & Security Review*, 32(5), 696-714. doi:10.1016/j.clsr.2016.05.002
- King, N., & Brooks, J. (2018). Thematic analysis in organisational research. In C. Cassell, A. L. Cunliffe, & G. Grandy. *The sage handbook of qualitative business and management research methods*. London, UK: Sage. doi: 10.4135/9781526430236
- Kissell, J. (2019). *Take control of your online privacy*. TidBITS Publishing.
- Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, 64, 122-134. doi:10.1016/j.cose.2015.07.002
- Kramer, J., & Wohlfarth, M. (2018). Market power, regulatory convergence, and the role of data in digital markets. *Telecommunications Policy*, 42(2), 154-171. doi:10.1016/j.telpol.2017.10.004
- Kshetri, N. (2014). Big data's impact on privacy, security and consumer welfare. *Telecommunications Policy*, 38(11), 1134-1145. doi:10.1016/j.telpol.2014.10.002
- Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons*, 60(3), 293-303. doi:10.1016/j.bushor.2017.01.004
- Li, B., Ch'ng, E., Chong, A. Y., & Bao, H. (2016). Predicting online e-marketplace sales performances: A big data approach. *Computers & Industrial Engineering*, 101, 565-571. doi:10.1016/j.cie.2016.08.009

- Libaque-Saenz, F. C., Chang, Y., Kim, J., Park, M., & Rho, J. J. (2016). The role of perceived information practices on consumers' intention to authorise secondary use of personal data. *Behaviour & Information Technology*, 35(5), 339-356. doi:10.1080/0144929X.2015.1128973
- Lipu, S., Williamson, K., & Lloyd, A. (2007). *Explore methods in information literacy research*. Wagga Wagga, AU: Centre for Information Studies.
- Lucas, K. (2018). Archive searching for research. In M. Allen, *The SAGE Encyclopedia of Communication Research Methods*. Thousand Oaks, CA: Sage. doi:10.4135/9781483381411
- Mack, N., Woodsong, C., MacQueen, M. K., Guest, G., & Namey, E. (2005). Qualitative research methods: A data collector's field guide. Retrieved from <https://www.fhi360.org/sites/default/files/media/documents/Qualitative%20Research%20Methods%20-%20A%20Data%20Collector%27s%20Field%20Guide.pdf>
- Mai, J. (2016). Big data privacy: The datafication of personal information. *The Information Society*, 32(3), 192-199. doi:10.1080/01972243.2016.1153010
- Mamonov, S., & Benbunan-Fich, R. (2018). The impact of information security threat awareness on privacy-protective behaviors. *Computers in Human Behavior*, 83, 32-44. doi:10.1016/j.chb.2018.01.028
- Mantelero, A. (2016). Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection. *Computer Law & Security Review*, 32(2), 238-255. doi:10.1016/j.clsr.2016.01.014

- Marreiros, H., Tonin, M., Vlassopoulos, M., & Schraefel, M. C. (2017). "Now that you mention it": A survey experiment on information, inattention and online privacy. *Journal of Economic Behavior & Organization* 140, 1-17.  
doi:10.1016/j.jebo.2017.03.024
- Matthias, O., Fouweather, I., Gregory, I., & Vernon, A. (2017). Making sense of big data: Can it transform operations management. *International Journal of Operations & Production Management*, 37(1), 37-55. doi:10.1108/IJOPM-02-2015-0084
- McIntush, E. K., Pierce, R., McIntush, E., Alcalá, A., Garza, A. K., Hardin, E., ... Burlbaw, M. L. (2019). Advancing archival research with technology: Application to Brazoria county school, 1917-1921. *American Educational History Journal*.
- McKee, H. A., & Porter, J. E. (2012). The ethics of archival research. *College Composition and Communication*, 64(1), 59-81.
- Merriam, B. S., & Tisdell, J. E. (2015). *Qualitative research: A guide to design and implementation*. San Francisco, CA: Jossey-Bass.
- Mendelson, D., & Mendelson, D. (2017). Legal protections for personal health information in the age of big data: A proposal for regulatory framework. *Ethics, Medicine and Public Health*, 3(1), 37-55. doi:10.1016/j.jemep.2017.02.005
- Meulen, R. (2017). Gartner say 8.4 billion connected things will be in use in 2017, up 31 percent from 2016. Retrieved from <https://www.gartner.com/en/newsroom/press-releases/2017-02-07-gartner-says-8-billion-connected-things-will-be-in-use-in-2017-up-31-percent-from-2016>

- Mills, J. A., & Mills, H. J. (2018). Archival research. In C. Cassell, A. L. Cunliffe, & G. Grandy. *The sage handbook of qualitative business and management research methods*. London, UK: Sage. doi:10.4135/9781526430236
- Milne, G. (2015). *Digital privacy in the marketplace*. New York, NY: Business Expert Press.
- Mukhiya, K. S., Wei, T., & Lee, J. (2018). *Hands-on big data modeling*. Birmingham, UK: Packt.
- Mulligan, P. S., Linebaugh, D. C., & Freeman, C. W. (2019a). Data protection law: An overview. Retrieved from <https://crsreports.congress.gov/product/pdf/R/R45631>
- Mulligan, P. S., Linebaugh, D. C., & Freeman, C. W. (2019b). Data protection and privacy law: An introduction. Retrieved from <https://crsreports.congress.gov/product/pdf/IF/IF11207>
- Myers, D. M. (2002). *Qualitative research in information systems: A reader*. Thousand Oaks, CA: Sage.
- Nada, R. S. (2016). How to use big data to drive your supply chain. *California Management Review*, 58(3), 26-48. doi:10.1525/cm.2016.58.3.26
- Nunan, D., & Domenico, D. M. (2013). Market research and the ethics of big data. *International Journal of Market Research*, 55(4), 505-520. doi:10.2501/ijmr-2013-015
- Oracle (2018). Oracle data cloud: Data directory. Retrieved from <http://www.oracle.com/us/solutions/cloud/data-directory-2810741.pdf>



- Organisation for Economic Co-operation and Development (2014). Protecting privacy in a data-driven economy: Taking stock of Current Thinking. Retrieved from [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=dsti/icc/p/reg\(2014\)3&doclanguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=dsti/icc/p/reg(2014)3&doclanguage=en)
- Ott, N., & Zylberberg, H. (2016). A European perspective on the protection of personal data in cyberspace: Explaining how the European Union is redefining ownership and policies of personal data beyond national borders. *Harvard Kennedy School Review*, 16, 69-75.
- Patterson, D., & Davis, K. (2012). *Ethics of big data*. Sebastopol, CA: O'Reilly Media.
- Perko, I., & Ototsky, P. (2016). Big data for business ecosystem players. *Our Economy (Nase Gospodarstvo)*, 62(2), 12-24. doi:10.1515/ngoe-2016-0008
- Petrow, S. (2018). Delete yourself from the Internet's people finder sites: Is it worth it. Retrieved from <https://www.usatoday.com/story/tech/columnist/2018/05/04/delete-yourself-internet-people-finder-sites-worth/553371002/>
- Prince, C. (2018). Do consumers want to control their personal data? Empirical evidence. *International Journal of Human-Computer Studies*, 110, 21-32. doi:10.1016/j.ijhcs.2017.10.003
- Ramirez, E., Brill, J., Ohlhausen, K. M., & McSweeney, T. (2014a). Big data: A tool for inclusion or exclusion. Retrieved from <https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf>

- Ramirez, E., Wright, D. J., Brill, J., Ohlhausen, K. M., & McSweeney, T. (2014b). Data brokers: A call for transparency and accountability. Retrieved from <https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf>
- Regnell, B., Rainer, A., Host, M., & Runeson, P. (2012). *Case study research in software engineering: Guidelines and examples*. Hoboken, NJ: John Wiley & Sons.
- Reinsel, D., Gantz, J., & Rydning, J. (2018a). Data age 2025: The digitization of the world from edge to core. Retrieved from <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>
- Reinsel, D., Shegawi, M., & Gantz, F. J. (2018b). Healthcare: DATCON level 3: An industry with a weak data management pulse. Retrieved from <https://www.seagate.com/www-content/our-story/trends/files/idc-seagate-datcon-healthcare.pdf>
- Riemer, J. F., Quartaroly, T. M., & Lapan, D. S. (2011). *Qualitative research: An introduction to methods and designs*. Hoboken, NJ: Jossey-Bass Wiley.
- Roderick, L. (2014). Discipline and power in the digital age: The case of the US consumer data broker industry. *Critical Sociology*, 40(5), 729-746.  
doi:10.1177/0896920513501350
- SAS (2018). Data privacy: Are you concerned. Retrieved from <https://www.sas.com/content/dam/SAS/documents/marketing-whitepapers-ebooks/sas-whitepapers/en/data-privacy-110027.pdf>

- Savage, W. C. (2019). Managing the ambient trust commons: The economics of online consumer information privacy. Retrieved from [https://law.stanford.edu/wp-content/uploads/2019/01/Savage\\_20190129-1.pdf](https://law.stanford.edu/wp-content/uploads/2019/01/Savage_20190129-1.pdf)
- Schudy, S., & Utikal, V. (2017). You must not know about me: On the willingness to share personal data. *Journal of Economic Behavior & Organization*, *141*, 1-13. doi:10.1016/j.jebo.2017.05.023
- Shenton, K. A. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, *22*, 63-75. doi:10.3233/EFI-2004-22201
- Smit, G. E., Noort, V. G., & Voorveld, A. M. H. (2014). Understanding online behavioural advertising: User knowledge, privacy concerns and online coping behaviour in Europe. *Computers in Human Behavior*, *32*, 15-22. doi:10.1016/j.chb.2013.11.008
- Smith, A., & Anderson, M. (2016). Online shopping and e-commerce. Retrieved from [https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2016/12/PI\\_2016.12.19\\_Online-Shopping\\_FINAL.pdf](https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2016/12/PI_2016.12.19_Online-Shopping_FINAL.pdf)
- Smith, A., & Anderson, M. (2018). Social media use in 2018: A majority of Americans use Facebook and YouTube, but young adults are especially heavy users of Snapchat and Instagram. Retrieved from [https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2018/02/PI\\_2018.03.01\\_Social-Media\\_FINAL.pdf](https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2018/02/PI_2018.03.01_Social-Media_FINAL.pdf)

- Smith, M. (2018). Cyber incident & breach trends report. Retrieved from [https://www.otalliance.org/system/files/files/initiative/documents/ota\\_cyber\\_incident\\_trends\\_report\\_jan2018.pdf](https://www.otalliance.org/system/files/files/initiative/documents/ota_cyber_incident_trends_report_jan2018.pdf)
- Spence, C. (2020). On the ethics of neuromarketing and sensory marketing. In J. T. Martineau, & E. Racine (Eds.), *Organizational Neuroethics: Advances in Neuroethics* (pp. 9-29). Switzerland: Springer Nature. doi:10.1007/978-3-030-27177-0\_3
- Stan, L. (2010). Archival records as evidence. In A. J. Mills, G. Durepos, & E. Wiebe (Eds.), *Encyclopedia of Case Study Research*. Thousand Oaks, CA: Sage. doi:10.4135/9781412957397
- State of California Attorney General (2020). Data broker registry. Retrieved from <https://oag.ca.gov/data-brokers>
- Stewart, L. (2019). Big data discrimination: Maintaining protection of individual privacy without disincentivizing businesses' use of biometric data to enhance security. *Boston College Law Review*, 60(1), 349-386.
- Sudtasan, T., & Mitomo, H. (2018). The internet of things as an accelerator of advancement of broadband networks: A case of Thailand, 42, 293-303. doi:10.1016/j.telpol.2017.08.008
- Swisher, K. (2018). Introducing the internet bill of rights. Retrieved from <https://www.nytimes.com/2018/10/04/opinion/ro-khanna-internet-bill-of-rights.html>

- Terry, G., Hayfield, N., Clarke, V., & Braun, V. (2017). Thematic analysis. In C. Willig, & W. Rogers. *The SAGE Handbook of qualitative research in psychology*. London, UK: Sage. doi:10.4135/9781526405555
- Tesar, M. (2015). Sources and interpretations: Ethics and truth in archival research. *History and Education, 44*(1), 101-114. doi:10.1080/0046760X.2014.918185
- Tikkinen-Piri, C., Rohunen, A., & Markkula, J. (2018). EU General Data Protection Regulation: Changes and implications for personal data collecting companies. *Computer Law & Security Review, 34*(1), 134-153. doi:10.1016/j.clsr.2017.05.015
- Tran, P. T. (2017). Personalized ads on Facebook: An effective marketing tool for online marketers. *Journal of Retailing and Consumer Services, 39*, 230-242. doi:10.1016/j.jretconser.2017.06.010
- Trepte, S., Teutsch, D., Masur, K. P., Eicher, C., Fischer, M., Hennhofer, A., & Lind, F. (2015). Do people know about privacy and data protection strategies? Towards the “online privacy literacy scale” (OPLIS). In S. Gutwirth, R. Leenes, & P. D. Hert (Eds.) *Reforming European data protection law* (pp. 333-365). doi:10.1007/978-94-017-9385-8\_14
- United Kingdom Information Commissioner’s Office (2017). Big data, artificial intelligence, machine learning and data protection. Retrieved from <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>
- U.S. Department of Labor (2019). Guidance on the protection of personal identifiable information. Retrieved from <https://www.dol.gov/general/ppii>

- U.S. Federal Trade Commission (2018). Competition and consumer protection in the 21st century. Retrieved from <https://www.ftc.gov/news-events/events-calendar/ftc-hearing-6-competition-consumer-protection-21st-century>
- U.S. House of Representatives, Committee on Financial Services (2017). Examining the Equifax data breach. Retrieved from <https://www.govinfo.gov/content/pkg/CHRG-115hrg30242/pdf/CHRG-115hrg30242.pdf>
- U.S. House of Representatives, Committee on Energy and Commerce (2017). Algorithms: How companies' decisions about data and content impact consumers. Retrieved from <https://www.govinfo.gov/content/pkg/CHRG-115hrg28578/pdf/CHRG-115hrg28578.pdf>
- U.S. House of Representatives, Committee on Energy and Commerce (2018). Facebook: Transparency and use of consumer data. Retrieved from <https://www.govinfo.gov/content/pkg/CHRG-115hrg30956/pdf/CHRG-115hrg30956.pdf>
- U.S. House of Representatives, Committee on Science, Space, and Technology (2018). Big data challenges and advanced computing solutions. Retrieved from <https://www.govinfo.gov/content/pkg/CHRG-115hrg30879/pdf/CHRG-115hrg30879.pdf>

- U.S. House of Representatives, Committee on Small Business (2019). A fair playing field: Investigating big tech's impact on small business. Retrieved from <https://www.govinfo.gov/content/pkg/CHRG-116hrg38314/pdf/CHRG-116hrg38314.pdf>
- U.S. House of Representatives, Subcommittee on Digital Commerce and Consumer Protection of the Committee on Energy and Commerce (2018). Understanding the digital advertising ecosystem. Retrieved from <https://www.govinfo.gov/content/pkg/CHRG-115hrg34638/pdf/CHRG-115hrg34638.pdf>
- U.S. House of Representatives, Subcommittee on Communications and Technology of the Committee on Energy and Commerce (2018). Protecting customer network proprietary information in the internet age. Retrieved from <https://www.govinfo.gov/content/pkg/CHRG-115hrg35164/pdf/CHRG-115hrg35164.pdf>
- U.S. House of Representatives, Subcommittee on Financial Institutions and Consumer Credit of the Committee on Financial Services (2017). Data security: Vulnerabilities and opportunities for improvement. Retrieved from <https://www.govinfo.gov/content/pkg/CHRG-115hrg30771/pdf/CHRG-115hrg30771.pdf>

- U.S. House of Representatives, Subcommittee on Oversight and Investigations of the Committee on Energy and Commerce (2017). Identity verification in a post-breach world. Retrieved from <https://www.govinfo.gov/content/pkg/CHRG-115hhr28714/pdf/CHRG-115hhr28714.pdf>
- U.S. Senate, Committee on Banking, Housing, and Urban Affairs (2018). FinTech: Examining digitalization, data, and technology. Retrieved from <https://www.govinfo.gov/content/pkg/CHRG-115shrg32753/pdf/CHRG-115shrg32753.pdf>
- U.S. Senate, Committee on Commerce, Science, and Transportation (2012). The need for privacy protections: Perspectives from the administration and the federal trade commission. Retrieved from <https://www.govinfo.gov/content/pkg/CHRG-112shrg81793/pdf/CHRG-112shrg81793.pdf>
- U.S. Senate, Committee on Commerce, Science, and Transportation (2013). *A review of the data broker industry: Collection, use, and sale of consumer data for marketing purposes*. Retrieved from [https://www.commerce.senate.gov/public/\\_cache/files/0d2b3642-6221-4888-a631-08f2f255b577/AE5D72CBE7F44F5BFC846BECE22C875B.12.18.13-senate-commerce-committee-report-on-data-broker-industry.pdf](https://www.commerce.senate.gov/public/_cache/files/0d2b3642-6221-4888-a631-08f2f255b577/AE5D72CBE7F44F5BFC846BECE22C875B.12.18.13-senate-commerce-committee-report-on-data-broker-industry.pdf)
- U.S. Senate, Committee on Commerce, Science, and Transportation (2015). The connected world: Examining the internet of things. Retrieved from <https://www.govinfo.gov/content/pkg/CHRG-114shrg99818/pdf/CHRG-114shrg99818.pdf>



U.S. Senate, Committee on Commerce, Science, and Transportation (2016). How will the FCC's proposed privacy regulations affect consumers and competition. Retrieved from <https://www.govinfo.gov/content/pkg/CHRG-114shrg24204/pdf/CHRG-114shrg24204.pdf>

U.S. Senate, Committee on Commerce, Science, and Transportation (2017). Protecting consumers in the era of major data breaches. Retrieved from <https://www.govinfo.gov/content/pkg/CHRG-115shrg33395/pdf/CHRG-115shrg33395.pdf>

U.S. Senate, Committee on Commerce, Science, and Transportation, & Committee on The Judiciary (2018). Facebook, social media privacy, and the use and abuse of data. Retrieved from <https://www.commerce.senate.gov/2018/4/facebook-social-media-privacy-and-the-use-and-abuse-of-data>

U.S. Senate, Subcommittee on Communications, Technology, Innovation, and the Internet of the Committee on Commerce, Science, and Transportation (2017). Digital decision-making: The building blocks of machine learning and artificial intelligence. <https://www.govinfo.gov/content/pkg/CHRG-115shrg37295/pdf/CHRG-115shrg37295.pdf>

U.S. Senate, Subcommittee on Consumer Protection, Product Safety, Insurance, and Data Security of the Committee on Commerce, Science, and Transportation (2018). Data security and bug bounty programs: Lessons learned from the Uber breach and security researchers. Retrieved from <https://www.govinfo.gov/content/pkg/CHRG-115shrg37302/pdf/CHRG->

115shrg37302.pdf

Ventresca, J. M., & Mohr, W. J. (2017). Archival research methods. In J. Baum (Ed.),

*The Blackwell Companion to Organizations* (pp. 805-828). Hoboken, NJ:

Blackwell. doi:10.1002/9781405164061.ch35

Verizon (2018). 2018 Data breach investigations report. Retrieved from

[https://enterprise.verizon.com/resources/reports/DBIR\\_2018\\_Report.pdf](https://enterprise.verizon.com/resources/reports/DBIR_2018_Report.pdf)

VPRO Documentary (2019). The age of surveillance capitalism. Retrieved from

<https://www.youtube.com/watch?v=hIXhnWUmMvw>

Volker, B. (2016). Big data in market research: Why more data does not automatically

mean better information. *GfK Marketing Intelligence Review*, 8(2), 56-63.

doi:10.1515/gfkmir-2016-0017

Wamba, F. S., Gunasekaran, A., Akter, S., Ren, J. S., Dubey, R., & Childe, J. S. (2017).

Big data analytics and firm performance: Effects of dynamic capabilities. *Journal*

*of Business Research*, 70, 356–365. doi:10.1016/j.jbusres.2016.08.009

Welch, C. (2014). The archaeology of business networks: The use of archival records in

case study research. *Journal of Strategic Marketing*, 8(2), 197-208.

doi:10.1080/0965254X.2000.10815560

West, M. S. (2019). Data capitalism: Redefining the logics of surveillance and privacy.

*Business & Society*, 58(1), 20-41. doi:10.1177/0007650317718185

White House (2013). Consumer data privacy in a networked world: A framework for

protecting privacy and promoting innovation in the global digital economy.

*Journal of Privacy and Confidentiality*, 4(2), 95-142. doi:10.29012/jpc.v4i2.623

- White House (2014). Big Data: Seizing opportunities, preserving values. Retrieved from [https://obamawhitehouse.archives.gov/sites/default/files/docs/big\\_data\\_privacy\\_report\\_may\\_1\\_2014.pdf](https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf)
- Williamson, K., & Johanson, G. (2013). *Research methods: Information, systems and contexts*. Prahran, VIC, AU: Tilde Publishing and Distribution.
- Willig, C. (2017). Interpretation in qualitative research. In C. Willig, & W. Rogers (Eds.), *The SAGE Handbook of qualitative research in psychology* (pp. 274-288). London, UK: Sage. doi:10.4135/9781526405555
- Wood, P. J. (2011). Understanding and evaluating historical sources in nursing history research. *Nursing Praxis in New Zealand*, 27(1), 25-33.
- World Economic Forum (2018). Identity in a digital world: A new chapter in the social contract. Retrieved from [http://www3.weforum.org/docs/WEF\\_INSIGHT\\_REPORT\\_Digital%20Identity.pdf](http://www3.weforum.org/docs/WEF_INSIGHT_REPORT_Digital%20Identity.pdf)
- Yeh, C. (2018). Pursuing consumer empowerment in the age of big data: A comprehensive regulatory framework for data brokers. *Telecommunications Policy*, 42(4), 282-292. doi:10.1016/j.telpol.2017.12.001
- Yilmaz, K. (2013). Comparison of quantitative and qualitative research traditions: Epistemological, theoretical, and methodological differences. *European Journal of Education*, 48(2). doi:10.1111/ejed.12014
- Young, C., & Brooker, M. (2006). Safeguarding sacred lives: The ethical use of archival data for the study of diverse lives. In J. E. Trimble, & B. C. Fisher (Eds.), *The*

*handbook of ethical research with ethnocultural populations & communities* (pp. 282-298). Thousand Oaks, CA: Sage. doi:10.4135/9781412986168

Zaki, M. (2019). Digital transformation: Harnessing digital technologies for the next generation of services. *Journal of Services Marketing*, 33(4), 429-435. doi:10.1108/JSM-01-2019-0034

Zhang, T., Wang, Y. C. W., & Pauleen, J. D. (2017). Big data investments in knowledge and non-knowledge intensive firms: What the market tells us. *Journal of Knowledge Management*, 21(3), 623-639. doi:10.1108/JKM-12-2016-0522

Zhu, Y., & Chang, J. (2016). The key role of relevance in personalized advertisement: Examining its impact on perceptions of privacy invasion, self-awareness, and continuous use intentions. *Computers in Human Behavior*, 65, 442-447. doi:10.1016/j.chb.2016.08.048

## Appendix A: Archived Document Selection Protocol

**Researcher:** Thanh Pham

**Archival Material Inquiry Method**

- Locate digital archive libraries and private company websites. Archival libraries include

Table A1

*Archival Library*

Archival Library	Website Uniform Resource Locator (URL)
The United States House of Representatives hearing transcripts	<a href="https://www.house.gov/">https://www.house.gov/</a>
The United States Senate hearing transcripts	<a href="https://www.senate.gov">https://www.senate.gov</a>
The U.S. Federal Legislative Information	<a href="https://www.congress.gov">https://www.congress.gov</a>
The U.S. Federal Trade Commission testimonies	<a href="https://www.ftc.gov/">https://www.ftc.gov/</a>
The U.S. National Archives	<a href="https://www.archives.gov">https://www.archives.gov</a>
Library of Congress	<a href="https://www.loc.gov/">https://www.loc.gov/</a>
Videos and transcripts of business executive officer interviews	<a href="https://www.youtube.com">https://www.youtube.com</a>
Catalog of U.S. Government Publications (CGP)	<a href="https://catalog.gpo.gov">https://catalog.gpo.gov</a>
GovInfo	<a href="https://www.govinfo.gov/">https://www.govinfo.gov/</a>
The U.S. Government Publishing Office (GPO)	<a href="https://www.gpo.gov">https://www.gpo.gov</a>
DocumentCloud	<a href="https://www.documentcloud.org/home">https://www.documentcloud.org/home</a>
American Cable Television Industry	<a href="https://www.c-span.org/">https://www.c-span.org/</a>
The World Privacy Forum	<a href="https://www.worldprivacyforum.org">https://www.worldprivacyforum.org</a>

Table A2

*Data Provider Company*

Company Name	Website URL
Acxiom	<a href="https://www.acxiom.com">https://www.acxiom.com</a>
Corelogic	<a href="https://www.corelogic.com">https://www.corelogic.com</a>
Epsilon	<a href="https://us.epsilon.com">https://us.epsilon.com</a>
Equifax	<a href="https://www.equifax.com/business/">https://www.equifax.com/business/</a>
Experian	<a href="https://www.experian.com">https://www.experian.com</a>
Facebook	<a href="https://www.facebook.com">https://www.facebook.com</a>
GfK	<a href="https://www.gfk.com/en-us/">https://www.gfk.com/en-us/</a>
Google	<a href="https://www.google.com">https://www.google.com</a>
i360	<a href="https://www.i-360.com/">https://www.i-360.com/</a>
Mastercard	<a href="https://www.mastercardservices.com">https://www.mastercardservices.com</a>
Oracle Data Cloud Service	<a href="https://www.oracle.com/cloud/data-hotline/">https://www.oracle.com/cloud/data-hotline/</a>
Salesforce	<a href="https://www.salesforce.com">https://www.salesforce.com</a>

- Identify types of archival materials that are available on the digital archive library.  
Types of archives include legislations, policies, hearing transcripts, hearing videos, videos and transcripts of business executive officer interviews, business documents, financial reports, policies, and statements.
- Identify search functionalities and filter options that are provided by the digital archive libraries.
- Search archived materials, recording search keywords or terms. The search keywords included data broker, data provider, data-driven business, collect consumer data, consumer data in online marketing, privacy concerns, consumer privacy, privacy violation, online privacy, consumer data, personal data, privacy information, data privacy, big data, ethics of big data, data ethics, privacy law, privacy protection, regulation control, and data breach. I opened to any possibility that can lead to relevant documents.

### **Document Archival Material Metadata to Manage Collected Documents**

- Title: The title of archival document.
- Author: The author of document.
- Published date: Published, created, or released date of archival document.
- Location: The URL links to documents
- Type of archived materials: Describes document type (e.g., document, text, video, voice, or image)
- Original organization: Organization that documents belong to
- Note: Notes on or descriptions of documents.
- Search keyword: Search keywords or terms which are used to find documents (e.g., data broker, data provider, data-driven business, collect consumer data, consumer data in online marketing, privacy concerns, consumer privacy, privacy violation, online privacy, consumer data, personal data, privacy information, data privacy, big data, ethics of big data, data ethics, privacy law, privacy protection, regulation control, and data breach).
- Content: Concepts which are relevant to the study.

### **Overview of the Data-Driven Ecosystem**

- What are current business models in the data-driven ecosystem? Businesses that use data for marketing, generating revenue, enhancing customer engagement, or improving business operations.
- How do data-driven businesses use information technologies to manage and process consumer data?

## **Identifying Data Sources, Data Types, and Ownership in the Data-Driven Ecosystem**

- Where are data sources in the data-driven ecosystem?
- What types of information do private companies collect from consumers?

## **Understanding Data Collection Methods**

- How do private companies collect consumer data?
- What methods and algorithms do private companies use to collect information from consumers?

## **Data Privacy Policy Statements**

- What privacy policy statements do companies provide to consumers?
- How do practices of collecting and using consumer data comply with current protection laws?

## **Current Data Provider and Data Processing Industry**

- What types of businesses operate in the data-providing industry?
- How do data provider companies manage, process, analyze, and use consumer data?

## **Classifying Data Products and Services**

- What types of data products and services do data providers offer to their clients and partners?

## **Types of Data Consumer Businesses**

- What types of businesses acquire or buy data products and services from data providers?



**Understanding Benefits, Risks, and Ethics in the Data-Driven Ecosystem**

- What are the benefits of using consumer information in the data-driven ecosystem?
- What are the risks of using consumer information in the data-driven ecosystem?

## Appendix B: Permission Letter for Use of ITU Figures

**From:** Thanh Pham <[redacted]>  
**Sent:** jeudi, 20 février 2020 00:21  
**To:** Pressinfo (ITU Press Service) <[redacted]>; ITUMAIL, ITU <[redacted]>  
**Subject:** Permission for Using Figures in the Dissertation paper

Dear International Telecommunication Union (ITU):

My name is Thanh Pham and I am a Ph.D. student at Walden University. I am in the process of writing my dissertation paper, researching business practices of collection and use consumer data without an individual's consent in the digital economy.

I am writing to obtain permission to use below figures in the article, "Measuring digital development: Facts and figures 2019" by ITU (2019), retrieved from <https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2019.pdf>

- Figure. Individuals using the Internet, 2005-2019 on page 1.
- Figure. Evolution of mobile and fixed subscriptions, 2005-2019 on page 5.
- Figure. Percentage of households with Internet access at home and with a computer, 2019 on page 7.

My dissertation paper will be published through Walden University repository and its publication databases.

Thank you for your consideration and support academic research in the information management system field.

Sincerely,  
 Thanh Pham.

**From:** Legal Affairs Unit, ITU <[redacted]>  
**Sent:** Thursday, February 20, 2020 2:35 AM  
**To:** Thanh Pham <[redacted]>  
**Subject:** RE: Permission for Using Figures in the Dissertation paper

Dear Mr Thanh Pham,

I am writing in follow-up to your message below which has been forwarded to ITU's Legal Affairs Unit for response and am pleased to confirm that the Union can accommodate your request on the following terms and conditions:

1. this authorization is limited to the ITU figures identified in your message below and for the sole purpose outlined therein;
2. this authorization is granted on a non-exclusive basis and is non-transferable to third parties; and
3. ITU will be clearly identified as the source of the material: "Source: ITU – Measuring digital development; Facts and figures 2019".

Please send confirmation of acceptance of the above terms by return email.

I do hope this is helpful and look forward to hearing from you.

Yours faithfully,

**Margaret CAMPION**  
*Assistante, Unité des affaires juridiques*  
**International Telecommunication Union**  
 Tel : [redacted]

## Appendix C: Permission Letter for Use of IDC Figures

**From:** Thanh Pham <[REDACTED]>  
**Sent:** Wednesday, February 19, 2020 5:10 PM  
**To:** Permissions <[REDACTED]>  
**Subject:** Written Permission Request for using Figures in the Dissertation paper

Dear International Data Corporation (IDC):

I am a Ph.D. student at Walden University. I am writing my dissertation paper, researching business practices of collection and use consumer data without an individual's consent in the digital economy.

I am writing to obtain permission for using Figure 1. Annual Size of the Global Datasphere on page 6 and Figure 4. Where is the data stored on page 10 in the article, "The digitization of the world from edge to core" by David Reinsel, John Gantz, and John Rydning, (2018), retrieved from <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>

My dissertation paper will be published through Walden University repository and its publication databases.

Thank you for your consideration and support academic research in the information management system field.

Sincerely,

Thanh Pham.

Permissions <[REDACTED]>  
Thu 2/20/2020 6:05 AM  
To: Thanh Pham



Hi Thanh,

This is fine. Please source it this way - IDC White Paper, sponsored by Seagate, Data Age 2025: The Digitization of the World from Edge to Core, November 2018.

Stephanie Geary  
Senior Inquiry Analyst, Content Permissions  
[REDACTED]