The Reading Behavior and Industry, Technology, and Communication of Data Journalism: A Literature Review

Mazaya Rizy Safira
Irwansyah

Abstract
This literature review wants to explain how the industry, technology, and communication of reading, especially on news and newspapers. The news reading behavior in society has also changed, from print newspapers to the latest reading through digital media, online and mobile devices. Media organizations and journalists who are already engaged in the news world must make several strategies to survive in the midst of competition and the presence of technology, especially in the era of industrial revolution 4.0. The results of this study provide an insight into the opportunities that can be built in the context of Indonesia, particularly media organizations and journalists in order to survive by making data-based journalistic work.

Keywords: data journalism, reading, news, reading behavior, data-based journalism

Introduction
Reading is still an efficient way to communicate words (Liu, 2005). Based from a study conducted by Central Connecticut State University in the US, Indonesia is in the 60th position, in terms of reading interest (Jakarta Post, 2016). Kurniasih (2016) explained that there are five elements to gain the reading interest, especially in digital era, such as the essence of reading, evolution of the technology, characteristic of the society, reading behavior, and how reading is seen as social activity. The flow of information and the presence of digital content, contributes in changing the human reading behavior from paper printed content then switches to digital format (Liu, 2005; Tveit & Mangen, 2014). From the Global Research Institution (GFK) and Indonesian Digital Association (IDA) survey in 2015, people are 96% tend to read online news (Ali, 2016; Kurniasih, 2016).
The behavior of reading news online is a complex thing with existing technology, compared to traditional news versions, now more flexible and rich environments (Yadamsuren & Erdelez, 2011). Regarding the time spent reading electronic documents, the behavior of screen-based reading or reading on the screen develops. This behavior is characterized by more time required for browsing and scanning, keyword searching, one-time reading, non-linear reading, and more selective reading (Liu, 2012).

The newspaper industry must adapt to the presence of digital by innovating for news products and industries to survive. Therefore, business model must be adjusted, with the industry conditions, use of technology, and communication, one way that can be done with data journalism. Data journalism is seen as a method for investigating and telling something with a large scale and high interest from the reader (Felle, 2016). Simon Rogers (2011) who founded Guardian Data Blog in 2009 (Hammond, 2017b) states that newspapers already do data journalism, but the difference is if the first data obtained with very expensive, now the data is scattered with a format that can be accessed by the computer. Rogers’ emphasis is technical, especially with existing digital technology, which shows that very large and open access to data is helpful. With the development of industry using technology in the era of industrial revolution 4.0, Big Data, cloud, mobile device, also have an important role in communication conducted by journalists in delivering news and data that provide new insights interesting and widespread to the public.

Methodology

This literature review is using the systematic literature review (SLR) approach, that aim to collect and analyze multiple literature and research studies (Ginieis, Sánchez-Rebull, & Campa-Planas, 2012). The various concepts which are used will be explained and illustrated how the linkages between reading behavior in the digital and mobile era and media workers, especially newspaper and news journalists, using data as sources. The main keywords are “data journalism” and “reading”. From the “data journalism” then we broke down the terms in specific “journalism” or “online journalism”. The “reading” also expanded to “reading behavior” and “reading activities”, which include the intensive, extensive, and also the pattern of reading in digital platform. The publication period about reading, included reading skills and behavior, period was since 2005 and the data journalism with the technology that are used to support the improvement of data journalism, in adapting the reading behavior nowadays.

Literature Review

Reading

Reading is an activity of seeing on written or printed writing in order to understand content or content, usually done in silence. On another understanding refer to reading the writing with a voice (Li, 2010). Reading is still an efficient way to communicate words...
Reading is not just a single activity, but an activity that requires a variety of behaviors, which can be done by voice or silence (Liu, 2005). More complex life of the community will provide a growing demand for reading (Liu, 2005).

Reading skills are the abilities and skills that a person has in reading judgments (Perfetti, 2001). In reading skills, there are two skills, namely intensive reading and extensive reading. The reader uses intensive reading techniques to extract specific information on short (Solak & Altay, 2014). Extensive reading is a form of reading instruction, in accordance with the quantity to get a general sense of what is read. This activity is done to increase reading habits to increase knowledge and encourage reading pleasure (Yamashita, 2008; Richards & Schmidt, 2002). Some types of extensive reading are skimming, scanning, rauding, learning, and memorizing. Rauding is a reading activity by looking at each word of the text sequentially and by comprehending the thoughts contained in every sentence (Carver, 1990 in Yamashita, 2008).

The existence of information and digital content, providing better performance in reading printed books and switching to reading digital content (Liu, 2005; Tveit & Mangen, 2014). Reading activities are not merely thoroughly changed and version-based (Tveit & Mangen, 2014), but programming activities, reading promotions, and developing activities due to media usage and reading habits. With regard to the time spent reading electronic documents, then screen-based reading behavior or reading on the screen develops. This behavior is characterized by more time required for browsing and scanning, searching for keywords, one reading, nonlinearity, and more selectively (Z. Liu, 2012).

Associated with the number of keyword searches and browsing or scanning of article volumes, readers skimming the article because it can’t keep reading speed. This reading trend in this way is quite intensive in a web-based environment, especially reading the text on the first page only. Scanning becomes an effective way to sort out the amount of information that spreads quickly. A keyword finding strategy helps to solve the amount of information. Individuals prefer to read overviews, such as graphs and brief tables to gain an understanding of what the text is reading (Liu, 2005).

The results of Liu's research (2005) show that individuals tend to do what is called one-time reading, which is the activity of reading a content or material that only read once. This is supported by research by Levy (1997) in Liu (2005) who says that because of the amount of information, the individual's attention to something becomes very rare. Therefore, in the context of reading with devices and digital content, individuals become more selective in choosing content that they think is more relevant. The result of selective behavior in reading into two things, the reader can fully understand or only understand the surface level alone (Topping, 1997 in Liu, 2005).

Irregular reading arises from the difficulty in reading with fixed attention. The existence of hypertext makes reading behavior by jumping to certain parts and move easily from one source to another. The digital age has brought individual behavior in...
reading at the surface level although the topic varies, according to research conducted by Birkerts (1994) and Stoll (1995) (in Liu, 2005). Hyperlinks are disruptive to the individual's focus on reading, concentrating, and thinking deeply about a topic (Liu, 2005).

**Data Journalism**

Journalism, according to Deuze (Deuze, 2005), is a work with the ideology of journalism and education studies, which is operationalized to analyze how socio-cultural and socio-economic issues emerge to alter the way of thinking and journalistic activities. In the work, journalists as actors of journalistic activities put forward the public service, objective, autonomous, ethical, and have the speed. What is needed in the industry today is innovation (Dekavalla, 2015). The practice of journalism is changing with the presence of the Internet, cloud computing, rapid development (agile), mobile devices, and open source software, leading to the emergence of a new term, data journalism (Howard, 2014).

Simon Rogers (2011) who founded Data Blog Guardian in 2009 (in Hammond, 2017) states that newspapers already do data journalism, only difference if the first data obtained with very expensive, now the data is spread with a format that can be accessed by the computer. Rogers' emphasis is technical, especially with existing digital technology, which shows that very large and open access to data is helpful.

According to Meyer (2002) in Splendore, Salvo, Eberwein, Groenhart, Kus, & Porlezza, (2016) data journalism is rooted in computer-assisted reporting (CAR) and precision journalism elements based on social knowledge. In some academic sources, data-based journalism is explained through various conceptualizations, as previously mentioned using computer-assisted reporting (CAR), journalism data, precision journalism, computational journalism, programming journalism, and algorithmic journalism (Borges-Rey, 2017).

Data journalism is a way to tell stories with numbers and vice versa (Howard, 2014; Gray, Chambers, & Bounegru, 2012). In the midst of the presence of digital technology, data journalism encourages media to adapt and respond to changes in information needed to become richer, interactive and multidimensional. In addition, journalists can also save time by using coding to see the interrelationship between data (Gray, Chambers, & Bounegru, 2012). According to Hammond (Hammond, 2017a) the transparency of journalism data is on sources and data used, as another form of journalist role as watchdog or observer for public information.

Data journalists operate all new forms within the scope of Big Data (Solop & Wonders, 2016b). The preferred issue of journalism data is the use of technology for gathering, processing, analyzing, and presenting large amounts of information. Three things that support journalism data are technological developments and digitalization, market incentives to compete for public attention, and government openness to news
channels (Splendore et al., 2016). Although driven by technology, journalists must maintain professionalism and provide quality news to readers.

Another way according to Gray, Chambers, & Bounegru (2012) news that comes from various sources is selected, by gathering then after it is filtered and visualized. Data that are scattered and viewed only from one side usually become irrelevant, but when it is collected and extracted into new form, it will provide a sharper and deeper view of what is happening and its impact on life. The task of journalists is no longer just in order to report an event, but how the development takes place and meaning, therefore the data becomes a dynamic and widespread opportunity. The obstacle is the training of journalists in processing the data (Gray, Chambers, & Bounegru, 2012).

Discussion

Technology

Big Data

Data (Clifton, 1986) is known as information with a specific structure length and pattern. The unit form of the data is the datum, which is individual, split, and separable, but in its arrangement, the form remains the same as the other (Rosenberg, 2013) in (Frith, 2017). Big Data refers to the collection and analysis of large amounts of data, provide insight or intelligence when analyzed properly (Sriramoju, 2017). Data processing is traditionally incapable of analyzing Big Data, requiring adequate software.

Five things that matter in Big Data are variety, veracity, volume (amount of data), visualization (readable data), variability (change in data), velocity (speed), and value (Sivarajah, Kamal, Irani, & Weerakkody, 2017). In terms of volume, Big Data requires large storage areas and organizations must accommodate by continuing to expand capacity. Speed, data is made quickly and responses must also be made in real time and for a variety of forms, such as text, images, video, and databases. Big Data is a broad term used for datasets that have size, ie dimensions, volume, and speed of generation and complexity, such as diversity and variability that exceed the capabilities of a traditionally used tool to capture, process, curate, and analyze data in a time frame can be tolerated (Beyer & Laney, 2012; Laney, 2001; Guo, Vargo, Pan, Ding, & Ishwar, 2016).

According to Frith (2017), Big Data collected digitally, may not necessarily represent the whole. Hence the question of the Big Data with what is called apophenia, the ability to see patterns that do not exist, simply because the quantity of data that displays connections and spreads in different directions (Boyd and Crawford, 2012 in Frith, 2017). Data mining is the process of discovering patterns that give broad, interesting, and new knowledge, and descriptive, understandable, and predictive models of large-scale data (Zaki & Meira Jr., 2014 in Firth, 2017).

Data journalism is mostly connected with Big Data, specifically related to automation, algorithms, and robots so that its existence becomes important (Hammond, 2017b). If reporting automatically evolves, the available data will be reported
automatically, thus reducing the need for the journalist profession. When data is connected with daily human interaction, it will be replaced with that information. Human existence will be less important and no longer a core role, if human beings are seen as having sides and faults.

According to Fairfield, & Shtein, (2014), some researchers focus on ethics and responsibility on participants. The technique applied by Big Data does not accommodate this with many topics, no informed consent or consent from the individual whose data is used. According to Boyd and Crawford (2012) in Frith (2017), judging that Big Data is not objective and suggests that apply the method of social science as a means to reduce the resulting subjectivity of algorithms and data sets. Ethical journalists should also think about how the data used can lead to the potential for harm and harm to others (Mcbride, 2016).

Cloud Computing

Cloud computing is a model that makes it possible to access networks anywhere, easily, and has on-demand network access to configured computing resources (for example, networks, servers, storage, applications, and services). This service can be quickly set and released with management efforts or service delivery. The five characteristics of cloud computing are extensive network access, source merging, high elasticity, on demand services, and scalable services (Mell & Grance, 2011).

According to (Cohen, Li, Yang, & Yu, 2011) cloud can help people in informing documents and facilitating collaborative tool development. The extraction of information can be shared and reused results from different data processing tasks by different users, which is meant here is the role of journalist, so there is an opportunity to collaborate. The journalism system envisioned by Cohen, Li, Yang & Yu (2011), based on cloud for crowd can store tools and data (both raw and derived, structured and unstructured), to run intelligent tasks and manage crowdsourcing. Merging resources to do tasks and having the opportunity to add human skills.

Cohen, Li, Yang & Yu (2011), provide input in the form of three things that can be done with cloud for crowd. First, the story becomes live, no longer static because the journalist gives a report at real-time. The existing system will provide data from the source by continuously updating and emphasizing when the answer changes. Second, stories and news can appear multiplied, although initially only focus on specific times, places, and contexts, but from collected data can be seen in other agencies, whether applicable or not. Third, the story can be checked quickly, as to whether there is negligence or errors of disclosure of information to maintain the drilling practices in accordance with applicable rules.

Mobile devices reinforce the possibility of journalists working and reporting from their location. The desired data can be collected with mobile and cross-media devices (Westlund, 2013). Internet connections and skills using search engines, easy access to
applications, give journalists the power to provide news. Every organization has a difference in using mobile devices and related to news making practices. Nevertheless, this opportunity is open to a variety of practices, one of which is the location function to organize assignments to journalists and news reporting guides (Westlund, 2013).

Data used in computer-assisted reporting prioritizes data collection and analysis to support a news or investigation, while journalism involves data publishing, reuse, and usability (Howard, 2014). App-based news is an important thing as a form of storytelling because it is accessible to all search engines, devices, and operating systems. App news-based refers to online apps or interactive features with a simple form (Howard, 2014).

**Industry**

Using data journalism as a task, it helps build some new jobs and industries for the environment. According to Gray, Chambers, & Bounegru (2012) the team needed to process the data until the presentation starts from the web developer or website developer along with the web designer. The task is to create a page or application friendly to the reader. If it refers to the pattern of community reading in the digital and mobile-based era, it must be adapted to the circumstances, so readers are comfortable with the display on their screen (Shimray, Keerti, & Ramaiah, 2015). After that, journalists who become leaders, tasked to determine what topics and how the interrelationship between data to be discussed, assisted by members of journalists who are in the team. To help journalists to gather important data, the researcher plays a role in extracting data, entering the data into Excel format, and clearing the available data, especially on Big Data forms.

Data Journalism also brings industry to consultants, who can be asked for opinions on what interesting things that can be made news and reports to the public. In addition, academic consultants may emerge, with skills in data mining, graphical visualization of data results, and high-level research capabilities (Gray, Chambers, & Bounegru, 2012). In addition to organizational and journalist levels, training on journalism data can be applied at the school level, such as high schools and universities. The amount of data that spreads it has a problem in the field of research and development of very few (Mcbride, 2016).

**Communication**

Data journalism can communicate many things and provide interesting insights from scattered data. For example in analyzing political issues in the presidential (Solop & Wonders, 2016b, 2016a). In Argentina, there is ‘Electoral Hack’ as an analysis and visualization of the results of political elections in October 2011. The system features election results and statistics on socio-demographics within the country. This platform gives some impact to the media, such as television, radio, print, and online. The map image of the project is used as a source of data by these media. Interactive maps created
by Argentine journalists appear in several types of print newspapers (Gray, Chambers, & Bounegru, 2012).

Given the data journalism and its application in Indonesia is very possible to do. Media organizations, especially newspapers can build industries. Journalists work with parties such as consultants, web designers, web sites and apps developers, research institutes to extract data to present news. The use of currently available technology can help to communicate data from both social, environmental, and natural resources. To gain information, readers gain a holistic understanding of a phenomenon by not having to open multiple links to other pages.

Topic which can be communicated and applied, is presenting the names of the political election candidates for regional in a map of Indonesia. If we click on an area, for example West Java, then the region will appear, with each name of the candidate, vision and mission, and whoever the supporter party. In terms of natural resources, mapping can be presented from each region, what percent of available resources, the scarcity, and what history and interesting stories about the resource. Other interesting points aimed at the health sector, such as public and private hospital service data available in Indonesia, how much the estimated cost needed for treatment and special treatment.

Case Study about Data Journalism in Indonesia

Data journalism has been applied by Tirto.id, giving the information with infographic (tirto.id, 2018) but should be maintained and improved in some aspect. For example in Electoral Hack as the political analysis from Argentina elections (Gray, Chambers, & Bounegru, 2012), this can be presented in Indonesian context, such as the relationship between the socio-demographic and result in legislative election 2018 and the 2019 presidential. The journalist and researcher could gather the data from each region which held the election. The source can be various, from the survey, the census, or qualitative data. After the data gathering, the analyst should analyze the relationship between the result and socio-demographic in each region. To maintain all the data, what we can learn from the Electoral Hack is developed as an open platform (Gray, Chambers, & Bounegru, 2012) so it can be updated over time. The technology which used by Electoral Hack are Google Fusion Tables, Google Maps, and vector graphics libraries (Gray, Chambers, & Bounegru, 2012). These technologies can be provided to give the display of geographic mapping and electoral demographics interactively. Another case could be made, in specific domain or culture in Indonesia. For example the deforestation in Indonesia, the number of flora and fauna’s scarcity in each province over time, and the data on the demographics and performance of all the public schools or private schools in specific city, such as Jakarta, Bandung, or Surabaya.
Conclusion

Community reading practices are unchanged, but with the presence of digital and mobile devices, the purpose and media have to conform in physical conditions, access, devices, and content to read. This has an effect on the news and news industry to adapt, both in the same technology, as well as the content of the presentation. This behavior is characterized by more time required for browsing and scanning, searching for keywords, one reading, non-linear reading, and more selectively (Liu, 2012). The visualization view of the data has to be adjusted with reading behavior in digital and mobile devices. Not just to please the reader with the interactive and attractive visual, but also develop the interest to read the articles and get the information.

The information data scattered on the various channels collected and filtered by the journalist can be presented following the pattern. The reader can choose the topic they want to know. Interactive visualizations can draw readers from scanning behavior then read and continue clicks to get more information. Non-linear reading patterns are assisted with data journalism that provides information, although the reading is not uniform because the reader does not have to follow words per long or paragraphs per paragraph. If the reader wants depth and structure, the visual presentation can be selected and then read a more complete narration to explain the data presented.

One of the things that can be done is data journalism, which helped build the industry and an opportunity for some parties to engage in the presentation, analysis, and visualization of the content. The technology used in the form of Big Data, cloud, and mobile devices to simplify the work and build interactivity and collaboration. Presentation of news topics can be very broad by linking available data, from natural, geographical, political, economic, to socio-cultural conditions. In addition to the advantages, it also has limitations and big questions about ethical issues in data journalism to be big questions, such as issues in Big Data and whether there is consent from individuals whose data is used.

References

Books

Journal Articles


**Online**