CHAPTER SEVEN

SOCIAL MEDIA RESEARCH: SAMPLING TECHNIQUES, DATA COLLECTION, ANALYSIS, AND DISCUSSION

Kenneth Ohei

Management Sciences, Mangosuthu University of Technology, South Africa
kennethohei@gmail.com

Joshua Ebere Chukwuere

Department of Information Systems, North-West University, South Africa
joshchukwuere@gmail.com

CHAPTER OBJECTIVES
- Overview of social media research sampling techniques.
- Data collection methods for social media research.
- Data analysis methods for social media research.
- Discussion of social media research.

ABSTRACT

Social media has skyrocketed and massively been incorporated into government, business sectors, universities, and individuals globally in the past years. The deployment of social media tools enabled by Web 3.0 technologies for content dissemination, information diffusion, and engagement by publishing thoughts and views to a targeted audience creates a new meaning to life and social interaction through social media research (SMR). Social media research (SMR) has given rise to massive data collection, analyses, and discussion using YouTube, Facebook, Twitter, Instagram, WhatsApp, and other visual content creators; however, they are less studied thus far. This chapter proposes a sampling technique, data analysis, and discussion for SMR. The chapter applied a narrative literature review (NLR) process in unpacking SMR sampling techniques, data collection, analysis, and discussion. Furthermore, the chapter looked into the critical area of research revolving around ethics and privacy and finally suggested general basics, key principles, and elements that guide future SMR.

OVERVIEW OF SOCIAL MEDIA (SM)

Social media (SM) can be regarded as a new and evolving area of research with the potential to reshape future research using social media platforms. Social media technologies have been at the forefront and gained considerable responsiveness among different users who continually post videos, pictures, text messages, and content online. Similarly, academics have started engaging empirical studies on the platform for decades (Snelson, 2016). Social media, as a channel for massive content-diffusion, distribution, information dissemination,

and broadcasting of individuals' thoughts and ideas, has been embraced and gathered attention across the heterogeneous area of studies, exploring online action from government and mass media to health and business sectors, and the list continues (Highfield & Leaver, 2015). The emergence of Web 3.0 as an enabler of social media plays a fundamental part in enhancing academic business processes towards information management, decision-making, integration of knowledge creation, and use (Ohei & Brink, 2019a; 2019b). Ohei (2019)asserts that Web 3.0 attempts to link, combine, and analyze several datasets to attain a new information stream. Web 3.0 can increase data management, sustain accessibility, provide creativity and innovation, and improve clients' satisfaction. Social media enabled by Web 3.0 technologies is becoming affordable, efficient, effective, cloud-based, standardized, global, and mobile platforms for meeting personal (individuals) and academic needs. It supports collective interaction in establishing effective communities and creating and trading content. Furthermore, the increase in using social media platforms can be attributed to its openness to everyone, including government, business, customers, suppliers, academics, students, and other users, to list but a few (Alberghini, Cricelli & Grimaldi, 2014).

The concept of social media denotes web-based technology platforms capable of turning interaction and communication into collaborative dialogue. Such SM tools entail wikis, weblogs, media-sharing applications, social networking applications (sites), Internet forums, magazines, podcasts, and social bookmarking (Ohei & Brink, 2019b; Alberghini et al., 2014). All these tools are made affordable through the arrival of Web 3.0 and the Internet. The SM platforms can encourage formal and informal conversation, collaborative content generation, dialogue, and knowledge sharing, giving academics more access to a wide range of ideas, networking, and representations. If deployed well, Web 3.0 provides groups of technologies, business, and social models, characterized by engagement, open participation in using trivial technologies, and decentralized application processes and distribution. These useful applications can facilitate collaborative knowledge distribution, interoperability, user-centred design, and interaction on the Internet (Lee & Ma, 2012; Kaplan & Haenlein, 2010) and application in the academic research community.

Emergent academic researchers (Alberghini et al., 2013; Ferron et al., 2011; Koo et al., 2011) have suggested numerous benefits that Web 3.0, in powering SM platforms and application technologies, can bring to any individual, educational, business, and government institutions. The utilization of SM has given rise to an advanced improvement in massive data analyses and amplified the awareness and gathering of pertinent information. However, there is strong evidence that it brings up pellucidity and group effort while encouraging data distribution and being proficient in rapidly dispensing it. Furthermore, inferring from SM's business perspective, it can increase the number of business operations due to their competence in generating brand awareness, the workforce's trustworthiness, and stimulating employee mindfulness. Additionally, SM platforms allow academia, government institutions, and people to work actively and, on the other hand, have a detailed knowledge of the concerns and interests raised by their audience. As the world is increasingly becoming a global village and connected, the focus has shifted to amplifying SMR and modes of data collection,

sampling techniques, analysing, and discussing SMR (Hansen et al., 2010). Therefore, this research objective has been formulated to provide direction to this chapter. The chapter's objective was to establish guidance on conducting an SMR focusing on data sampling techniques, data collection, analysis, and discussions.

This chapter discusses the SMR sampling techniques, data collection, data analysis, and discussion to offer considerable research strategies and approaches to be followed when conducting SMR activities. This chapter presents a narrative literature review (NLR) as a research method for conducting SMR, which is scientific and philosophical. According to Booth, Sutton and Papaioannou (2016), recognising philosophies in the research literature validates the explicit approach to the theme and the choice of methods. The literature review aids the choice of suitable approaches, but it also encourages the expansion of knowledge and makes a meaningful contribution to the body of knowledge. In a nutshell, the drive was to build a theoretical foundation utilising a narrative literature review process, offering a comprehensive elucidation of the research available in the literature. Consequently, this section presents the overview of SM purported to contextualise by offering an overview of connected academic activity in the developing area of SMR (*read detailed SMR in Chapter 2*). Furthermore, the overview is accompanied by a methodology discourse, which entails sampling techniques and data gathering.

BACKGROUND CONTEXT AND LITERATURE REVIEW

SMR is gradually evolving with research practitioners in building methodologies and techniques to appreciate how the audiences use SM platforms, their interactive nature, behaviour, and discussions using such platforms. Therefore, SM plays an important role, as it has become an integral part of daily livelihood and lifestyle. Consequently, it becomes a crucial area to channel research activities to comprehend this new opportunity for discourse through SMR. SMR is the basic procedure of gathering, interpreting, analysing, and discussing SM data. The procedure can be undertaken using either a quantitative, qualitative, or mixed-methods research approach, with the essential drive to comprehend how audiences narrate topics by merely deploying tools and data extraction practices. Thanks to innovative SM platforms, technologies, and other tools, which have enabled academics and consultants to collect data relating to definite events, subject matters, or topics within a precise audience cluster from YouTube, Facebook, Twitter and Instagram, WhatsApp, and many more. SMR may be conducted in the form of social listening and audience intelligence by noting the content and evidence generated online and discovering predispositions and ideas that emerge from such data sources (Snelson, 2016).

Nonetheless, SMR is growing and gaining academic attention as supported by the proliferation of academic literature and an increasing number of all-inclusive literature studies that have led to probing several facets of SMR. The probe's purpose on SMR is that SM is becoming more and more attractive for the user. The platform provides quicker access to converse ideas and may be classified as a key dataset or data source regarding how individuals interact. For this notion, SM platforms are crucial for deriving peoples'

ideologies, contents, and interactions for research purposes (Snelson, 2016; Bredl, Hünniger & Jensen, 2012). Furthermore, a summarised review of a few listed SMR academic literature is in Table 7.1. Even though not a full list, however, Table 7.1 articles below provide a narrative literature review and analysis defined by the respective scholars.

Table 7.1: A systematic literature review on SMR

Author (s)	Short title	Focus on literature	Field
Stieglitz *et al.* (2018; 2014)	"Social media analytics-Challenges in topic discovery, data collection, and data preparation".	The paper focuses on social media research analytics and processes that involve data discovery, gathering, planning, and analysis. It reviewed the literature on the challenges, and difficulties emerging in involving specific data analysis methods.	Information management
Bredl *et al.* (2012)	"Methods for analysing social media: Introduction to the special issue".	The emphasis is on research strategies and methods for analysingSMR. It reports on the advent results onadvanced social science areas.	Social media analysis
Snelson (2016)	"Qualitative and mixed methods social media research: A review of the literature".	This article exploresthe methodological analysis of mixed-methods approaches covering SMR. A review of literature collections was published from 2007 through 2013.	Educational technology
Kivunja (2015)	"A convenience sampling investigation into the use of social media technologies"	This paper contributes to articulating the theoretical framework in pedagogy as potential methodologies of convenience sampling case study.	Education
Boulianne (2015)	Social media use and participation: A meta-analysis of current research	It shows the results of a meta-analysis of SMR and its use and engagements. The meta-data demonstrates a goodassociation between SM use and participation.	Sociology
Kern *et al.* (2016)	Gaining insights from social media language: Methodologies and challenges	The study focuses on SM language research by showcasing descriptive, and predictive language data analyses.	Social media for psychology
Camacho *et al.* (2020)	New trends and applications in social media analytics	This presents special research that focuses on practical usage of data science (DS) and artificial intelligence (AI) in social media analytics (SMA). The focus covers disciplines like natural language, machine learning, processing, evolutionary computation and sentiment analysis.	Computer systems
Sivarajah *et al.* (2020)	Role of big data and social media analytics for business to business sustainability: A participatory web context.	This study's focus presents a paradigm shift on passive information source moving intoa collaborative, and intelligent participatory web, that commend active engagement of users and contributors.	Marketing management.

Several research themes covered across the assortment of literature review evidence suggest the multiplicity in prominence and fields of study from which the evidence emerges. In their studies, some researchers (authors) have riveted on the classification of trends and methods in scholastic literature pertinent to definite SMR platforms such as Facebook and others(Bredl et al., 2012; Błachnio, Przepio´rka & Rudnicka, 2013; Caers, De Feyter, De Couck, Stough, Vigna & Du Bois, 2013; Dhir, Buragga & Boreqqah, 2013; Hew, 2011; Snelson, 2016; Terras & Warwick, 2013; Manca & Ranieri, 2013; Nadkarni & Hofmann, 2012; Wilson, Gosling & Graham, 2012; Williams).

On the other hand, some studies were substantiated within the precise subject of study to explore SMR. The existing literature reviews are shown in Table 1 to signify the work previously undertaken regarding SMR trends. However, there is little information about social media research (SMR) sampling techniques, data collection, analysis, and discussion

trends. It is evident that the preceding literature reviews have provided trends in discourse in research methods but have offered a global categorisation of universal trends (Snelson, 2016), whereby an author provided an overview of the diverse tacit approach for studying SMR. The contribution of the study focused on the categorisation between statistical, computational, and ethnographical perspectives. The inquiry elucidated the use of YouTube, Facebook, and Twitter studies.

Some authors' contributions focused mainly on the use of Twitter. The study presents a relative study validating quantitative approaches that compare microblogging platforms' communication practices and forms. Based on the research findings, a distinct type of discussion was acknowledged. Therefore, some forms of metrics concerning communications on Twitter were studied. In the following article, another author's study concentrated on the 'social' in social media research and analysed communities' aspects. Employing an analytic framework begins with the digital tracking of every user's life on the web. Particularly, it is conceivable to observe the communication practice in the networks. These emerging approaches induce different challenges, that can be tackled by applying ground-breaking methodologies, as in the case of this chapter and book. The discovery of "proximity in groups" is the foundation for future network analysis.

Academic researchers continue to discover the underlying literature linked to topical network analysis to study and understand online activity. In the author's contribution, the topical networks are units of analysis associated with the big datasets. By applying qualitative and quantitative methods, the method can complement large datasets' analysis to deliver additional facts into what is being tracked. This approach enables scholars to unravel why and when connections are established. Therefore, the findings emerging from the discovery depict an important method for refining the insight into online communication structures. Several academic authors' research focuses on analysing large bodies of data, and media use patterns, looking at the effect of relationship-related content and information, how YouTube can be used for seeking learning and information between young people, and the similar topic often not recognised. Therefore, the study believes such an approach could be able to deliver information that is valuable about the use, which is often not recognised in social media's dominant research.

These authors discuss the methods of archiving web data in a qualitative SMR. The authors depict that 'social media archives (for example, set of posts, text, and comments on different social media platforms) contain important datasets for exploring the social media users' communication pattern. They present a theoretical exploration of Internet (web) archives and provide an example-based analysis and discussion of the technical, methodological, and ethical challenges of analysing and mining archived social media data. Therefore, it is crucial to note that these literature reviews increase knowledge level or base concerning how the methods and sampling techniques were used in SMR. There are different reasons why such might be essentially important today. For one, SMR can be considered a new research area or field that emerged because of the usage of social media, its applications (technologies), and

its growing usage (Duggan, Ellison, Lampe, Lenhart & Madden, 2015). Therefore, little is known about how many approaches to conducting SMR are in the academic literature, where and how they originated, and scholarly journals or publishers that keep publishing them. Furthermore, the trends in choosing the right research design, data collection techniques or instruments, and analytic approaches are still few or not well known. Therefore, this chapter attempts to explain further how SMR is carried out (Snelson, 2016).

SAMPLING TECHNIQUES IN SMR

Sampling is a statistical process used to predetermine the number of persons or observations within a population to participate in a study. According to the study, sampling provides the subset of the population to estimate the attributes of the population. Sampling is essential in carrying out research, but it remains challenging for SMR projects. According to Piña-García, Gershenson and Siqueiros-García (2016), conducting large-scale research such as social media is challenging in determining a suitable sampling and concise size. Social media platforms are housing billions of data daily and analysing these data can be challenging and unachievable for any researcher. Academic researchers are continuously researching social media events. Based on the amount of data available on social media, researchers are using them in their study, but the determination of the appropriate population sample and size is vital. According to this chapter, sampling is an essential part of any research in determining the necessary population and sample size. In determining a sampling, the properties such as "user age distribution, net activity, net connectivity, and node degree" are used (Piña-García et al., 2016: 3). There are several sampling methods (types of sampling) that can be used on the SMR as presented in the next section of the chapter.

Sampling methods (types of sampling) for SMR

Here are some of the sampling methods that can be used on the SMR. Research can be conducted using the Internet and social media platforms. Whether probability or non-probability sampling, it can be done using the social media platform. According to Fricker (2008), sampling using the Internet in form of an email address requires only participants' contact details. Different types of Internet-based surveys can be done on probability and non-probability sampling (Fricker, 2008), the same types as intercept (pop-up), list-based sampling frame, non-listed-based random sampling, mixed-mode survey with Internet-based option, pre-recruited panel surveys within probability sampling while unrestricted self-selected survey, entertainment polls, volunteer (opt-in) panels and harvested email lists (data) are categorized within non-probability. These types of Internet-based surveys and others below can be applied to social media research. There are two types of sampling in SMR. The probability and non-probability sampling.

Probability sampling

Probability sampling is a sampling method that allows the researcher to set criteria to determine the population randomly. This method provides the opportunity for participants with an equal chance to be selected with selection criteria. This sampling method can be

applied to SMR to determine the criteria for the sample population (participants) to be included in the study.

Types of probability sampling
a.   *Simple random sampling:* This is the most used probability sampling method to randomly obtain data from the sample population. In this technique, every sample population has an equal opportunity to be selected. For example, in a study with a total population of 300 people, there is a probability that everyone will be randomly selected. The chance of everyone being chosen is possible in the study.

b.   *Systematic sampling:* In this sampling method, the researcher select sample size on a sample interval basis. The researcher selects the participants through sample intervals, for example, with a population of 3000 (N) and a sample size of 300 persons (n), the researcher selects the 10th person (k) as part of the study (total population /sample size = k (3000/300 = 10)).

c.   *Stratified random sampling:* This sampling method divide the total population into small groups to draw findings and conclusion. For instance, in a study on social media fake news, the research may group participants according to the effects, practical examples, and challenges of social media fake news. The grouping ensures that meaningful data are collected from each group to define practical solutions for fake news on social media.

d.   *Cluster sampling:* Through this sampling method, the research divides the total population into sections. For example, in a study on social media fake news in a particular, the researchers might divide the participants based on country, gender, education, and others to better gain insightful findings.

e.   *Intercept (pop-up), river, or real-time sampling:* This sampling is used in a systematic study to collect data on visitors who visit a website which is commonly used for the customer-satisfaction study (Fricker, 2008). The author believed that this type of Internet-based survey is non-response in nature and it is restricted to those who visited a particular web page and the visitor's Internet Protocol (IP) addresses (Fricker, 2008). The methods provide the survey post or information to the respondents using the web link, ads, and others (Liu, 2016). This kind of Internet-based survey can be used on social media platform to conduct SMR. Some researchers like Lehdonvirta, Oksanen, Räsänen and Blank (2021) view this sampling method as non-probability sampling. The authors further believed that this sampling method recruited participants through invitation using a web link. The link could be shared on the website, email, or other social media platforms like WhatsApp, Facebook, Twitter, and Telegram. Respondents follow the link to participate and share with others. Members of the population recruit others to join in the survey. However, this sampling method can be biased on coverage because not alltargeted populations will be represented (Lehdonvirta et al., 2021) because of lack of access to the Internet, data, and other forms of digital divides.
Researchers in information systems (IS) can use this method to recruit participants through social media platforms, web links, and so many others. Then, participants further recruit others to the study. Social media platforms today are actively used by

billion globally and it presents a perfect opportunity to conduct research. However, the researcher will only stop collecting data when enough data is collected.

f. *List-based sampling frame:* It seeks the email or social media handles of the participants. However, when the email details of the populations are not available the sample can be homogeneous like the universities, organizations, and government organizations and many other emails can be used (Fricker, 2008). The collected email details can be a source to recruit, share and conduct the survey.

When the sample is complicated, stratified sampling can be used. Stratified sampling is applied in dividing the sample population into small groups known as strata. For example, a researcher who wants to conduct a particular situation on the social media platform, like the fake news of the COVID-19 pandemic. The researcher need**s** to identify the particular region to consider as a country. To start. profile information on the social media account of the users in that country. For example, in a study that has a population of real-estate journalists as the participants, the researcher can assemble the publication necessary first and contact the editors of the publications (Fricker, 2008). The kind of study can be done on stratified samplingon the Internet.

g. *Non-listed-based random sampling:* This method is used without determining the sampling frame. Social media research and survey can be regarded as non-listed-based because the sampling frame (population) is not known. Fricker (2008) Internet-based and social media is non-list-based sampling. The study on social media or Internet-based research lies on the respondents. Traditionally, this type of Internet-based survey is used on telephone surveys for recruitment and others.

h. *Mixed-mode survey with Internet-based option:* This survey method involves the use of different means than email. Fricker (2008) states that telephones or malls could be used to collect data. Through this survey, the researchers will be able to collect data from different modes to accommodate people using other means, for example, a researcher collects data using Twitter but involves the use of Facebook, WhatsApp, and others at the same time.

i. *Pre-recruited panel surveys:* This involves potential participants who accepted to participate in the study and can be recruited through different means like email, web, or telephone (Fricker, 2008). This kind of Internet-based survey can be used on social media platforms where the participants can be informed about a particular study using social media platforms.

Non-probability sampling

This is sampling method is used to select participants without any fixed criteria for selection. It shows that participants (sample size) are selected non-randomly or subjectively. But the outcome from the non-probability sampling method may not represent the expectations of the sample size or the population because the selection criteria are subjective to the researcher. Etikan and Bala (2017) believed that non-probability is costly to be implemented. According to Showkat and Parveen (2017), the non-probability sampling method is judgmental because the researchers select participants who are easily accessible. It is convenient to use because only those that are available and accessible are considered. For example, on SMR, a set of

sample sizes are available and have the opportunity to be included, but only those who are easily accessible will be contacted. In this study, the non-probability sampling method is the opposite of probability sampling.

Types of non-probability sampling
   a. *Web panels:* This type of non-probability sampling method is used to conduct online or offline studies like social media research. It is used to select participants who are willing to engage in an online questionnaire. When using web panel sampling, all the participants must have social media account and a workable email address. Participants are sourced online using social media, email, phone addresses, banners, posters, group members, websites, community forums or sites, and other channels. While participants can be sourced offline using magazines, newspapers, radio, television, customer registers, outdoor poster, and others. There are sourced or recruited to participate in an online questionnaire or survey. This method is suitable for SMR because it allows researchers to recruit a large number of willing participants on social media platforms.
   b. *Convenience sampling:* This method applies the principle of reachable. The researcher has no authority or influence on whom to select, rather, the selection is purely based on accessibility and proximity. Only those who are convenient will be selected. For example, in a study conducted on social media impact on students, the researcher invited participants or share a questionnaire randomly and only those who are reachable and convenient to participate will be selected. However, this non-probability method is used when the budget is low and other limited resources and there is enough time to conduct the study.
   c. *Quota sampling:* This is a rapid method for gathering samples with the principle of pre-set criteria or standards in choosing participants in a study. The m.ethod is used in ISR in providing a set of attributes and criteria upon which participants must qualify. However, the attributes of the sample size must march the same attributes of the total population. As IS discipline and research keep changing and innovating, quota sampling is suitable for managing such changes by speeding up the sampling process.
   d. *Judgmental or purposive sampling:* The researcher applies his or her discretion in determining the sample size. This method is used for a deep understanding of the research objective and the researcher applies personal discretion to select those who can provide the data. For example, when a researcher wants to understand the effects of social media bullying, the researcher can only select those who have experienced bullying on social media. Those who have no experience with social media bullying will be excluded.
   e. *Snowball sampling:* Controlling or determining the sampling frame could be challenging. However, when a study targets the entire population, coverage errors are used (Fricker, 2008); while Baltar and Brunet (2012) recommend that snowball sampling can be used on "hard to reach" samples as in the case of social media platforms. SMR can be challenged in determining the population and sample size, then a sampling method that can cover a wide population should be adopted. Hard to reach is a situation where the population of the study could be hard to participate. Social media

platforms have increased hard to reach research population and making it challenging to get the population that can constitute the sample size. According to Baltar and Brunet (2012), social media especially Facebook have increased "chain-referral" techniques in research questionnaires through snowball sampling. Snowball sampling is a non-probability sampling method. Snowball involves participants sharing research invitation web links with others in the target population (Dusek, Yurova & Ruppel, 2015). Snowball sampling or chain-referral sampling is a method that focused on referrals. It means that the data is incremental, or the data size increases continuously till enough data is collected. The method is used in research when the population size is known or difficult to assemble.

To assemble and reach the target population, each participant (a node) adds another person till the node completes (Wagner, Singer, Karimi, Pfeffer & Strohmaier, 2017). For example, through snowball, researchers continue to increase data size and collect data till enough is collected for data analysis. SMR has no sampling frame, .and the total population is unknown and difficult to determine. In this case, snowball sampling methods can use in collecting data till the researcher is satisfied with the collected data.

f. *Random walk (RW) sampling:* This sampling method involves starting with a node and going to the network beginning with the random node and selecting a one-step link randomly and navigating through the network (Wagner et al., 2017). RW as a sampling process, the random movement expands in the same direction towards achieving one goal. Each node is independent of the other with one purpose and discovering an unknown network. Researchers using RW randomly chose a node and it expand in the same directions in the data collection. According to Wagner et al. (2017), this sampling method is used in conducting social media research such as Twitter, Facebook, and many others. Researchers can use this method to collect data and analyze it and be able to reach enough data.

g. *Unrestricted self-selected survey:* This kind of survey can be posted on the website and is open to the public and allows everyone to participate (Fricker, 2008). The openness to the public makes it part of convenience sampling with the ability to be generalizable. However, Fricker (2008) believes that negatively, the respondents may not be eligible (possessing not the right experience) to participate.

h. *Entertainment polls:* This is unscientific polls involving a survey done for entertainment through a website and anyone can participate in the survey by visiting the survey website post (Fricker, 2008). However, it can be adopted for SMR in looking at various issues in the research area.

i. *Volunteer (opt-in) panels:* This method is like the pre-recruited panel, but it is different in that participants choose to participate willingly (Fricker, 2008). SMR is usually voluntary and individuals participate willingly. Then, this method is suitable for SMR because of its ability to accommodate a large number of volunteers. Individuals decide to participate willing, and the survey is usually conducted on market research to gather customers' views (Fricker, 2008).

j. *Harvested email lists (data):* According to Fricker (2008), harvested email deals with the collection of email addresses through posting on the Internet. The emails are

collected from the people informed or uninformed; however, some email brokers sell the email addresses of the people to willing buyers. Through this process, social media handles (account names) of people could be harvested with or without the knowledge of the users for the purpose of conducting SMR.

According to Piña-García et al. (2016: 3), Metropolis-Hastings Random Walk (MHRW) provide the node for a candidate which shows: "q(y|x) = {Brownian, Illusion, Reservoir}":

1. **Brownian walk:** this is normal distribution working with node on Java language of Math. random.
2. **Illusion spiral:** it presents a geometric shape showing sequence of points "plane such that they are equitably and economically spaced" as shown in Figure 7.1. The formula: z ← az + bz/ |z| can be used.

Figure 7. 1: Pattern visualisation of the illusion spiral, adopted from Piña-García et al. (2016)
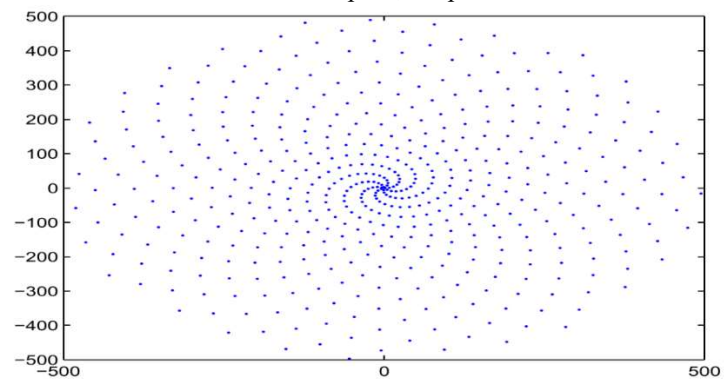
3. **Reservoir sampling** uses algorithm random sampling for *n* as the size. However, the *N* is not known in the algorithm. While the selected item (k) is fewer than the source array S(i).

DATA COLLECTION METHODS FOR SOCIAL MEDIA RESEARCH (SMR)

An empirical study cannot be complete without data collection. Data collection is a process of collecting evidence to answer a given research question/s.Kern, Park, Eichstaedt, Schwartz, Sap, Smith and Ungar (2016) believed that having the correct data is normally better than having a great deal of data, which is more reason to bear in mind as a researcher when conducting any kind of analyses or an SMR, it necessitates that data ought to be gathered. Careful deliberation must be directed towards which type of data can be more relevant and applicable for any inquiry at hand and whether the availability of useful data is attainable and available or accessible. Many social media platforms provide researchers to have access to data, whereas the costs and the amount of information accessibility differ. While some data might not necessarily be obtainable from companies like Google and Microsoft will not able to share users' private information through search query data, emails, and others. These challenges can be attributed to privacy issues that remain a concern and have continued to emerge in regulatory policies. Data access possibilities will remain a challenge in the foreseeable future. Consequently, it is vital to assert that when planning to

conduct an SMR, it may be helpful to consider the flexibility in terms of the sampling techniques and data collection methods to analyse and interpret the findings based on the questions asked.

Data forms a necessary part of the research process, and obtaining data involves unique techniques for researchers and organisations. Data can be sourced from "value-added data, cleaned data, raw data, a real-time feed, and holistic data" (Batrinca & Treleaven, 2015). This data can be collected for business and academic purposes. According to Alfantoukh and Durresi (2014), data collection is an essential technique or step in any given research. Data collection can be regarded as a process of obtaining valuable information for research. Presently, social media platforms are the most significant data source for researchers (Alfantoukh & Durresi, 2014), and scholars are beginning to explore these platforms for research data collection. Social media platforms house a vast amount of user-generated content (data) because of billions of active users daily. The user-generated content deposited on social media platforms has gained scholars' and researchers' attention to explore the platforms' roles in humans and society.

The platforms provide access to data collection. The benefits of collecting data using social media are encouraging because it limits biasness, open sampling, tracking, and directing sampling (Mirabeau, Mignerat & Grangé, 2013). Social media data allows easy accessibility of raw and massive data and longitudinal studies (Oh, 2012). According to Snelson (2016), several research studies on the contents of social networking activities conducted revealed in 2015 that 1.96 billion users are active on these platforms, with an estimated 2.44 billion users predicted in 2018, and the increase continues. WhatsApp, Facebook, YouTube, and Twitter were the most recognised and encouraged among various social networking sites. These platforms were popularly ranked among the top 10 most profoundly visited and used sites on the Internet (Snelson, 2016). The grouping of prolific individual activities and the creation of user-generated information and content have engrossed academic attention who pursue to appreciate SMR and its protagonist in modern society. Oh (2012) revealed that there are various kinds of data collection methods used in research. For SMR, content analysis is the second most applied SMR data collection method on YouTube, Twitter posts (Tweet), Facebook posts, WhatsApp chats, and many others as a data sources (Snelson, 2016; Drula, 2012).

SMR data collection methods
Data collection methods for SMR can involve name generation, name interpretation, and mining data (Ricken, Schuler, Grandhi & Jones, 2010).
- **Name generation:** This data collection focuses on recording names and relationships rather than understanding the "social ties" (Ricken et al., 2010). This is a form of "diary method" of collecting data; for example, participants record people they interact with and contact over a period. This method works well for a small group of people in a social media circle of friends.

- ▪ **Name interpretation:** Scholars use this method to gather deeper information from the participants, such as "relationship types, context, communication methods, frequency of contact," and many more (Ricken et al., 2010: 2). These details assist the research in understanding social connection, social class, strengths, and others. This data collection method can allow a researcher to gain respondents' information by connecting friend-of-friends on the social media platform.
- ▪ **Mining data:** Different data can be collected on social media, such as interpersonal connections, organizational and online mass communication, or interaction (Ricken et al., 2010). The social media friend list sets name generation in understanding friends' social relationships and all that through mining data.

Conventional data collection techniques (methods)

Conventionally, there are different types of data collection techniques (methods) that are categorized according to qualitative and quantitative research methods. Qualitative data collection techniques are interviews, observation, focus groups, and case studies, while quantitative data collections are questionnaires. Each of these techniques whether qualitative or quantitative can be collected on SMR. The researchers discuss the data collection methods suitable for SMR.

Qualitative data collection methods
a. *Interview:* an interview involves the collection of participants' opinions, beliefs, expressions, experiences, and ideas (Parveen & Showkat, 2017). Interview questions are presented to participants who are allowed to provide detailed information on a particular question. Participants express themselves with deeper information on a given phenomenon. Interview questions are categorized into structured (predetermined questions are given and participants are restricted to the questions), semi-structured (open-minded questions, giving the interviewer or interviewed to divert to any ideas not included in the list of questions) and unstructured (no designated range of questions to ask).
b. *Observation:* this qualitative data collection method allows researchers to observe the situation as it happens. Parveen and Showkat (2017) suggest that observation is used to deeply understand phenomena under investigation to observe and record the observed results. For example, a researcher can observe things or events on social media and record the results.
c. *Focus group:* this method gathers groups of people to engage in a discussion relating to the research underway. This method gathers participants' expressions, opinions, ideas, views, and reactions within the research under investigation (Parveen & Showkat, 2017). While Canals (2017) believe that participants are invited to engage and participate in a research discussion. For example, social media groups can be a starting point to collect data on a particular issue under study.
d. *Case study:* it involves a deeper understanding of actors such as occurrence, events, situation, group, and person (Parveen & Showkat, 2017). It is used to understand the social situation by providing detailed information and probing a situation. For instance,

to understand the impact of social media on the social well-being of students, SMR can apply a case study to explore the situation.

Quantitative data collection methods
a. *Questionnaire/survey:* It is a method used to collect data from a large group of people. It is widely used in quantitative research in reaching a wider audience. A questionnaire/survey can be done online or offline. In the case of SMR, a researcher can use this method to cover a range of topics on social media and reach a wider audience.
b. *Observation:* its method which can be used to collect quantitative data in counting and recording the number of occurrences of event/s or phenomena.

DATA ANALYSIS METHODS FOR SOCIAL MEDIA RESEARCH (SMR)

SMR data is collected through posts, words, and text. In instances where the use of a social media platform is deployed as a data collection method, Twitter, as illustrated in Figure 7.2, demonstrates how to name generation and tagging, to mention but a few, are used in SMR.

Figure 7.2: Tweet and contents, adopted from Highfield and Leaver (2015)

The process of using a Twitter account is that, after a dataset is deemed appropriate, the data can be obtained (downloaded) and used. Social media data is classically retrieved over an "application programming interface (API)", which is often regarded as the layout that identifies whether the posted words are copyrighted and used by the "American Psychological Association" or one of its associated publishers. Furthermore, APIs are similar to converters or translators that enable application designers to generate applications on their systems and effortlessly disseminate the content to clients, improving the client or user's experience. An essential component of the APIs is that it offers platforms that enable research scholars to generate massive amounts of information without interfering with the user's activities and experiences. APIs allow for easy and accessible big data in terms of live streaming data in an approach that cannot be suitable via the Internet browser like any other download and streaming method. Importantly, there is a growing number of data analyses on SMR.

Methods of data analysis in SMR

Different methods and techniques are applied in analyzing data in SMR. Some of them are discussed in the succeeding paragraphs.

Content analysis: content analysis is a qualitative research method used to analyse voice recording. According to Bengtsson (2016), content analysis brings together the meanings of the data collected to conclude. The decision on selecting an appropriate data collection method is important for researchers towards content analysis credibility (Elo, Kääriäinen, Kanste, Pölkki, Utriainen & Kyngäs, 2014). A content analysis means interpreting what is contained in qualitative research data collected especially through interview recording and existing documents. It is done with quantitative or qualitative, or mixed research methods (Snelson, 2016). According to Hsieh and Shannon (2005) and Drula (2012), content analysis is the most used qualitative data analysis technique. By applying content analysis, researchers analyse social media platforms' text, posts, and videos. The analysis seeks to understand, describe, and quantify a situation (Elo et al., 2014). The process involves data cleaning (analysis) and interpreting into meaning. Based on the attribute, content analysis is always unstructured, which is seen in social media data. The unstructured nature of SM data makes content analysis necessary, and appropriate. Content analysis can be done manually (human coding) or using computer-aided text analysis (CATA). There are steps in content analysis (Luo, 2019) of SMR:

   a. **Content selection:** This involves the selection of social media data content to analyse. The basis for content selection depends on the research topic, the problem under investigation, and the question.
   b. **Define units and category of analysis:** The unit looks into the frequency of the social media text or words. According to Pulido, Redondo-Sama, Sorde´-Martı and Flecha (2018), the units involve all the content organized and analyzed together and not individually.
   c. **Define rules for coding:** Coding deals with organising the defined social media units into categories. To ensure consistency, transparency, credibility, and data reliability, it is necessary to determine the rules guiding the included and excluded social media text or words.
   d. **Code by the rules:** Assess every social media text or word and record all necessary data that meets the categories. The coding can be done manually or using computer programs such as Atlas.Ti, NVivo, and others.
   e. **Analyse data findings and make recommendations and conclusions:** After coding, the collected and coded social media data is evaluated and interpreted according to the study and question. Furthermore, the recommendation is made to stakeholders, and a conclusion is drawn from the findings.

According to Hsieh and Shannon (2005), there are three kinds of content analysis, which are conventional content analysis used to describe and interpret a given phenomenon; directed content analysis which covers the validating and extending existing theory or phenomenon; while summary content analysis focuses on "identifying and quantifying" text or words to understand the actual meaning behind the text or words. The summary of content analysis goes beyond words or text in quantifying and counting the frequency of occurrence used in quantitative research. Also, summary content analysis involves latent content analysis, which covers the content's interpretation and unfolding of the meaning behind the text and words

(Hsieh & Shannon, 2005). These content analysis types can be applied to social media research in unfolding events taking place on social media platforms.

a. **Semantic analysis:** It involves the extraction of meaningful content from a trending issue rather than the frequency of the text or words (Bayrakdar, Yucedag, Simsek & Dogru, 2020). This is achieved by unlocking meaning from social media content. To achieve semantic analysis, text or word, image, and video analyses are performed (Bayrakdar et al., 2020). Semantic text analysis allows the organisation of written words or text based on the relationship associated with each other.

b. **Network analysis:** This method is used when researching social media platforms (Drula, 2012). Text, images, and videos are analysed using network analysis in mining the activities and connections among bloggers. As a mining technique, network analysis is grounded in social media analysis, web mining, games theory, network theory, and graphs theory. It allows the analysis of a large amount of data (Drula, 2012). This shows that network analysis can be used in social media data mining and analysing text, images, and video, whether they are trending.

c. **Social media analysis:** It is a method used to collect and analyse data from different social media platforms like Instagram, Twitter, Facebook, WhatsApp, and many others. Data collection, monitoring, analysis, interpretation (discussion), summarising, and envisioning of social media data are performed (Fan & Gordon, 2014) to understand conversations, engagement, and interaction on the platforms. The application of social media analytics allows social media data analysis to extract intellectual content and pattern. Social media analytics can help academic researchers and businesses understand and access the instance insight, experience, and opinions of users and customers in monitoring hashtags and other trending conversations. On the other hand, social media analytics is among the new and emerging methods of analysing social media data (Pulido et al., 2018). It involves four steps, for example, "data discovery, collection, preparation, and analysis" (Stieglitz, Mirbabaie, Ross & Neuberger, 2018:1).

d. **Data mining:** This is a method used in collecting and analysing data. It is used to sort out the meaning behind a large dataset and identify the relationship that provides solutions to business challenges through data analysis. Social media data is different from the actual traditional data (Zafarani, Abbasi & Liu, 2014). However, the mandate of data mining in traditional research data is the same as social media data and research. The collected and analysed data informs the decision-making process. In this study, data mining can be called social media mining. The growth of social media data has created what is called social media mining. According to McCourt (2018) and Zafarani et al. (2014), social media mining involves interpreting, analysing, and extracting relevant content from social media data. It allows researchers and organisations to collect data and analyse it towards understanding the meaning to make a conclusion and inform decision-making. In social media research, data mining or social media mining is used to gather, analyse, and discuss (interpret) social media data to arrive at a conclusion that informs decision-making.

Social media research analysis and procedure

To fully address this chapter's main core objective, there is a need to offer guidance on conducting an SMR and approaches to adhere to for data sampling techniques, data collection, and analytics. This chapter appreciates the framework of social media analytics (SMA) proposed by Stieglitz (Stieglitz et al., 2018). An SMA framework was adopted to elucidate this procedure, which builds a mutual foundation for conveying social media analytics. Stieglitz et al. (2018) explain the advantages that could emerge of social media analytics and, therefore, deemed it fit to suggest a research framework approach for establishing the guidance and considering the relations among humanity, business organisations, and social media. According to the Stieglitz et al.'s (2018) framework, it entails four (4) practices that ought to be followed, and these practices relate to social media-kind of activities, and three (3) different stages of analysis which any research scholar could pay attention to when tentatively exploring these practices (see Figure 7.3).

On the contrary, a few research studies, for example, Fan and Gordon (2014) and van Osch and Coursaris (2014), discover a procedural practice for social media analytics encompassing three (3) stages "capture," "understand," and "present." They further explained that capturing entails assembling data and handling it, where information relevant to the underlying study is extracted from the data in this stage. This is followed by dealing with irrelevant or noisy information, and if such information exists, it should be removed. Conversely, this stage fundamentally entails using a key system approach, namely the sentiment analysis, or social network analysis for interpreting data, whereas, in the final stage, the discoveries are reported in summaries and presentations (Fan & Gordon, 2014). However, of all the framework discoveries or studies that have been conducted or proposed by academic scholars, the proposed SMA framework by Stieglitz et al. (2018)is among the preferred in the information systems (IS) discipline; this was judged based on the references or citation that the paper had received in information systems (IS) literature. As described by the authors, the adapted framework provides the SMA process as entailing three stages (see Figure 7.3). Figure 7.3 shows the process of conducting SMR.

Figure 7.3: Social media research analysis and procedure as adapted by Stieglitz et al. (2018)
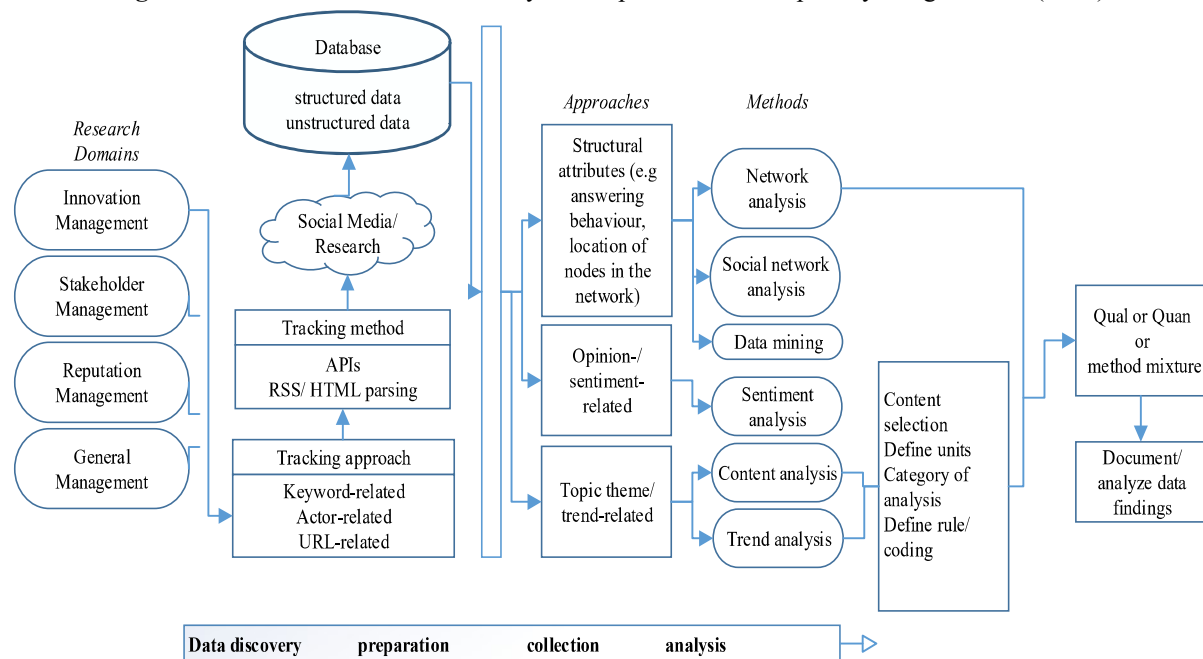
In this chapter, the researchers adapted the Stieglitz et al. (2018) SMA framework, though modified by including the data discovery and collection stage. They excluded the tracking stage that appeared after and before the tracking stage for the following justifications. The initial framework developed was grounded in a political context. In origin, it can straightforwardly be tailored for any other research investigation and different research areas. Nevertheless, the objectives and study methods are likely dissimilar, though the practical procedure and process are principally the same approach.

The framework above in Figure 7.3 depicts four different stages of conducting SMR, namely the data discovery, preparation, collection, and analysis. The first stage entails the data discovery phase, where predefined terms are searched for and stored in the database, either structured or unstructured, as shown in Figure 7.3 above. This stage allows tracking posts, tweets, and text or words used to establish their relevance. The searched words, text posts, and tweets are then selected in an attempt to classify the social media platforms chosen from the dimension of social media analytics that unambiguously mentions the tweets 'provided' or any other phrase that may be applied within the context of the study (see Figure 7.3). According to the study being probed, the researcher might decide to increase the search with additional synonymous search terms, which may be best known to the researcher. After this, the researcher may enhance the search terms iteratively and prepare the words to eliminate irrelevant or noisy information and embrace various pertinent ones. After the discovery and preparation, the researchers then applied various collection methods, as mentioned in the previous section of the chapter. There are various ways in which SMR data collection methods can be carried out. This approach deals with selecting the most appropriate method and strategies that suit the research investigation and inquiry. Next is the analysis stage; the original framework proposes that in the analysis stage, "statistical analysis, social network analysis, sentiment analysis, content analysis, and trend analysis". However, in Figure 3, the

researchers added a few analysis methods to the already existing framework. This included data mining in place of statistical analysis. The researchers also modified the content analysis by including content selection, defining units, category of analysis, and defining rules/coding.

Therefore, overall, the data discovery depicts the uncovering of the underlying latent structures and patterns. This stage includes choices of data source (Twitter and Facebook), data sampling techniques, approach, and output. A specific subdivision of this stage can be traced to Stieglitz et al. (2018). In numerous academic discoveries and studies, diverse Twitter sources' wholeness was compared (Driscoll & Walker, 2014; Morstatter, Pfeffer & Liu, 2014). The initial framework does not discuss the preparation stages needed. Finally, a data analysis depends solely on the research aim, comprising social network analysis, opinion mining, and many more.

DISCUSSION OF RESULTS IN SMR

Discussion of research findings is an essential aspect of the overall research process in any research, including social media research. It is a process expected of a researcher to translate the analysed data into a common understanding for the audience (Chukwuere & Chukwuere, 2020). It allows researchers to deduce findings from the analysed social media data and build meanings, informing audiences to understand the study better. It also involves extracting and interpreting the study's relevant importance with support from the existing literature (Gemayel, 2016). As an essential aspect of research writing, social media research discussions clarify the research findings and make them relevant to the audience by highlighting the research's implications, the alignment with existing literature, and paving the way forward for future studies.

For example, after data analysis, the findings of the collected data must be discussed in adding meaning to the data in line with existing literature. SMR findings are discussed and aligned with existing literature.

CHALLENGES, ETHICS, AND RELIABILITY WITH PROCESSING AND ANALYSING SMR

There are several challenges that any research scholar may experience when conducting SMR. The challenges can be:
  a. **Storage capacity or size:** For the preliminary handling of data, some category of server or database management system (DBMS) is desirable to store the raw data that necessitates memory and storage capacity or size. However, storing, and processing the tweets and their tokenized usages necessitated over one terabyte of storage capacity (Kern et al., 2016).
  b. **The language used:** Another issue or challenge that might surface relates to language use and vagueness; language could be ambiguous; and therefore, consideration can be given in particular on how to define "words" in instances where informal text punctuation follows specific rules, while in social media this can reproduce actual punctuation (for example, a period ending a sentence), the use of emotion to express

(for instance, emoticons), emphasise a point, mixture of languages (for example, use of English and home language or jargon) or expose typing errors and misspellings, alongside, overfitting, regularisation, variable assortment, and model error. In a model error, there is a definite degree of error happenings and is conceded throughout the procedure of formulating and data scrutinizing (Kern et al., 2016).

c. **Reliability and validity question:** Social science research is subjected to several errors that test the reliability of the data collected. According to Cheliotis, Lu and Yi (2015), reliability deals with the consistency of measurements used and each other, while correlations ensure measurement reliability. SMR is collected using social media platforms, while reliability and validity can be questioned. Validity focuses on determining whether the data collected is produced through a valid instrument or measurement. In social media research, methods are used to assess the reliability of data collected, for example, correlation and Cronbach alpha (Cheliotis et al., 2015). Correlation measures the relationship strength between two variables by applying Pearson's *rho* and Spearman's *rho*. These methods are applied to SMR data collection. Though these methods are commonly used on quantitative data, it is still a question of how to apply them to text, specifically qualitative research. Cheliotis et al. (2015) test correlation values are used to determine whether the collected text recorded the correct users' posts on a social media platform. The researcher suggested normalized edit distance (NED) ($NED = (l - \delta)/l$) and the longest common subsequence (LCS) ($LCS = m/l$). Mining and analysing social media research have some ethical and legal implications (McCourt, 2018).

d. **Ethical issues:** SMR is fast becoming a norm but there is a number of challenges to confront. The privacy, consent, and confidentiality of participants' details are in doubt in conducting social media research. The three issues and many others always confront the reliability and validity of the findings.

CONCLUSION

Social media datasets are available in the public domain, mined by applying different research approaches and methods. The chapter unpacked the processes involved in SMR sampling techniques, data collection, analysis, and discussions. The chapter presented different approaches that necessitate conducting SMR with a clear definition of steps. SMR stands as an emerging research field with huge attention from scholars across disciplines. This kind of attention can be attributed to the growing usage of social media platforms that keep collecting large datasets that can be mined and analysed for decision-making. This chapter paves the way for Chapter 8, which presents the guideline and principles that should be followed in SMR, emphasizing ethical conduct and compliance.

REFERENCES

Alberghini, E., Cricelli, L., & Grimaldi, M. (2013). KM versus enterprise 2.0: a framework to tame the clash. *International Journal of Information Technology and Management*, *12*(3-4), 320-336.

Alberghini, E., Cricelli, L., & Grimaldi, M. (2014). A methodology to manage and monitor social media inside a company: a case study. *Journal of Knowledge Management*.

Alfantoukh, L., & Durresi, A. (2014, September). Techniques for collecting data in social networks. In *2014 17th International Conference on Network-Based Information Systems* (pp. 336-341). IEEE.

Baltar, F., & Brunet, I. (2012). Social research 2.0: virtual snowball sampling method using Facebook. *Internet research*.

Batrinca, B., & Treleaven, P. C. (2015). Social media analytics: a survey of techniques, tools and platforms. *Ai & Society*, *30*(1), 89-116.

Bayrakdar, S., Yucedag, I., Simsek, M., & Dogru, I. A. (2020). Semantic analysis on social networks: A survey. *International Journal of Communication Systems*, e4424.

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus open*, *2*, 8-14.

Błachnio, A., Przepio ́rka, A., & Rudnicka, P. (2013). Psychological determinants of using Facebook: A research review. International Journal of Human-Computer Interaction, *29,* 775-787. doi:10. 1080/10447318.2013.780868.

Booth, A., Sutton, A., & Papaioannou, D. (2016). *Systematic approaches to a successful literature review*. Sage.

Boulianne, S. (2015). Social media use and participation: A meta-analysis of current research. *Information, communication & society*, *18*(5), 524-538.

Bredl, K., Hünniger, J., & Jensen, J. L. (2012). Methods for analyzing social media: Introduction to the special issue. *Journal of Technology in Human Services*, *30*(3-4), 141-144.

Caers, R., De Feyter, T., De Couck, M., Stough, T., Vigna, C., & Du Bois, C. (2013). Facebook: A literature review. *New media & society*, *15*(6), 982-1002.

Camacho, D., Luzón, M. V., & Cambria, E. (2020). New trends and applications in social media analytics.

Canals, L. (2017). *Instruments for Gathering Data*. Research-publishing. net. La Grange des Noyes, 25110 Voillans, France.

Cheliotis, G., Lu, X., & Yi, S. (2015). Reliability of Data Collection Methods in Social Media Research. In *ICWSM* (pp. 586-589).

Chukwuere, J. E., & Chukwuere, P. C. (2020). Guide for Postgraduate Student Research and Publications: A Step-by-Step Approach. In *Handbook of Research on Connecting Research Methods for Information Science Research* (pp. 529-550). IGI Global.

Dhir, A., Buragga, K., & Boreqqah, A. A. (2013). Tweeters on campus: Twitter a learning tool in classroom?. *J. UCS*, *19*(5), 672-691.

Driscoll, K., & Walker, S. (2014). Big data, big questions| working within a black box: Transparency in the collection and production of big twitter data. *International Journal of Communication*, *8*, 20.

Drula, G. (2012). Social and online media research–data, metrics and methods. *Review of Applied Socio-Economic Research*, *3*(1), 77-86.

Duggan, M., Ellison, N. B., Lampe, C., Lenhart, A., & Madden, M. (2015). Social media update 2014. *Pew research center*, *19*, 1-2.

Dusek, G. A., Yurova, Y. V., & Ruppel, C. P. (2015). Using social media and targeted snowball sampling to survey a hard-to-reach population: A case study. *International Journal of doctoral studies*, *10*(1), 279-299.

Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). Qualitative content analysis: A focus on trustworthiness. *SAGE open*, *4*(1), 2158244014522633.

Etikan, I., & Bala, K. (2017). Sampling and sampling methods. *Biometrics & Biostatistics International Journal*, *5*(6), 00149.

Fan, W., & Gordon, M. D. (2014). The power of social media analytics. *Communications of the ACM*, *57*(6), 74-81.

Ferron, M., Massa, P., & Odella, F. (2011). Analyzing collaborative networks emerging in Enterprise 2.0: The Taolin Platform. *Procedia-Social and Behavioral Sciences*, *10*, 68-78.

Fricker, R. D. (2008). Sampling methods for web and e-mail surveys. *The SAGE handbook of online research methods. London: SAGE Publications Ltd*.

Gemayel, R. (2016). How to write a scientific paper. *The FEBS journal*, *283*(21), 3882-3885.

Hansen, D., Shneiderman, B., & Smith, M. A. (2010). *Analyzing social media networks with NodeXL: Insights from a connected world*. Morgan Kaufmann.

Hew, K. F. (2011). Students' and teachers' use of Facebook. *Computers in human behavior*, *27*(2), 662-676.

Highfield, T., & Leaver, T. (2015). A methodology for mapping Instagram hashtags. *First Monday*, *20*(1), 1-11.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research*, *15*(9), 1277-1288.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, *53*(1), 59-68.

Kern, M. L., Park, G., Eichstaedt, J. C., Schwartz, H. A., Sap, M., Smith, L. K., & Ungar, L. H. (2016). Gaining insights from social media language: Methodologies and challenges. *Psychological methods*, *21*(4), 507.

Kivunja, C. (2015). Innovative methodologies for 21st century learning, teaching and assessment: A convenience sampling investigation into the use of social media technologies in higher education. *International Journal of Higher Education*, *4*(2), 1-26.

Koo, C., Wati, Y., & Jung, J. J. (2011). Examination of how social aspects moderate the relationship between task characteristics and usage of social communication technologies (SCTs) in organizations. *International Journal of Information Management*, *31*(5), 445-459.

Lee, C. S., & Ma, L. (2012). News sharing in social media: The effect of gratifications and prior experience. *Computers in human behavior*, *28*(2), 331-339.

Lehdonvirta, V., Oksanen, A., Räsänen, P., & Blank, G. (2021). Social media, web, and panel surveys: using non-probability samples in social and policy research. *Policy & internet*, *13*(1), 134-155.

Liu, M. (2016). Comparing data quality between online panel and intercept samples. *Methodological Innovations*, *9*, 2059799116672877.

Luo, A. (2019). What is content analysis and how can you use it in your research. *Scribbr, July 18th*.

Manca, S., & Ranieri, M. (2013). Is it a tool suitable for learning? A critical review of the literature on Facebook as a technology-enhanced learning environment. *Journal of Computer Assisted Learning*, *29*(6), 487-504.

McCourt, A. (2018). Social Media Mining: The effects of big data in the age of Social Media. *Media Freedom and Information Access Center, Yale Law School*.

Mirabeau, L., Mignerat, M., & Grangé, C. (2013). The utility of using social media networks for data collection in survey research.

Morstatter, F., Pfeffer, J., & Liu, H. (2014). When is it biased? In Proceedings of the 23rd International Conference on World Wide Web – WWW '14 Companion 555-556. https://doi.org/10.1145/2567948.2576952.

Nadkarni, A., & Hofmann, S. G. (2012). Why do people use Facebook?. *Personality and individual differences*, *52*(3), 243-249.

Oh, S. (2012). Methods of Social Media Research: Data Collection & Use in Social Media.

Ohei, K. N. (2019). Integration of social media technologies and applications to serve as blended approaches to traditional teaching and learning method: a case study of South African universities. *International Journal of Social Media and Interactive Learning Environments*, *6*(2), 150-167.

Ohei, K. N., & Brink, R. (2019). The use of instructional web technologies in higher education systems to support students' academic success. *Mousaion*, *37*(2), 1-27.

Ohei, K. N., & Brink, R. (2019). Web 3.0 and web 2.0 technologies in higher educational institute: Methodological concept towards a framework development for adoption. *International Journal for Infonomics (IJI)*, *12*(1), 1841-1853.

Parveen, H., & Showkat, N. (2017). Data collection. *Pathshala Quadrant-I (e-Text)*, 17-18.

Piña-García, C. A., Gershenson, C., & Siqueiros-García, J. M. (2016). Towards a standard sampling methodology on online social networks: collecting global trends on Twitter. *Applied network science*, *1*(1), 1-19.

Pulido, C. M., Redondo-Sama, G., Sordé-Martí, T., & Flecha, R. (2018). Social impact in social media: A new method to evaluate the social impact of research. *PloS one*, *13*(8), e0203117.

Ricken, S. T., Schuler, R. P., Grandhi, S. A., & Jones, Q. (2010, January). TellUsWho: Guided social network data collection. In *2010 43rd Hawaii International Conference on System Sciences* (pp. 1-10). IEEE.

Showkat, N., & Parveen, H. (2017). Non-probability and probability sampling. *Media and Communications Study*, 1-9.

Sivarajah, U., Irani, Z., Gupta, S., & Mahroof, K. (2020). Role of big data and social media analytics for business to business sustainability: A participatory web context. *Industrial Marketing Management*, *86*, 163-179.

Snelson, C. L. (2016). Qualitative and mixed methods social media research: A review of the literature. *International Journal of Qualitative Methods*, *15*(1), 1609406915624574.

Stieglitz, S., Dang-Xuan, L., Bruns, A., & Neuberger, C. (2014). Social media analytics-an interdisciplinary approach and its implications for information systems. *Business & Information Systems Engineering*, *6*(2), 89-96.

Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). Social media analytics–Challenges in topic discovery, data collection, and data preparation. *International journal of information management*, *39*, 156-168.

Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). Social media analytics–Challenges in topic discovery, data collection, and data preparation. *International Journal of Information management*, *39*: 156-168.

van Osch, W., & Coursaris, C. K. (2014). Social media research: An assessment of the domain's productivity and intellectual evolution. *Communication Monographs*, *81*(3), 285-309.

Wagner, C., Singer, P., Karimi, F., Pfeffer, J., & Strohmaier, M. (2017, April). Sampling from social networks with attributes. In *Proceedings of the 26th international conference on world wide web* (pp. 1181-1190).

Williams, S. A., Terras, M. M., & Warwick, C. (2013). What do people study when they study Twitter? Classifying Twitter related academic papers. *Journal of Documentation*.

Williams, S. A., Terras, M. M., & Warwick, C. (2013). What do people study when they study Twitter? Classifying Twitter related academic papers. *Journal of Documentation*

Wilson, R. E., Gosling, S. D., & Graham, L. T. (2012). A review of Facebook research in the social sciences. *Perspectives on psychological science*, *7*(3), 203-220.

Zafarani, R., Abbasi, M. A., & Liu, H. (2014). *Social media mining: an introduction*. Cambridge University Press.