



## Digital Transformation, Big Data and Research Landscape in Digital Communication

<http://dx.doi.org/10.25008/jkiski.v3i2.220>

**Imam Syafganti<sup>1\*</sup>**

The Department of Business Administration, Jakarta State Polytechnic  
Jl. Prof. Dr. G.A Siwabessy, Kampus UI Depok, Indonesia

\*Correspondent author: [imam\\_syafganti@bisnis.pnj.ac.id](mailto:imam_syafganti@bisnis.pnj.ac.id)

**Submitted:** 05 August 2018, **Revised:** 24 December 2018, **Accepted:** 27 December 2018

Accredited by Kemristekdikti No. 30/E/KPT/2018

---

### Abstract

The digitization of communication technology has led to an intense interaction between human and digital-based technology. A large number of digital data traces were produced by humans as a result of the activity. Such data was commonly referred to as big data. The availability of the big data as a digital data source in turn opened opportunities for communication scientists to use the data to identify the patterns and trends of human activities in social research. It was necessary to understand the basic concept of the big data using appropriate tools and adequate access to the data and appropriate research method in order to be able to conduct research using the digital data. The paper aimed at describing the potential of the data for the purposes of communication research using appropriate tools, techniques and methods and at identifying potential research directions in the digital realm. Some limitations and critical issues related to the research validity, population and sample and ethics in digital media research method were also discussed.

**Keywords:** *big data, communication research, digitization, new method, research direction*

### Abstrak

Digitalisasi teknologi komunikasi telah mendorong terjadinya interaksi yang intens antara manusia dengan teknologi berbasis digital tersebut. Sejumlah besar jejak data digital diproduksi oleh manusia sebagai hasil dari aktifitas itu. Data tersebut lazim disebut sebagai Big Data. Ketersediaan Big Data sebagai sumber data digital tersebut pada gilirannya membuka peluang bagi para ilmuwan komunikasi untuk dapat menggunakan data itu untuk mendapatkan pola-pola dan kecenderungan dari aktifitas manusia yang selama ini dilakukan melalui penelitian sosial. Untuk dapat melakukan penelitian dengan menggunakan data digital itu diperlukan adanya pemahaman terhadap konsep dasar Big Data, menggunakan alat yang tepat dan memadai untuk mengakses data, serta metode penelitian yang tepat pula dalam pengolahan dan analisis. Tulisan ini bertujuan untuk mendeskripsikan potensi data digital bagi keperluan penelitian komunikasi, penggunaan alat, teknik dan metode yang tepat serta mengidentifikasi arah penelitian yang potensial dalam ranah digital ini. Selain itu dibahas juga beberapa keterbatasan dan isu-isu kritical terkait validitas, populasi dan sampel, serta etika dalam metode penelitian media digital.

**Kata kunci:** Big Data, Riset komunikasi, Digitalisasi, Metode baru, Arah penelitian.

## Introduction

The digital revolution has brought a great substantial change in many aspects of technology in human communication process (Kaul, 2012). The change was especially characterized by the proliferation of new inventions in the field of information and communication technology used and applied in various aspects of human life. One of the most phenomenal technological inventions in the changing environment was the invention of internet. The technology allowed the connection between a computer and other enormous computer networks in various parts of the world (Abbate, 2017), which stimulated the occurrence of both textual and virtual interactions among human and in turns facilitated global human communication in a more effective and efficient way (Matusitz, 2007).

The increasingly wide spreading use of the internet and its applications has changed the world we live today in a dramatic way and modified the world into an interconnected cyberworld (Tsou, 2011). Subsequently, such condition also encouraged more people in our society to get involved in various online communication activities. Sending and receiving e-mail, visiting websites, watching video in youtube, chatting in social media, updating status in social media, writing online reviews, and browsing information related products were some examples of popular online activities in our daily life.

According to Manyika, the above-mentioned online activities of human could produced a large number of digital traces and data (Lewis, Zamith, & Hermida, 2013), which recorded and stored permanently in the computer servers of particular providers that facilitated and enabled such online activities and services (Power, 2014). A simple example of how digital trace and data were produced by human daily activities was through human interaction in various platforms of social media. Everyday the users of social media across the world communicated each other by updating their status (in response to various current social issues, complaining or commending product performance, expressing particular feelings like happiness or sadness, lament, anger, responding to other person status, arguing on social issues, giving likes, etc.). Amongst other, there were several popular social media that widely used by users across the world such as Facebook, Instagram, Twitter and Google+. The social media produced petabytes of data on a daily basis (Housley, Williams, Williams, & Edwards, 2013). The

amount of the data would be bigger when we also took other digital data into account, which were produced by humans in various aspects of life such as health, education, socio-economic and politics.

In the context of technology and information study, the data was referred to as "big data". The term "big data" referred to a number of data sets that were very large in capacity and because of its huge size, the data could not to be processed using personal computers or simple software that was commonly used by most of computer users (Eynon, 2013).

The availability of such large amount of digital data traces has attracted the attention of various parties such as business organizations (Zhao, Fan, & Hu, 2014), governments (Eynon, 2013) and also various researchers of many disciplines (Boyd & Crawford, 2012). They tried to get access to the data and started to analyze the large digital datasets to understand general trends and patterns of human behavior (Tsou, 2015).

In the context of communication research, "the big data" has emerged as a new approach in collecting and analyzing digital data in internet (Boyd & Crawford, 2012). Furthermore, the phenomenon certainly provided communication scientists with many opportunities and benefits to explore and to find patterns of human communication in a more practical and concise manner (Mahrt & Scharkow, 2013). As a consequence, various initial studies related to "the big data" in the field of communication began to appear sporadically (Burgess, Bruns, & Hjorth, 2013).

One of the earliest studies that dealt with the importance of "the big data" in communication research was conducted by Papacharissi & De Fatima Oliveira (2012). In the study, they applied computational discourse analysis to assess the news on Twitter in Egypt related to the resignation of Hosni Mubarak as the president in the period of January - February 2011. Also, Burgess & Bruns (2012) began to suggest the use of the term "big data" in communication and media studies by explaining the possibility of using the Twitter API (Application Programming Interface) and Twitter archives as two important aspects in gathering digital data traces from Twitter. Other studies related to the topic were conducted by Lewis *et. al.*, (2013) who combined computational and manual methods in making content analysis of massive digital datasets.

Despite the importance and the potential use of the data in communication study, the use of "the big data" as a source of digital data in the communication study was still very rare in Indonesia. The situation was very unfortunate given that the development of internet and digital technology went very fast in Indonesia. Based on APJII (Indonesia Internet Service Provider) data, Indonesia had 132.7 million internet users (around 52% of the total population) with internet penetration rate of 34.9% (2016). Moreover, the data also indicated that social media, entertainment, news, education and commerce were the contents most frequently accessed by internet users in Indonesia. Consistent with the findings of the previous studies, other data also confirmed that there were approximately 79 million active social media users in Indonesia (wearesocial, 2016). Furthermore, the data also showed that Facebook, Twitter and Instagram were the most popular social media platforms widely used by Indonesians.

The aforementioned facts and statistics of the internet usage and the availability of digital data clearly showed that Indonesia was one of the countries whose citizens did massive activities in internet. As a consequence, Indonesia produced huge digital footprints that communication scientists could use as sources of digital data in predicting attitudes and patterns of human behavior related to certain aspects of life. Therefore, it was necessary to be able to handle the digital method in conducting internet-based research (data retrieval) (Hutchinson, 2016).

This paper aimed at generally describing the concept of "big data", the potential use of digital data traces for communication research in Indonesia in the context of popular social media (Facebook, Twitter and Instagram). Additionally, it would also try to briefly explain the tools and the methods to retrieve, to extract and to analyze data, to analyze sentiments, the direction and the tendency of research in communication studies, and to address challenges and weaknesses.

## Theoretical Framework

### 1. The conceptualization of "big data"

Digitalization has brought radical changes in the landscape of media environment and also offered interdisciplinary scientists new methods in the process of data collection (Burgess *et al.*, 2013). The change took place especially because of the availability of a large number of digital data

tracks left by humans as a result of human interaction with digital-based technology. In the context of computer science, the data was known as "big data".

Basically, "the big data" could be defined as a large amount of data that could not be stored, managed and processed adequately by using a standard computer only (Kaisler, Armour, Espinosa, & Money, 2013). A more comprehensive definition proposed by Zou (2015) suggesting that "big data" referred to a very large (structured and unstructured) data set resulting from human interaction with digital-based technology in communication, movement and behavior. According to Duan & Xiong (2015), there were three main issues in the discussion of "big data", including volume, speed and variation. The volume usually related to a substantial measure of data provided by certain sources. The speed related to real-time applications and data processing speed. Meanwhile, variation referred to the diversity of data formats or different unstructured data such as text, graphics, series data, and other related things.

In addition to the above-mentioned issues, a general map of where the digital data could be obtained must also be addressed appropriately. In general, "the big data" sources could be found in various places where human interaction with digital technology occurred. Tsou (2015) identified several examples of "the big data" sources in everyday life of human activities, which were conversations that took place on social media platforms, electronic medical records in hospitals, health centers or insurance companies, records of business transactions such as credit card records and online shopping transactions, movements and traffic data using the GPS system, scientific research data such as earthquake records, weather records and census records. Similarly, Edwards, Housley, Williams, Sloan, & Williams (2013) formulated several human activities that had the potential to generate digital data traces, including retail transactions, telephone communications, financial expenditures and insurance claims, data of population censuses, general household surveys, police record crimes, victims of crime surveys and labor market surveys.

### 2. "The big data" as a source of data in communication research

Although digital data trace could be found in various forms of human interaction with technology, not all types of digital sources and data could be used in social disciplines such as

communication studies. Considering the type of the digital data that was usually in the form of words, sentences, conversations, numbers and emoticons, there were at least two different fields of study in which communication study could be conducted, which were social media-based interactions and web-based interactions.

In the context of social media-based interactions, social study of a variety of popular social media (social networking sites, micro blogs and blogging) could be conducted (Edwards *et. al.*, 2013). For social networking sites, Facebook, Instagram and Youtube were some common examples of favorite social networking sites that could serve as digital data sources of communication study and Twitter was a potential data source for micro-blogging platforms. In case of blogs, communication scholars could use various types of public blogs available on internet. In addition to the categories mentioned above, communication scientists could also use web-based interactions as a source of large digital data for learning purpose. In this category, general news sites (online news portals) and online review sites could be used as main source of digital data traces for the purpose of communication studies.

### Material and Methodology

Unlike various dominant methods in social study that generally used questionnaires or various qualitative techniques (interviews, focus group discussions, etc.) as the tools in data collection process, the study related to “big data” in communication science required a completely different tool. One of the techniques commonly used in “the big data” study was text mining. According to Feldman & Sanger (2007), text mining might be defined as “a knowledge-intensive process in which a user interacted with a document collected over time by using a series of analysis tools”. Also, the text mining was also considered as a method used to uncover knowledge in a computerized process of filtering information from unstructured text documents involving a combination of several techniques, including data mining, machine learning, natural language processing (NLP), information retrieval and knowledge management (de Fortuny, Smedt, Martens, & Daelemans, 2012). In the context of social sciences like communication, the text mining approach could be useful when a researcher wanted to find valuable information from a conversation or a text available in various platforms of social media.

In order to be able to perform the data collection process in a “big data” study, researchers usually needed adequate tools capable of capturing, processing and analyzing the data from a particular computer server. The need for the tools could be fulfilled by using a Data analytics software.

Basically, data analysis software was software that could be used for the process of withdrawal (collection), processing and analysis of digital data contained in various sources of “big data”. Just like a search engine (i.e Google), the software was able to index various data in online environment. In addition to the feature, it was also able to pull/to retrieve data and then to process the data and subsequently to display the results of data processing.

The way how the software worked as described above required researchers to customize the definition of retrieving data in the source of “the big data”. The customization of search engines was made by using several related key words related (included) or not related (excluded) to a particular research theme. The more accurate the key words (included and excluded) related to the specific research theme defined by the researcher, the higher the precision of the results obtained by the software used.

For example, a study of “Tweets containing SARA (*Suku, Agama, Ras, and Antar-golongan*) or ethnicity, religion, race, and inter-group relation issues during the Jakarta governorship election campaign 2017” would retrieve the data from Twitter platform by specifying several key words related to themes such as “the names of all candidates participating in the election”, “campaign”, “indigenous”, “non-indigenous”, “Chinese”, “Islam”, “Christians”, “Muslims”, “non-Muslims”, “infidel” etc. In addition to the use of the key words, the periodization of data retrieval was also an important factor that should be carefully addressed by researchers. In the case of the above-mentioned example, the period of data retrieval was set up during the 2017 Jakarta Election campaign.

In addition to the text mining technique described above, the study of communication science in the context of “the big data” could also be conducted using sentiment analysis. A sentiment analysis techniques could be useful if a study had to reveal the tendencies and the attitudes (the emotions) in a conversation or a post in the online environment (including social media) (W. Duan, Cao, Yu, & Levy, 2013).

Sentiment analysis also referred to specific computing applications based on the emotional polarity of communicators (Li & Wu, 2010). Also, they suggested that the main purpose of the analysis was to determine the emotions contained in a text posted or written by someone related to a particular topic in online environment. Liu (2010) stated that sentiment analysis was a computerized calculation that tested the opinions, the sentiments and the emotions expressed in a text expressed in the online world.

As the social media and internet users continued to grow in Indonesia, the availability of “the big data” was also increasingly significant. Such situation provided communication scholars in Indonesia with an opportunity to elaborate the new field of study. Consequently, the text mining techniques and the sentiment analysis were very important tools for communication scholars in analyzing various data available online, especially the data related to the theme of communication study.

## **Results and Discussion**

### **1. Political Communication**

In the context of political communication, various actual research themes might be raised in response to the abundance of human-generated digital data. Some examples of such issues were as follows: predicting the winner of general elections (local, national, legislative and executive) through the level of popularity and sentiment in the realm of social media and online news portals, making comparisons of official social media accounts of political actors and political institutions (political parties, presidents, house of representatives, etc.), reviewing and comparing the thoughts and statements of political actors as expressed in the online environment, assessing political issues that were popular, interesting and became the main issues of public in general.

### **2. Marketing Communication**

The study theme in the domain of marketing communication was given a good attention because of the availability of data produced by human interaction with digital technology. In general, the major themes of the study that could be conducted in the context included measuring consumer attitudes and sentiments towards particular products, detecting the weaknesses of a product by elaborating consumers' complain of the product, measuring and

comparing the popularity of several products and services in market, comparing and assessing the interactivity level of social media accounts for certain products, assessing the popularity and level of public trust in freelancer services and testing the link between the level of popularity in cyberspace (social media) and public acceptance and sales levels.

### **3. Other (Intercultural, Organizational and Interpersonal) Fields in Communication Study**

In addition to the two fields discussed above, other fields of communication could also give new directions in various studies. Some examples included comparing the differences in the use of cross-cultural adjectives (emotions) on various different accounts (intercultural communication), measuring the use of social media as an official communication medium internally and externally (organizational communication), assessing the depth of interpersonal communication that occurred in social media platforms and testing the link between online closeness and in real world closeness (intercultural communication).

### **4. Content analysis**

In the context of study technique, one of the study techniques (methods) in the discipline of communication science that posed a new challenge and development was content analysis technique (Lewis *et. al.*, 2013). The technique was widely used in communication study. Classical content analysis research technique was generally applied to describe the contents, the message characteristics and the content development (trends) in various written texts such as newspaper, advertisements, important documents like rules and regulations (Eriyanto, 2011).

The increasingly intense human interaction in the online environment also stimulated the availability of digital data in the forms of written text. Consequently, the condition provided communication scholars with a great opportunity to use various types of data as materials of study in applying content analysis method. In addition to quantitative content analysis, qualitative techniques could also be used in various directions of study.

### **Challenges and Limitations**

The digital transformation in one hand provided an opportunity to use of the latest study models in communication science, while in other hand the study using the digital data also had

vulnerability in its implementation. There were several important issues concerning with the methodology that should be addressed in conducting digital data-based study.

### 1. Validity issue

Validity was one of the main issues that arose when data/text mining or sentiment analysis was made in communication study (Boyd & Crawford, 2012). The validity issue could arise during the data/text mining and hence it was necessary to test of the validity of data collected from the sources in internet. Various softwares used to collect and to analyze the data still had considerable errors. Most of software developers that developed the tools claimed that the accuracy level of the application was only about 70%. The percentage indicated that there was still a large amount of data that might not yet be retrieved and further analyzed. It might certainly be serious problem when it came to the generalization of results. On the contrary, a "conventional" survey could generally tolerate errors in the sampling process up to 5%. Consequently, the condition could certainly affect the level of precision of the study.

### 2. Population and sample

The issue of sample and population surfaced in digital data-based communication study because the exact number of populations of posts and conversations related to certain issues in the internet domain could in general be very large and hence it was difficult to predict the exact number. Consequently, it was difficult to ascertain the exact number of samples drawn using random sampling technique (Mahrt & Scharkow, 2013). Meanwhile, justifiable generalizations of the results of the study based on the rules of statistics and social study became dubious.

### 3. Ethical issue

Another factor that could cause serious issue in conducting a study of digital data in cyberspace was ethical factor. The issue especially became prominent because of the use of public and semi-public data. There was not any standard procedure and any ethical standard that served as the basis of the above types of data usage in the context of scientific study (Mahrt & Scharkow, 2013).

## Conclusion

The wide open opportunities resulting from human interaction with the internet or the digital world have resulted in the availability of large amount of digital data that could be processed in a communication study. Consequently, it was possible to start and develop new ways of conducting a study in the field of communication. In the context of Indonesia as a country with a relatively high level of internet usage, the opportunity must be taken by communication scholars.

Also, the communication scholars in Indonesia must pay a good attention to the issue and began to develop new methods to face the challenges as a result of the study/the use of digital data. Such effort should be made by strictly following research method procedures and adequate research design. Additionally, the study must also pay a good attention to various issues previously described to obtain useful results using the right process (population and sample, validity and ethics).

The limitations of the study resulting from the limitations of the tools used in retrieving data should also be the main concern of all communication scholars and hence additional techniques should be developed to optimize the results of a digital media study.

## References

- Abbate, J. (2017). What and where is the Internet? (Re)defining Internet histories. *Internet Histories*, 1(1-2), 8-14. <https://doi.org/10.1080/24701475.2017.1305836>
- Boyd, D., & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, 15(5), 662-679. <https://doi.org/10.1080/1369118X.2012.678878>
- Burgess, J., & Bruns, A. (2012). (Not) the Twitter election: The dynamics of the # ausvotes conversation in relation to the Australian media ecology. *Journalism Practice*, 6(3), 1-40. <https://doi.org/10.1080/17512786.2012.663610>
- Burgess, J., Bruns, A., & Hjorth, L. (2013). *Emerging Methods for Digital Media*

- Research: An Introduction. *Journal of Broadcasting & Electronic Media*, 57(1), 1–3. <https://doi.org/10.1080/08838151.2012.761706>
- de Fortuny, E. J., Smedt, T. De, Martens, D., & Daelemans, W. (2012). Media coverage in times of political crisis: Faculty of Applied Economics, 32(0).
- Duan, L., & Xiong, Y. (2015). Big data analytics and business analytics. *Journal of Management Analytics*, 2(1), 1–21. <https://doi.org/10.1080/23270012.2015.1020891>
- Duan, W., Cao, Q., Yu, Y., & Levy, S. (2013). Mining Online User-Generated Content: Using Sentiment Analysis Technique to Study Hotel Service Quality. *2013 46th Hawaii International Conference on System Sciences*, 3119–3128. <https://doi.org/10.1109/HICSS.2013.400>
- Edwards, A., Housley, W., Williams, M., Sloan, L., & Williams, M. (2013). Digital social research, social media and the sociological imagination: surrogacy, augmentation and re-orientation. *International Journal of Social Research Methodology*, 16(3), 245–260. <https://doi.org/10.1080/13645579.2013.774185>
- Eynon, R. (2013). The rise of Big Data: what does it mean for education, technology, and media research? *Learning, Media and Technology*, 38(3), 237–240. <https://doi.org/10.1080/17439884.2013.771783>
- Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analysing unstructured data* (Vol. 1). <https://doi.org/10.1017/CBO9781107415324.004>
- Housley, W., Williams, M., Williams, M., & Edwards, A. (2013). Special issue: Introduction. *International Journal of Social Research Methodology*, 16(3), 173–175. <https://doi.org/10.1080/13645579.2013.774164>
- Hutchinson, J. (2016). An introduction to digital media research methods: how to research and the implications of new media data. *Communication Research and Practice*, 2(1), 1–6. <https://doi.org/10.1080/22041451.2016.1155307>
- Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). Big Data: Issues and Challenges Moving Forward. *2013 46th Hawaii International Conference on System Sciences*, 995–1004. <https://doi.org/10.1109/HICSS.2013.645>
- Kaul, V. (2012). The Digital Communications Revolution. *Online Journal of Communication and Media Technologies*, 2(3), 113–130. Retrieved from <http://www.ojcm.net/articles/23/237.pdf>
- Lewis, S. C., Zamith, R., & Hermida, A. (2013). Content Analysis in an Era of Big Data: A Hybrid Approach to Computational and Manual Methods. *Journal of Broadcasting & Electronic Media*, 57(1), 34–52. <https://doi.org/10.1080/08838151.2012.761702>
- Li, N., & Wu, D. D. (2010). Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision Support Systems*, 48(2), 354–368. <https://doi.org/10.1016/j.dss.2009.09.003>
- Liu, B. (2010). Sentiment Analysis and Subjectivity. *Handbook of Natural Language Processing*, (1), 1–38. <https://doi.org/10.1145/1772690.1772756>
- Mahrt, M., & Scharrow, M. (2013). The Value of Big Data in Digital Media Research. *Journal of Broadcasting & Electronic Media*, 57(1), 20–33. <https://doi.org/10.1080/08838151.2012.761700>
- Matusitz, J. (2007). The Implications of the Internet for Human Communication. *Journal of Information Technology Impact*, 7(1), 21–34. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.99.5597&rep=rep1&type=pdf>
- Papacharissi, Z., & De Fatima Oliveira, M. (2012). Affective News and Networked Publics: The Rhythms of News Storytelling on #Egypt. *Journal of Communication*, 62(2), 266–282. <https://doi.org/10.1111/j.1460-2466.2012.01630.x>
- Power, D. J. (2014). Using “Big Data” for analytics

and decision support. *Journal of Decision Systems*, 23(2), 222–228. <https://doi.org/10.1080/12460125.2014.888848>

Tsou, M.-H. (2011). Mapping cyberspace: Tracking the spread of ideas on the internet. In *25th International Cartographic Conference*. Paris. Retrieved from [http://icaci.org/files/documents/ICC\\_proceedings/ICC2011/Oral\\_Presentations\\_PDF/D3-Internet, web services and web mapping/CO-354.pdf](http://icaci.org/files/documents/ICC_proceedings/ICC2011/Oral_Presentations_PDF/D3-Internet,_web_services_and_web_mapping/CO-354.pdf)

Tsou, M.-H. (2015). Research challenges and

opportunities in mapping social media and Big Data. *Cartography and Geographic Information Science*, 42(sup1), 70–74. <https://doi.org/10.1080/15230406.2015.1059251>

Zhao, J. L., Fan, S., & Hu, D. (2014). Business challenges and research directions of management analytics in the big data era. *Journal of Management Analytics*, 1(3), 169–174. <https://doi.org/10.1080/23270012.2014.968643>