DATA-DRIVEN JOURNALISM BASED ON BIG DATA ANALYTICS: A MODEL DEVELOPMENT FROM INDONESIA'S EXPERIENCE

Z. Hidayat

Communication Department, Strategic Marketing Communication – BINUS Graduate Program Bina Nusantara University, Jakarta, Indonesia

Debra Hidayat

Communication Department, Strategic Marketing Communication – BINUS Graduate Program Bina Nusantara University, Jakarta, Indonesia

ABSTRACT

This study aims to analyze the understanding and practice of data-driven journalism in Indonesia using big data analytics in media institutions such as data infrastructure availability, data structuring, data deepening, enrichment, presenting in the media, resource development, and management, and develop the model of data-driven journalism. The qualitative approach is used by in-depth interviews with 45 participants from media leaders, journalists, vloggers, and audiences. Besides, contents analyses are conducted on 48 samples of various media series along one year. The results showed that big data analytics was conducted for data-driven journalism step by step at national and local media in Indonesia. However, the knowledge and understanding of big data analytics and data-driven journalism in the newsroom and research supporters are still not adequate for both policy and technology. This study suggests the model development of the data-driven journalism process from Indonesia's experience and has managerial implications on empowering data-driven journalism in Indonesia.

Keywords: Data-driven Journalism, Big Data Analytics, Indonesian Media Newsroom.

INTRODUCTION

Changes in the technological environment continue without interruption for each In generation. viewpoint generation, technology is called new media, but for generations after that, there is the emergence of newer and newer media. Therefore, changes in the technological environment must be part of an increasingly better and more beneficial social change. Technology also changes the groups, communities, way individuals, organizations, and communities in large aggregates interact, communicate, and transact. Information and communication technology specifically has changed the way business is done, how to produce, and to distribute its results throughout the world.

Journalism is one of the essential tasks of modern society to maintain each other so that continuity is preserved in modern democratic human relations harmoniously to protect planet earth. The way of working, producing, distributing, and presenting in the world of journalism is also changing. Developments also occur in journalists, media organizations, work tools, data sources and resources, data collection, search and crafting, and presentation platforms. Likewise, innovations occur part of the public, and the audience concept becomes participants due to massive collaboration in all the above processes among all stakeholders. This transformation, from the perspective of Bhatt, Goyal & Yadav (2018), especially concerning social media and the society, is an "important aspect which is authentic and helpful for the development and growth of the society."

Inevitably there is intense competition between media institutions themselves and media institutions with grass-roots media. Now everyone is dealing with other individuals and media organizations fighting over viewers. The landscape of journalism is evolving, and each participant must be able to adjust to that terrible change. On the other hand, technology has been helping and shaping the way of producing content. Carlson (2015) said that most news content is automated, and this determines how this situation creates new challenges for journalists. This development became a "technological drama" of the potential of news technology related to the future problems of the journalistic workforce, rigid conformity between the form of news composition, and the normative foundation of journalistic authority.

The work of a journalist - like all other professions in all fields of life - is assisted by information and communication technology devices. Computer-mediated communication is a part of life that is inseparable from journalism. Previous studies have given indepth attention to the availability of information resources in the cloud data. Information, the contents of earth and planets, is in the cloud that can be accessed by anyone, including journalists. According to Cheruiyot, Baack & Ferrer-Conill (2019), data journalism is now beyond legacy media, which needs a variety of actors within and outside the old media organizations. However, this previous research still discussed the new challenges for journalism, while our paper is more focused on practical models in the field.

Meanwhile, big data appears and grows as part of human life, and any field is always associated with this concept because all human activities on planet Earth are recorded and documented in it. Thus, big data is a and technological social. cultural, phenomenon - which, according to Lewis &Westlund (2015), is a complex amalgamation of digital data abundance, emerging analytic techniques, mythology about data-driven insights, and growing criticism of the overall consequences of big-data practices for democracy and society. However, Lewis and Westlund's view is still very philosophically fundamental about big data and journalism.

One of the most relevant issues regarding the growth and adoption of big data in journalism is about policy and the way models and streams of data journalism work practically. The model and flow of data journalism become essential at the information gathering, screening, production, and news presentation stages. Data journalism practices, especially the use of large data sets, could be used in many newsrooms. Uskali&Kuutti (2015) suggested three different models are needed for regulating data journalism work practices such as the traditional data desk model, the flexible data projects model, and the entrepreneur model (sub-contractor model). This thinking needs to be developed, and this study discusses models and streams of data journalism based on the principles of big data analytics.

Previous studies show a diversity of perspectives to discuss journalism data or the relationship between big data and journalism. Parasie&Dagiral (2013) who researched the role of "computer-assisted-reporters" and "programmer-journalists" in Chicago, USA, computer-mediated of called the era journalism as the revival era of "programmerjournalist" because journalists were armed with information technology skills and artifacts that influenced the way traditional journalism was expected to contribute to the public welfare. This research is essential as a reference because the direction was to find a "programmer-journalist" model, but the focus was more on efforts to strengthen democracy and transparency.

In Canada, Hermida& Young (2017) tried to open unicorn data as the basis of journalism data research by mentioning it as a hierarchy of hybridity in data and computational journalism. This unicorn basis has roles in influenced norms, practices, and news organizations. Then the hierarchy of hybridity is essential to see the concept of developing journalism data as a new way of working.

Data journalism is a form of investigative journalism that continues to grow. As in many other countries in the world, on the European continent, such as in Sweden, Appelgren& Nygren (2014) explained that data journalism began to be adopted into "old" organizations because journalism was something newly introduced as a genre of journalism. Previously, this form of data journalism was known as computer-based reporting and databased journalism, as well as journalism of precision, computing, or database. However, Appelgren& Nygren only introduced new methods and genres of journalism into "old" organizations. As new methods and genres of journalism into "old" organizations may be similar in various countries, but our paper wants to formulate a model of Indonesia's experience.

Problems encountered are in the form and content. In the UK, Knight (2015) found data journalism practiced for its visual appeal and the quality of its investigations, and the overall impact, especially in the tabloid format, was decorative and informative. In the Guardian's big data journalism, content as evidence from creative reportage work in the UK media, according to Tandoc Jr. & Oh (2017), was related to aspects of norms, values, and routines. As the most popular media in the UK, The Guardian is a pioneer in contemporary big data journalism. His findings showed that big data journalism exhibited new trends in how sources were used, but still, in general, adhered to traditional news values and formats such as objectivity and visual use.

Further, in subsequent studies, Borges-Rey (2017) highlighted media at the local level in the UK, such as Scotland, Wales, and Northern Ireland, and discovered different characters based on three conceptual lenses such as materiality, performativity, and reflexivity. This perspective was considered in developing the model of this paper. However, research by De Maeyer et al. (2015) in Belgium, was still discussing the fundamental concept of the growth of data journalism, specifically in the French-language media. According to De Meyer, data journalism could be understood as a socio-discursive practice, not only in the production of artifacts alone but also in the eloquent efforts of all the actors involved in the newsroom.

As a comparison, other previous papers, and the development of similar research in Asia, especially in China, will provide a different perspective because the system and ideology of the country are different from the general Western democratic system. However, in China, the diffusion of data journalism as journalistic innovation also develops. The issue is more complicated than in Western countries if data journalism is applied in socialist-communist countries. In China, the media system has a complexity problem where the inheritance of Chinese quantitative reporting traditions is inadequate, and practitioners have difficulty with the availability of data. Nevertheless, researchers such as Zhang & Feng (2019) had different beliefs and opinions about data journalism that were apparently locally adapted and historically embedded in China, and what facilitated or inhibited them. However, the strengths and weaknesses of previous papers are used as contributions to develop a model in this study.

All things can be reported through social media, and it is often difficult to distinguish between fake and fact news so that journalism data can be the answer. In the South Asian region, such as India, news that is based on facts and quality needed to deal with the tide of fake news circulating in the community. Narula (2019) stated, "data journalism is the need of the hour to move from fake news to verifiable and reliable news rooted in facts. It would allow reporting an incident based on the facts as well as allow readers to examine the facts." Social media tends to be preferred by netizens over official channels. WhatsApp is considered more effective for sharing news (Kabha et al., 2019). Jamil (2019) also examined the practice of data journalism in Pakistan, whose use is increasing following the development software, algorithms, of programming, and data processing techniques by media organizations. Nielsen et al. (2016) see the challenges and opportunities for news media and journalism is increasingly digital, mobile, and social media environments. However, our study does not explicitly discuss the role of social media but rather the broader context that underpins social media: using big data.

As an analogy at the global level, research conducted by Heravi (2018) was a summary of results from the global data journalism survey of data journalism at the global level, which included the participation of journalists from 43 countries. Heravi explained various aspects of data journalism practices throughout the demographics, world, including skills, education, and the formation of data teams, as well as opportunities and values related to data journalism. In its current growth, the best practice of data journalism is needed, as a mechanism of expertise and skills to develop this environment.

Overall, in the context of big data and journalism, media and communication scholars have begun examining and theorizing about big data in the circumstances of media and public life at large, both regarding the specific implications for journalism, and the consequences of collaboration in presenting information among all stakeholders. All previous papers showed great attention to the development and practice of data journalism in various places around the world. However, they had not yet paid attention to how the conception and practice of journalism in big data analytics as a database of journalism in media institutions, and to understand the habits of journalists working with big data and the policy in building resources for data journalism. This study takes a position in the gap that has not been filled.

Based on the background above, the research aims, first, to analyze the understanding and practice of journalists regarding infrastructure data. Second, to analyze media institutions conduct data structuring. Third, to analyze journalists and media institutions conduct data deepening, enrichment, and presentation in the media. Finally, to analyze media institutions in perceiving the importance of resource management and developing data journalism models.

CONCEPTUAL FRAMEWORK

Big data and Data-driven Journalism

This concept of explication analyzes the relationship between big data and data journalism. Big data is the basis for almost all fields of business, including advocacy and public services, through the power of journalism. However, operational institutions, public institutions such as media corporations consider the management of their organizations of the companies. The term "big data" in media corporation has attention in fields of life and is discussed in broad scope, including in its derivation for the term data journalism. Kalyvas& Overly (2014)summarized what big data is:

Big data is a process to provide insight into decision making. This process uses people and technology to quickly analyse large amounts of data of various types (structured data with traditional tables and unstructured data, such as images, videos, emails, transaction data, and social media interactions) from various sources to produce actionable knowledge flows.

Based on the definition, it reflected how much data is collected every second and continues to

grow to be accommodated in cloud computers. Contemporary journalists use various digital tools and services to collect, manage, and process information for public consumption. Even data journalism is considered as a new genre that brings change. Data journalism is welcomed by many because of the availability of data in digital form and the number of efficient online tools. Everything can help users to analyze, visualize, and publish large amounts of data (Veglis&Maniou, 2018).

Meaning according to Mayerdata, Schönberger&Cukier (2013), is no longer considered static, rigid or staled – the benefits are completed once the collection objectives are reached, such as after the aircraft has landed (or in the case of Google, after a search request has been processed). Instead, data becomes the raw material of business, vital economic input, which is used to create new forms of economic value. Even with the right mindset, data can be cleverly reused, chosen with humility, willingness, and listening instruments. data contains All that information and is suitable for the public to know is the raw material for journalism data.

Large amounts of raw data are stored on internal servers and web servers around the world, but only what is needed has valuable information to be withdrawn and presented to the public. Referral data journalism is the search and processing of data for journalism that relies on computer machines. It is like a journalist assisted by a machine to guide him in producing news, even Carlson (2015) called it a robot reporter. A nickname was given because the automatic style of journalism forced a redefinition of the workforce, the form of composition, and changing authority journalism. The term "automatic of journalism" indicated an algorithmic process that transforms data into narrative news texts with limited human intervention outside of programming initial choices. The consequences of increasing the ability of news texts that are written by machine signify a new possibility for a news content field that far exceeds the ability of human journalist production.

Journalists must realize that in this era of big data, the news reports will be more influential for public trust and transparency. Then, today, big data has become a necessity as the basis of journalism. Mayer-Schönberger&Cukier (2013) also explained that big data caused a revolution that would change the way we live, work, and think about things that were of an people's individual other interest. or Therefore, big data had the potential to influence society and science, how to produce and share experiences, and how to make decisions receiving near-perfect after information.

Debate about determination information technology into journalism, and it makes it a dichotomy between before and after the adoption of technology. However, it is essential to follow the thinking of Witschge et al. (2016), that instead of drawing a dividing line between traditional journalism and the future, mainstream and alternative, digital and nondigital journalism - where this condition shows the complexity and diversity of contexts and journalistic practices - it is better to understand fundamental changes in the field of journalism over several last decades due various factors. to Not merely technological issues as a causal factor, but also cultural, social, or economic changes make journalism find its place.

Another term put forward beside data journalism is digital journalism. It appears earlier, and the scope of meaning is broader than data journalism, but the practice of both in the field is the same. Josephi in Witschge et al. (2013) stated that digital journalism was not born from anywhere, but rather grew from technical possibilities and, in a short period, has changed the face of journalism, especially in North America, Europe, and Australia. Globally, traditional journalism is still in control, and for this reason, digital journalism seen concerning traditional must be journalism rather than in isolation.

Further and more contemporary, researchers like Lewis (2017) explained that journalism in the era of big data is a way of viewing journalism as an interpolation through a conceptual and methodological approach to calculation and quantification in the form of rows of numbers or visual graphics. However, the way to discuss journalism data must be comprehensive, including its practices and philosophies. Thus, it also means paying attention to the socio-cultural dynamics of previous calculations and quantifications. According to Anderson (2016), the breadth of the horizon means recognizing that big data algorithms and computational tools and techniques are not entirely material nor fully human but hybrids way of journalism. This issue, by Lewis and Usher (2013), was discussed as the role of technology in the newsroom, specifically the introduction of open-source culture.

Data-driven Journalism and Its Role in Society

The distribution and adoption of digital handheld devices can be stated to be evenly distributed throughout the world. Likewise, in Indonesia, smartphones and other gadgets are part of everyone's attributes, whether in urban, suburban, or even rural areas because telecommunications infrastructure is generally evenly distributed in a country consisting of thousands of islands in Southeast Asia. The distribution of digital devices to tens of millions of populations in countries with rapidly increasing education levels provides opportunities for the growth and development of big data in the national context of Indonesia.

Every smartphone owner is a productive person to create content and then share it with others, groups, communities, and the public. Currently, institutional media in Indonesia produce content by considering the role of viewers or similar collaborative mechanisms journalism. citizen The public in is increasingly media literate and increasingly smart as both citizen-producers and viewers to handle public issues. Tong & Zuo (2019) said that because of the social constructivism nature of data journalism, serving the public interest and building democracy is a more appropriate principle and objectivity for data journalism.

Journalism is undergoing digitization, changing from analog to digital bits. Besides, it is also experiencing digitalization, which captures how journalism is restructured by digital communication and media infrastructure (Brennen &Kreiss, 2018). The purpose of data journalism is threefold, "enabling the audience to find relevant information personally, uncovering extraordinary and previously unknown stories and helping the audience to understand better

complex problems" (Hermida& Young, 2019). This explication is more practical because, previously, according to Royal & Blasingame (2015), the term "data journalism" began to be used instead of the more traditional "computer-assisted reporting" or "computational journalism"— a term that might be frightening to some journalists.

Royal & Blasingame (2015) described the development of journalism data, which later became a topic of academic research in various developed countries. Royal (2010; 2012) also explicitly examined and published a case study team of The New York Times Interactive News, spending one-week interviewing team members and studying the process of adopting journalism data. Identification of unique skills, workflows, and culture of an organization was identified and integrated with storytelling, combining quantification and quality concern. In another case, Royal (2013) also examined the diffusion analysis of the interaction of Olympic activities in The New York Times from time to time. Other researchers such as Parasie and Dagiral (2013) studied data-based journalism in the Chicago found that practitioner Tribune and programmer-journalists had provided new ways for journalism to deal with social kindness.

As Coddington (2015) stated, professional journalism was historically built from two elements, such as textual and visual. From this primary character, three forms of quantitative journalism emerged, namely computerassisted reporting, data journalism, and computational journalism, all of which often overlap and diverge in the context of values practices. Hammond (2015)and also mentioned the use of data and computers in contemporary journalism, which gave rise to various terms such as data (-driven) journalism, computational journalism, journalism, programmer algorithmic journalism, robot reporting, and automated journalism. However, it is also characterized by a post-objective professional mindset that emphasizes the transparency of editorial processes, user integration, and new qualitative epistemological values from large Transparency is a data sets. quality management strategy by sharing an approach to gathering information with users. The new epistemological value of large data sets is that knowledge is obtained inductively by exploring data, discussing initial findings, and detecting meaningful data patterns with the assistance of algorithms (Rinsdorf& Boers, 2016). However, all the support instruments basically must be aimed at building a society that grows its democracy through transparency.

METHODS

This paper focuses on answering the following research questions: first, what is the understanding and practice of journalists regarding infrastructure data? Second, how do media institutions conduct data structuring? how do journalists and media Then, institutions conduct data deepening, enrichment, and presentation in the media? Furthermore, finally, how do media institutions perceive the importance of resource management regarding data journalism?

A qualitative approach is used to collect data in addition to the literature review. In-depth interviews were conducted with participants from several categories or levels in organizations and audiences such as media institutions, journalists, and ordinary viewers and news or events vloggers. At a media institution category, participants were taken from the chief editor or editor in chief of the media institution or manager or at the level of the reporting coordinator. In the journalist category, participants were taken from various types of media, such as television, online, radio, and newspapers. In the viewer category, ordinary people were taken as readers, news viewers in television, print, or online media. A few viewers were also members of the community who were actively producing and sharing content on YouTube as vloggers.

Based on these categories, this study gathered 45 participants consisting of seven managers or coordinators of national media coverage, five from local media in several provinces, five television journalists, two from radio, seven online, and five newspaper journalists. There are seven non-vloggers viewers and seven vloggers-viewers.

Open-ended questions were arranged in general and specifically addressed to each participant category. Some open-ended question items generally related to the research objectives discussed in this paper, namely, to find out precisely and in-depth about a few groups of indicators such as first, journalists' understanding and practice regarding infrastructure data. Second, practices of media institutions in conducting data structuring, then the practices of journalists and media institutions in conducting data deepening, enrichment, and presenting in their media. Moreover, finally, opinions and policies of media institutions the importance regarding of resource management related to data journalism.

In addition to the in-depth interview technique, there were also observations and contents analyzes on various media that were the work of active groups of workers and viewers. Observations were made at each media office, and content analysis was carried out on news manuscripts in each media whose samples were taken from 12 months in 2019 with a sample of one edition each month, to obtain a sample of four types of media multiplied by 12 editions, which produced 48 editions media content. In every media content, the news is searched using data journalism style such as included quantitative data and visual, and weekly trending topics rank.

In-depth interviews and content analysis pay attention to coding on each indicator (Kalyvas& Overly, 2014; Henry & Venkatraman, 2015), such as first, data infrastructure includes data domain, intelligent business domain, statistical domain, the home page, sentiment page, and weather page. Second, data structuring includes keyvalue stores, column-based stores, graphbased stores, docs-based stores, sentiment analysis, trending topics, and recommender systems, social media-language detection, social media-named entity recognition, data validation-parameters, data validationmethods, and data validation-results. Third, deepening and qualitative enrichment data includes news craft and improvement, indepth interviews and confirmation, and performance optimization – finally, the resource management, and sustainable development policy of data-driven journalism. Data processing of the results of the interview is the primary data, which is the subject of analysis in addition to content analysis. Both are combined to describe adoption in practice and reasoning about data journalism in Indonesia. Data triangulation was carried out on interview transcripts, and summaries related to each context were presented in the results.

RESULTS AND DISCUSSION

Data Infrastructures in Media Institution

As the concept of explication before, big data is concerned with the existence of large, complex and ever-growing data sets that are enriched and derived from a variety of

Media	Data	Business intelligent	Statistic	Home	Sentiment	Weather
	domain	domain	domain	page	page	page
Koran Tempo	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
Kompas	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
Media Indonesia	\checkmark		\checkmark	\checkmark	\checkmark	\checkmark
Republika	\checkmark		\checkmark	\checkmark	\checkmark	\checkmark
Jawa Pos	\checkmark		\checkmark	\checkmark	\checkmark	\checkmark
Bali Post			\checkmark	\checkmark		\checkmark
Sriwijaya Pos			\checkmark	\checkmark		\checkmark
Harian Fajar			\checkmark	\checkmark		\checkmark
TV One	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
Metro TV	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
i-News			\checkmark	\checkmark		\checkmark
Radio Elshinta			\checkmark	\checkmark		\checkmark
Radio Sindo Trijaya	\checkmark		\checkmark	\checkmark		\checkmark
Detik.com	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
Kompas.com	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
Liputan6.com	\checkmark		\checkmark	\checkmark	\checkmark	\checkmark
Kumparan.com	\checkmark			\checkmark	\checkmark	\checkmark

Table 1. Availability of Data Infrastructure at Media Institutions in Indonesia

Note: ✓ : available

independent sources (Wu et al., 2014), such as from environmental and body sensors, cellular devices, administrative claims data, social media, email, laboratory studies, electronic medical records, internet, business transactions, and geospatial devices or sensors (Gerard et al., 2014, Barton, 2016; Chan, 2013). Big-data growth rates are high-speed, and it can be produced by machines or humans (Henry & Venkatraman, 2015). From this understanding, it means that all journalism activities are categorized as the work to build big data from time to time. Journalists' job is to create and store data on a continuity basis.

Most established media institutions have a data infrastructure that is used as an operational base to produce, distribute, and present news or articles on various topics. However, some media institutions are not pioneering to build their infrastructure, and the work takes time in line with the development of their respective media. Some local media in the area have been acquired or established by national media business groups so that the infrastructure data is also centered on the holding company. One of the influential media institutions in Indonesia, Tempo Media Corporation, which publishes Tempo Newsletter or Koran Tempo, has a robust data infrastructure base and has long been known as the Tempo Data Centre (TDC). B. Setvarso, editor in chief of Koran Tempo explained:

In the media institutions where we work, the existence of data is essential. journalists However, often are overwhelmed to search, find, and process them. We have a research team to serve the editorial team. The seconded researchers are tasked with processing unstructured data into mature data and ready to be displayed in news presentations. They are also a capable visualization team that is equipped with the latest applications, including animation.

Tempo Data Center (TDC) is accessed by the public to search for data and information from various fields, so that educated people often come to conduct studies. Several other media institution groups are working hard to build their infrastructure data and thus have not yet opened public access for this purpose. N. B. Nugroho, head of the multimedia department at *Media Indonesia* newspaper explained:

We often discuss big data as a topic of internal discussion, but indeed among the journalists, no one mastered what big data is and how it works. We had invited experts to discuss it, but in the discussion, there was a kind of gap, because the experts were information technology people while we as journalists were very diverse in their backgrounds.

Many practitioners of journalism, especially journalists themselves, are not aware that they have a role in building big data that is concentrated in the media institutions where they work. Journalists feel that they only work by collecting data through interviews, checking the internet to enrich writing, and submitting it to proofreading for later publication. However, some of them also realize the importance of building big data for data journalism.

Implementation of big data systems in media institutions is critical because, in that system, we will find home page views, sentiment pages, or even weather pages as a standard measure. However, it was found that the display of quantitative data often lacks in the narrative that gives deep meaning to graphic visuals. Research by Cushion, Lewis & Callaghan (2017) also found that statistics are often referred to in news coverage, but their role in interpreting situations is often unclear, lacking, and inaccurate. F. Pratama, the managing editor of *Detikcom*, explained the online news homepage he led:

Like our print newspaper edition, the online edition that is sought on the homepage is a graphic display that is attractive and easily understood by the public. This data display is the result of processing from the graphic team whose materials are prepared by the research team or from the journalists themselves. Often, we also cite outside sources to be presented on the front page.

The data infrastructure of media institutions or the public can always rely on open access, including social media and institutions of statistical centers and the like. However, what is most talked about is digital journalism and social media producing large amounts of digital content every day, journalism scholars faced with new challenges to describe and analyze the wealth of information. Günther & (2016) illustrated this practice Ouandt workflow of assuming а rule-based approaches, dictionaries, supervised machine learning, document grouping, and topic models. Arifin Asydhad, editor in chief of Kumparan.com explained the appearance of the online news homepage he led:

Our online news homepage must be attractive, so viewers do not get bored visiting our site all the time. There is always something new and variations on the headlines and visual appearance. The main points of the identity and visual character of our website must still be there to be easily recognized by anyone. Our online news homepage is in the top ten on the number of viewers and their active involvement in the issues that interest them.

Often, sentiment analysis is carried out on data collected from the Internet and various social media platforms. Politicians and governments often use sentiment analysis to understand how people feel about themselves and their policies. Ilham Safutra, head of the *Jawa Pos* newspaper coverage which also has a regional newspaper network in various provinces in Indonesia, explains:

Our media is very concerned about a topic that has an impact on public sentiment. However, so far, we have not been independent in processing sentiment analysis. The analysis needs external parties who usually do regular releases and hold press conferences for many media, and we are there. We try to do a text analysis process to interpret various data sources, such as viewers' opinions, but they are still limited.

Data Structuring Practices in Media Institutions

Data structuring is a big job that must be done by the media as an institution that advocates for the community, government, legislative, and judiciary institutions. Structuring owned by the media must be several steps ahead of other institutions because the media supervision is inherent as a watchdog for the Therefore, state administration policy. understanding, adoption, and practice of keyvalue databases in media institutions is a necessity for storing, retrieving, and managing associative arrays, and having data structures as a dictionary or hash table. Media institutions have dictionaries that contain a collection of objects, or historical editorial

Media	Key-values	Column-based	Graph-based	Docs-based
	stores	stores	stores	stores
Koran Tempo	\checkmark	\checkmark	\checkmark	\checkmark
Kompas	\checkmark	\checkmark	\checkmark	\checkmark
Media Indonesia	\checkmark	\checkmark	\checkmark	\checkmark
Republika	\checkmark	\checkmark	\checkmark	\checkmark
Jawa Pos	\checkmark	\checkmark	\checkmark	\checkmark
Bali Post	\checkmark	\checkmark	\checkmark	\checkmark
Sriwijaya Pos	\checkmark	\checkmark	\checkmark	\checkmark
Harian Fajar	\checkmark	\checkmark	\checkmark	\checkmark
TV One	\checkmark		\checkmark	\checkmark
Metro TV	\checkmark		\checkmark	\checkmark
i-News	\checkmark		\checkmark	\checkmark
Radio Elshinta	\checkmark		\checkmark	\checkmark
Radio Sindo Trijaya	\checkmark		\checkmark	\checkmark
Detik.com	\checkmark	\checkmark	\checkmark	\checkmark
Kompas.com	\checkmark	\checkmark	\checkmark	\checkmark
Liputan6.com	\checkmark	\checkmark	\checkmark	\checkmark
Kumparan.com	\checkmark	\checkmark	\checkmark	\checkmark

Table 2a. Data Model Building Practices at Media Institutions in Indonesia

Note: ✓ : available

records. All documents are stored and retrieved if needed by using a key that uniquely identifies the records and finds them quickly.

Media institutions in Indonesia relatively follow contemporary developments in the adoption of journalism, although it is still limited to large established media groups. However, significant work in data structuring as a key in data journalism is carried out by almost all institutional media, depending on the type. F.H. Santoso, general manager of research and development at *Kompas.id* said:

Our institution has long building data storage, and this is our basis, that in the development process, there is a digitization of the library from the conventional form, so we have gone through it. We have been building sophisticated data centers, crossing the developments from time to time, especially regarding Indonesia and the relevant international world. Now the development is at its optimum so that it has added value and is continuously being updated. The public can access our data center with digital services with only a small fee.

Some established media institutions such as *Tempo Media* and *Kompas Gramedia* in Indonesia are known as media institutions that have advanced big-data structures because they have been built over decades and become models of big data media institutions. Besides, several groups of media institutions in the region also have sufficientbig data because of their excellent inter-island networks, such as the *Jawa Pos Network Group*, the *Media IndonesiaGroup*, and the *Media Nusantara Citra* (MNC) Group. Ample data storage in media institutions illustrates its ability to handle data journalism at high speed, high volume, and high data varieties.

Technically, media institutions have distributed file systems, NoSQL databases, graph databases, and NewSQL databases where data is stored securely, and privacy protected. According to Strohbach et al. (2016), databases that are built must pay attention to several things such as databases in memory, and column databases usually outperform traditional relational database systems. Also, the main technical barriers to widespread retrieval of large data storage solutions are missing standards. Moreover, there is a need to address open research challenges related to graphical scalability and database performance. Problems faced by media the especially skills institutions, and knowledge of individual journalists, are generally still low to utilize the availability of big data creatively and optimally. M. Hafizni, a senior journalist at *MetroTV* station, said:

We are used to processing qualitative data from interviews so that if there is data in the form of tables or graphics, we are often overwhelmed to discuss it in an argumentative language. The news material is too severe and scientific for us. We often leave it to the research team whose part is to discuss it.

The same finding was also stated in research conducted by Young, Hermida & Fulda (2018), which showed that journalists had a lack of understanding and were less able to achieve the standard of excellence. Two main factors, which they should be familiar with (such as the use of free online options on Google Maps) turned out not to be easily adjusted, and the number of practitioners working on data projects was mainly within the framework of traditional journalism. According to him, the most widely used visual elements were graphics, maps, dynamic and videos. Concerning interactivity, all but one of the projects contained interactive elements. The most popular interaction techniques were inspection and screening, which is considered an entry-level technique in the field of information visualization.

As a comparison of other research findings, Loosen et al. (2017), examined, among other aspects, data sources and types, visualization, interactive features, topics, and producers in their research. The potential of data-based journalism was mainly aimed at community advocacy as a form of the social function of journalism. The existence of bar-charts was also commonplace in newspapers, magazines, or online news. Stalph (2018) also found the same thing after analyzing daily news from data-driven, which showed that daily journalism data illustrated two visualizations that tend to be bar charts. Most of these

visualizations were not interactive, while maps turned out to be the most interactive type of visualization. more professional news broadcast platform and a more participatory online platform. Researchers such as Johnson & Dade (2019) refer to it as local broadcast journalism with

Media	Sentiment	Trending	Recommender
	analysis	topics	systems
Koran Tempo	\checkmark	\checkmark	✓
Kompas	\checkmark	\checkmark	\checkmark
Media Indonesia	\checkmark	\checkmark	\checkmark
Republika	\checkmark	\checkmark	\checkmark
Jawa Pos	\checkmark	\checkmark	\checkmark
Bali Post		\checkmark	
Sriwijaya Pos		\checkmark	
Harian Fajar		\checkmark	
TV One	\checkmark	\checkmark	\checkmark
Metro TV	\checkmark	\checkmark	\checkmark
i-News	\checkmark	\checkmark	
Radio Elshinta	\checkmark	\checkmark	
Radio Sindo Trijaya	\checkmark	\checkmark	
Detik.com	\checkmark	\checkmark	\checkmark
Kompas.com	\checkmark	\checkmark	\checkmark
Liputan6.com	\checkmark	\checkmark	\checkmark
Kumparan.com	\checkmark	\checkmark	\checkmark

Table 2b.	Data Model	Building	Practices at	Media I	Institutions in	Indonesia:Tex	xt-meaning
10010 20.	D'atta moaci	Dananis	1 Iucucco ut	inicala i	intonio int	indoncola.ic/	w meaning

Note: ✓ : available

Sentiment analysis, trending topics, and recommender systems are the most vital elements and are widely adopted and discussed in journalism data. Likewise, in the eyes of viewers or online news users, trending topics are timely and are always sought every minute. The public is very enthusiastic to know and share about trending topics that occur on a daily or weekly basis. G. M. Sigit, the chief editor at *MNC Radio Networks-MNC Sindo Trijaya FM*, said:

At the news radio station, we work based on trending topics that occur, per minute, daily, up to a week. Even routinely in the week, we count trending topics that have the highest audience reception. Based on that, we do face-to-face events with customers or listeners such as live or on-air discussions on weekends on the hottest topics of the week.

News and talk show radio pay attention to issues that are developing in society, and as such, journalists must work on the issue. Broadcasters need to be sensitive to what in the minds of the public or the community of listeners. *SindoTrijaya Radio* seeks to develop a

more considerable attention to user-generated content and boundary work. Not only in the radio broadcast, television, and online broadcast industries, print media also routinely include trending topics from social media databases (especially Twitter) whose measurements are obtained from internal or external data. The journalistic algorithmic engine allows media institutions to map quickly.

Dörr (2016) did mapping in the field of algorithmic journalism based on the evolution of natural language generation (NLG). He explained that with software that automatically generates text from structured data, NLG transforms traditional news production into fast computer-based. The public wants to know what is happening around them by turning on the radio, even actively participating on-air. Appelgren (2018) called this combination of human and computer machine sensitivity as hybrid journalism, namely contemporary data journalism that was inherent due to verbal and visual elements. Trending topics and sensitivity analysis were processed to be presented visually.

Modia	Social modia:	Social modia:	Data	Data	Data
Media		Social Ineula.			
	Language	Named entity	validation:	validation:	validation:
	detection	recognition	Parameter	Methods	Results
Koran Tempo	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
Kompas	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
Media Indonesia	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
Republika	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
Jawa Pos	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
Bali Post			\checkmark	\checkmark	\checkmark
Sriwijaya Pos			\checkmark	\checkmark	\checkmark
Harian Fajar			\checkmark	\checkmark	\checkmark
TV One	\checkmark		✓	\checkmark	\checkmark
Metro TV	\checkmark		\checkmark	\checkmark	\checkmark
i-News	\checkmark		\checkmark	\checkmark	\checkmark
Radio Elshinta			\checkmark	\checkmark	\checkmark
Radio Sindo Trijaya			\checkmark	\checkmark	\checkmark
Detik.com	\checkmark	\checkmark	\checkmark	✓	✓
Kompas.com	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
Liputan6.com	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
Kumparan.com	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark

Table 2c. Data Model Building Practices at Media Institutions in Indonesia: Social Media and Data Validation

Note: ✓ : available

The tendency of social media to be the basis for information retrieval is very dominant in the newsroom. Heravi& Harrower (2016) also stated that Twitter formed a source and verification process in the editorial room. Carlson (2018) mentioned that the role of social media, especially Facebook, in writing news tends to be confidential and even haphazard. According to him, Facebook reacts to controversies with personalized writing, organized according to the popularity and involvement of its users. Most journalists do see the benefits of social media (Weaver &Willnat, 2016). The most common short cut for individual journalists was to verify social media content because there was an emerging journalistic verification practice concerning social media (Brandtzaeg et al., 2016). Therefore, the source of social media is vital, and the media must conduct language detection and named entity recognition in the data structuring process. Lewis, Zamith&Hermida (2013) argued that an approach that combined computational and manual methods throughout the content analysis process could produce more useful and exciting results in a case study of news sources on Twitter to illustrate the hybrid approach.

Owned technology is directed to build social engineering, and the two are inseparable. Cheruiyot, Baack & Ferrer-Conill (2019) said that the practice and role of nonprofit public institutions and civil technology, such as journalism, were essential for building society. Research conducted by Baack (2018) said something similar that it needed the terms between data journalism and civic tech that describe progressive datafication in social life. Public aspirations that reflected in social media, the implications of data journalism are, therefore, increasingly strong to pay attention to the public interest.

Furthermore, meaningful work in journalism is data validation, both regarding parameters or indicators set according to journalism ethics and internal organization, including methods and results. Data journalism practices in Indonesia show that data validation is essential so that it must be present in every media. Quantitative statistical data are essential, but the visual artifacts must be balanced with the qualitative side. Related to this, Rinsdorf & Boers (2016) analyzed the occurrence of qualitative changes in newsgathering caused by the availability of open data, big-data sources, and the consequences for quality management. Therefore, it is necessary to align and expand individual skills and inventory of new

resources. Likewise, the management of newsroom quality must be adjusted to provide accurate news in data journalism.

Big data in social media exists and grows in large numbers but must be structured according to needs. Ilham Junaidi, the editor in chief of the *Republika* daily newspaper, said:

We often do not focus on social media trending topics because, in every editorial meeting in our newsroom, we discuss and debate about content planning following Republika's vision and mission. We still monitor the public aspirations or events that are being discussed by many netizens and use them to choose various alternative issues. However, we put forward Republika's mission as a national newspaper that wants to continue to bring a change towards a better direction with a new spirit.

Data Deepening and Enrichment

The data deepening phase in data journalism is essential because it is at this stage that a balance is made between the work of computer machines and the narratives interpreted by humans, namely journalists. After a computer assists the work of a journalist with the results in the form of visual statistical data, illustrated images, and charts, then news craft and quality improvements are the lives of every journalistic work. All media in Indonesia operate with or without help visual statistical illustrations in the written news series. However, every journalist at a time or periodically must combine the visual artifacts and verbal narratives of the news.

In certain conditions, journalism data presentation often does not reveal much for a type of investigative reporting. Therefore, the journalist has the function to dig deeper into what happens behind the visual presentation of quantitative statistical data. Journalists must believe that the quantification system has shortcomings that must be complemented by quality.

In-depth interviews and confirmations in the check and re-check series are needed to avoid publication and broadcast errors. Journalists know who is eligible to be re-interviewed and confirmed for the interim results obtained. The snowball interview method, for example, is significant to be adopted to deepen the news side. Irna Gustiawati, editor in chief of *Liputan6.com* online news explained:

We need to study and then determine the exciting side of the news to be discussed from journalism data published by various institutions, consultants, or researchers. We

Media	News craft and	In-depth interview	Performance
	improvement	and confirmation	optimization
Koran Tempo	\checkmark	\checkmark	\checkmark
Kompas	\checkmark	\checkmark	\checkmark
Media Indonesia	\checkmark	\checkmark	\checkmark
Republika	\checkmark	\checkmark	\checkmark
Jawa Pos	\checkmark	\checkmark	\checkmark
Bali Post	\checkmark	\checkmark	\checkmark
Sriwijaya Pos	\checkmark	\checkmark	\checkmark
Harian Fajar	\checkmark	\checkmark	\checkmark
TV One	\checkmark	\checkmark	\checkmark
Metro TV	\checkmark	\checkmark	\checkmark
i-News	\checkmark	\checkmark	\checkmark
Radio Elshinta	\checkmark	\checkmark	\checkmark
Radio Sindo Trijaya	\checkmark	\checkmark	\checkmark
Detik.com	\checkmark	\checkmark	\checkmark
Kompas.com	\checkmark	\checkmark	\checkmark
Liputan6.com	\checkmark	\checkmark	\checkmark
Kumparan.com	✓	\checkmark	\checkmark

Table 3. Data Deepening and Enrichment at Media Institutions in Indonesia

Note: ✓ : available

consider all published data as a trending topic because we must pay attention to many aspects to create a booming issue for netizens. We consider the depth of data to be processed and published gradually and its continuity. This performance depends on the public and viral response obtained from sharing links through social media. The involvement of netizens is significant to us.

All media samples in this research have a reference to achieve performance optimization. This achievement indicates that the media is growing with qualitative as well as quantitative measures. Journalists work with orders that are internally and externally outlined in the association of print press, television, and radio journalists.

On the one hand, journalism data writing practices are often considered narrative dry or lacking in life. Therefore, the data model makes it possible to juxtapose knowledge with writing narratives that are driven by events. Caswell & Dörr (2018) found several technical and editorial challenges. Two components, such as a structured database and narrative, are needed to provide journalistic knowledge writing platform from simple to the descriptions to real stories. So, the integration of quantitative artifact data and qualitative narratives is a way to achieve performance optimization. Lewis, Zamith&Hermida (2013) also emphasized that analyzing big-data content requires an approach to computational and manual methods because an extensive collection of communication data challenges the traditional, human-driven approach to content analysis.

Model Development of Data-driven Journalism

The following development model of datadriven journalism process is a summary of all the steps taken to prepare data journalism. The experience is not only among media institutions in Indonesia but also throughout the world, and therefore this model is expected to be an essential practical guide for developing data-driven journalism in various places and times. The stages presented begin with the process of big data analytics and continue with the process of data journalism for the sake of reporting, transparency, and democratic growth.

The six stages in this model are infrastructure building, data structuring, data data deepening and enrichment, data visualization, classification, clustering and composing, and finally, presentation and publishing. At each stage, newsroom must perform tasks for both journalist and researcher, or specific as big data research team. This model can still be developed according to the conditions of each media, but overall it has been built based on experience stakeholders in Indonesia to practice and develop data-driven journalism.



Figure 1. Model development ofdata-driven journalism process (Adapted model of Rinsdorf& Boers, 2016)

Resource Policy and Management

Data journalism is a new way that places the ability and accuracy of computers to collect data and information to be processed, enriched, and presented on several media platforms. The length of the process makes it impossible for a journalist to do everything on his/her own. Therefore, we need organizations that pay attention to the new direction of journalism data. Editors-in-chief must learn quickly, identify technical and policy issues for medium to long term development for media institutions. In terms of technical issues, Wu, Tandoc Jr. & Salmon (2019) saw the importance of technology companies that support media institutions to transform from traditional conceptions to data journalism. Furthermore, social media is now the foundation of citizens to participate in democracy. Interaction building of communication and reportage as citizen journalism enables citizens to transform the society, and according to Pain & Chen (2019) from the results of their research in India, this is a form of participatory monitoring and evaluation.

One aspect that needs attention is the policy and management of resources, including human capital in the field of media institutions. Parasie (2015) argues that the efforts of news organization resources are needed to generate valid trust because data processing artifacts constitute a significant component of the big new data editorial space environment. According to Parasie (2015), even the use of this artifact, for journalists, must follow two opposite paths to produce reliable news, namely the "hypothesis-driven" path and the "data-driven" pathway. Along with that, Heravi (2019) also paid attention to the human capital of data journalism, which was considered weak in interdisciplinary insight.

Experience in forging data journalism allows each journalist to become proficient and indepth in knowledge quickly. The quality of works that continues to improve is a good benchmark. In journalism data, the quality of the work and presentation is understood as the result of better, faster, and more open data access. According to Hammond (2017), in the context of big data, human subjectivity tends to be downgraded, and that is an important issue. However, the increase in 'datafication' is not what drives change in the profession but is a continuation of the preexisting, shaped by the broader context of contemporary conditions.

Media institutions must also continue to innovate, have influence, and solve the constraints as stated by Hewett (2017, the technical capabilities and knowledge of human resources in the world of journalism must adopt data journalism. The complexity associated with the demands of data journalism is very urgent and must be able to identify critical problems and play a role in information technology. Usman Kansong, Media Group News Director in charge of *MetroTV* and *Media Indonesia*, said:

At Media Group, we pride ourselves on being national and regional media with diverse cultures that are brought to the life of our organizations. We pay close attention to human capital policy and management as the main aspects of the operational and strategic media institutions. Many Metro TV and Media Indonesia alumni were successful, and they were promoted in various similar companies. We are proud to be very concerned about the continuous improvement of the ability of journalists, especially in data journalism.

As a comparison, one of the researches conducted in the United States was conducted by Fink & Anderson (2015) on beyond the "usual suspects" with the finding that there was great variety of educational а backgrounds, skills, tools, and goals among data journalists. However, many of them faced similar struggles, such as trying to define their roles in their organizations and manage scarce resources. Borges-Rey (2016) discussed data journalism practices in the UK with semistructured interviews with data journalists, data editors, and news managers working for the UK mainstream media. Innovation is needed to be synergistic with data journalism to storytelling, news, and news dissemination. Journalist resources are significant, and some of the large groups of journalists must focus as data-journalists as Wright & Doyle (2019) concern, who investigated how and why data journalism had developed in Australia.

Research, through semi-structured interviews with journalists, led them to suggest that there should be a core group of data journalists in media institutions. Previously, from research in the UK, Knight (2015) found that the skills of journalists following computer technology were still considered low. The findings of Borges-Rev (2016)showed that most professional journalists demonstrated a degree of doubt about numbers and computational literacy. There are demands of an industry that is growing with technology, which must also be followed by human competencies.

Appelgren, Lindén& Van Dalen (2019) also paid attention to the future of cross-continent data journalism resources. By comparing data journalism in Africa, Arabic, Italy, England, and Argentina, they concluded that three important contextual factors shape data journalism throughout the world, namely journalistic culture, media markets, and the political environment. In the future, the wave of post-cellular technology referred to as the progress of the vector of virtual reality (VR) by Hassan (2020) creates an 'empathy machine' that will form the basis of new journalism as used by the New York Times and the United Nations.

Finally, the collaboration between media institutions and citizens takes place, especially when news organizations struggle in difficult economic times. Konieczna (2020) calls a community collaboration with the term Stepladder Collaboration, namely "simple sharing, collaboration on stories, collaborating on methods, organizing events, allocating money, and working to solve a problem in the community."

CONCLUSION

Big data as a basis for compiling news and articles is an inevitable source in the online digital era. Data is continuously growing in various cloud computing centers from all walks of life. Journalists in the media and even citizen journalists must be able to analyze big data to be presented in data-driven journalism mode. Several steps must be taken to process large amounts of unstructured data into structured data to be validated, combined with qualitative aspects. The crafting steps of credible news or articles are trusted and used as a reference by the public, government, or related communities.

The results of this research showed that big data analytics was conducted for data-driven journalism at national and local media in Indonesia. The data-driven journalism was routinely used in the production process and presentation. Data journalism is applied to online and conventional media, both in national and local media scopes in various regions. However, the knowledge and understanding of big data analytics for datadriven journalism were still inadequate in the newsroom and research supporter division. Several media in Indonesia apply data journalism by taking data from outside media institutions and are still lacking in qualitative enrichment. Journalists need understanding, knowledge, and skills to produce visual quantification so that journalism data has a future in an increasingly competitive market. Ordinary people as new media participants also have not apprehended journalism data but only practice doing it to share information and entertainment.

This study suggests the model development of the data-driven journalism process from Indonesia's experience, which can be a guide for journalists. The managerial implication is that the newsrooms must empower the datadriven journalists' team and research division at media institutions.

Due to the observation and content analysis that was limited on media series document samples, future research on this topic could be performed with ethnographic methods in various media institutions, including in the process and interaction in the newsroom.

REFERENCES

Appelgren, E., Lindén, C.-G., & van Dalen, A. (2019). Data Journalism Research: Studying a Maturing Field across Journalistic Cultures, Media Markets and Political Environments. *Digital Journalism*, 7(9), 1191–1199.

Appelgren, E., & Nygren, G. (2014). Data Journalism in Sweden: Introducing new methods and genres of journalism into "old" organizations. Digital Journalism, 2(3), 394-405.

Appelgren, E. (2018). An illusion of interactivity: The paternalistic side of data journalism. Journalism Practice, 12(3), 308-325.

Baack, S. (2018). Practically Engaged: The entanglements between data journalism and civic tech. Digital Journalism, 6(6), 673-692.

Barton, A. (2016). Big data. Journal of Nursing Education, 55(3), 123–124. http://doi.org/10.3928/01484834-20160216-01

Bhatt, V., Goyal, K., & Yadav, A. (2018). The authenticity of social media information among youth: an Indian perspective. Journal of Content, Community & Communication, 4(8), 42-45.

Borges-Rey, E. (2016). Unraveling data journalism: A study of data journalism practice in British newsrooms. Journalism Practice, 10(7), 833-843.

Borges-Rey, E. (2017). Towards an epistemology of data journalism in the devolved nations of the United Kingdom: Changes and continuities in materiality, performativity, and reflexivity. Journalism, 1464884917693864.

Bradshaw, P. (2014). Data journalism. Ethics for digital journalists: emerging best practices, 202-219.

Brandtzaeg, P. B., Lüders, M., Spangenberg, J., Rath-Wiggins, L., & Følstad, A. (2016). Emerging journalistic verification practices concerning social media. *Journalism Practice*, 10(3), 323-342.

Brennen, J. S., & Kreiss, D. (2016) Digitalization. In: Jensen, K. B., Rothenbuhler, E. W., Pooley, J. D., & Craig, R. T. (eds.). The International Encyclopedia of Communication Theory and Philosophy. Chichester, West Sussex: Wiley-Blackwell, pp. 556-566.

Coddington, M. (2015). Clarifying journalism's quantitative turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. Digital journalism, 3(3), 331-348.

Cushion, S., Lewis, J., & Callaghan, R. (2017). Data journalism, impartiality, and statistical claims: Towards more independent scrutiny in news reporting. Journalism Practice, 11(10), 1198-1215.

Carlson, M. (2018). Facebook in the news: Social media, journalism, and public responsibility following the 2016 trending topics controversy. *Digital journalism*, *6*(1), 4-20.

Carlson, M. (2015). The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. Digital journalism, 3(3), 416-431. Caswell, D., & Dörr, K. (2018). Automated Journalism 2.0: Event-driven narratives: From simple descriptions to real stories. Journalism Practice, 12(4), 477-496.

Chan, J. O. (2013). An architecture for Big data analytics. Communications of the IIMA, 13(2), 1.

Cheruiyot, D., Baack, S., & Ferrer-Conill, R. (2019). Data journalism beyond legacy media: The case of African and European civic technology organizations. Digital Journalism, 7(9), 1215-1229.

Dörr, K. N. (2016). Mapping the field of algorithmic journalism. Digital Journalism, 4 (6): 700-722.

De Maeyer, J., Libert, M., Domingo, D., Heinderyckx, F., & Le Cam, F. (2015). Waiting for data journalism: A qualitative assessment of the anecdotal take-up of data journalism in French-speaking Belgium. Digital journalism, 3(3), 432-446.

Fink, K., & Anderson, C. W. (2015). Data Journalism in the United States: Beyond the "usual suspects." Journalism Studies, 16(4), 467-481.

Gerard, G., Haas, M., & Pentland, A. (2014). Big data and Management. Academy of Management Journal, 57(2), 321–326.

Günther, E., & Quandt, T. (2016). Word counts and topic models: Automated text analysis methods for digital journalism research. Digital Journalism, 4(1), 75-88.

Gray, J., Chambers, L., & Bounegru, L. (2012). The data journalism handbook: how journalists can use data to improve the news. " O'Reilly Media, Inc.."

Hammond, P. (2017). From computer-assisted to data-driven: Journalism and Big data. Journalism, 18(4), 408-424.

Hassan, R. (2020). Digitality, virtual reality and the 'empathy machine'. *Digital Journalism*, 8(2), 195-212.

Henry, R., & Venkatraman, S. (2015). Big data Analytics: The next big learning opportunity. Academy of Information and Management Sciences Journal, 18(2), 17–29.

Heravi, B. R. (2018, March). Data journalism in 2017: A summary of results from the global data journalism survey. In International Conference on Information (pp. 107-113). Springer, Cham.

Heravi, B. R. (2019). 3Ws of Data Journalism Education: What, where, and who? Journalism Practice, 13(3), 349-366.

Heravi, B. R., & Harrower, N. (2016). Twitter journalism in Ireland: sourcing and trust in the age of social media. *Information, communication* & society, 19(9), 1194-1213.

Hermida, A., & Young, M. L. (2019). *Data journalism and the regeneration of news*. Routledge.

Hermida, A., & Young, M. L. (2017). Finding the data unicorn: A hierarchy of hybridity in data and computational journalism. Digital Journalism, 5(2), 159-176.

Hewett, J. (2016). Learning to teach data journalism: Innovation, influence, and constraints. Journalism, 17(1), 119-137.

Jamil, S. (2019). Increasing Accountability Using Data Journalism: Challenges for the Pakistani Journalists. *Journalism Practice*, 1-22.

Johnson, B. G., & Dade, C. (2019). Local broadcast journalism, user-generated content and boundary work. *Media Practice and Education*, 20(3), 260-276.

Kabha, R., Kamel, A., Elbahi, M., & Narula, S. (2019). Comparison Study between the UAE, the UK, and India in Dealing with WhatsApp Fake News. *Journal of Content, Community & Communication*. 10(5), 163-186.

Kalyvas, J. R., & Overly, M. R. (2014). Big data: A business and legal guide. CRC Press.

Knight, M. (2015). Data journalism in the UK: a preliminary analysis of form and content. Journal of Media Practice, 16(1), 55-72.

Konieczna, M. (2020). The Collaboration Stepladder: How One Organization Built a Solid Foundation for a Community-Focused Cross-newsroom Collaboration. *Journalism Studies*, 1-18.

Lewis, S. C. (2017). Journalism in an Era of Big data: Cases, concepts, and critiques. New York, NY: Routledge.

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. 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.

Lewis, S. C., & Westlund, O. (2015). Big data and journalism: Epistemology, expertise, economics, and ethics. Digital journalism, 3(3), 447-466.

Loosen, W., Reimer, J., & De Silva-Schmidt, F. (2017). Data-driven reporting: An on-going (r) evolution? An analysis of projects nominated for the Data Journalism Awards 2013–2016. Journalism, 1464884917735691.

Mayer-Schönberger, V., & Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.

Narula, S. (2019). India and Data Journalism Revolution, an editorial. *Journal of Content, Community & Communication*. 10(5), 1.

Nielsen, R. K., Cornia, A., & Kalogeropoulos, A. (2016). Challenges and opportunities for news media and journalism in an increasingly digital, mobile, and social media environment. *Mobile, and Social Media Environment* (December 1, 2016).

Pain, P., & Chen, G. M. (2019). 'A chance for me to do good, make a real difference': how citizen journalists in India view their role in social transformation. *Media Practice and Education*, 20(4), 334-349.

Parasie, S. (2015). Data-driven revelation? Epistemological tensions in investigative journalism in the age of "big data." Digital Journalism, 3(3), 364-380.

Parasie, S., & Dagiral, E. (2013). Data-driven journalism and the public good: "Computerassisted-reporters" and "programmerjournalists" in Chicago. New media & society, 15(6), 853-871.

Rinsdorf, L., & Boers, R. (2016, July). The need to reflect: Data journalism as an aspect of disrupted practice in digital journalism and journalism education. In Proceedings of the Roundtable Conference of the International Association of Statistics Education (IASE), Berlin, July 2016.

Royal, C., & Blasingame, D. (2015). Data journalism: An explication. In International

Symposium in Online Journalism (Vol. 5, No. 1, pp. 24-46).

Stalph, F. (2018). Classifying Data Journalism: A content analysis of daily data-driven stories. Journalism Practice, 12(10), 1332-1350.

Strohbach, M., Daubert, J., Ravkin, H., & Lischka, M. (2016). Big data storage. In New horizons for a data-driven economy (pp. 119-141). Springer, Cham.

Tandoc Jr, E. C., & Oh, S. K. (2017). Small departures, big continuities? Norms, values, and routines in The Guardian's big-data journalism. Journalism Studies, 18(8), 997-1015.

Tong, J., & Zuo, L. (2019). The Inapplicability of Objectivity: Understanding the Work of Data Journalism. *Journalism Practice*, 1-17.

Uskali, T. I., & Kuutti, H. (2015). Models and streams of data journalism. The Journal of Media Innovations, 2(1), 77-88.

Veglis, A., & Maniou, T. A. (2018). The Mediated Data Model of Communication Flow: Big data and Data Journalism. *KOME: An International Journal of Pure Communication Inquiry*, 6(2), 32-43.

Weaver, D. H., & Willnat, L. (2016). Changes in US journalism: How do journalists think about social media? *Journalism Practice*, 10(7), 844-855.

Witschge, T., Anderson, C. W., Domingo, D., & Hermida, A. (Eds.). (2016). The SAGE handbook of digital journalism. Sage.

Wright, S., & Doyle, K. (2019). The Evolution of Data Journalism: A Case Study of Australia. *Journalism Studies*, 20(13), 1811-1827.

Wu, S., Tandoc Jr, E. C., & Salmon, C. T. (2019). When journalism and automation intersect: Assessing the influence of the technological field on contemporary newsrooms. *Journalism Practice*, *13*(10), 1238-1254.

Wu, X., Zhu, X., Wu, G.-Q., & Ding, W. (2014). Data mining with big data. IEEE Transactions on Knowledge and Data Engineering, 26(1), 97–107.

Young, M. L., Hermida, A., & Fulda, J. (2018). What makes for great-data journalism? A content analysis of data journalism awards finalists 2012–2015. Journalism Practice, 12(1), 115-135.

Zhang, S., & Feng, J. (2019). A Step Forward? Exploring the diffusion of data journalism as journalistic innovations in China. Journalism Studies, 20(9), 1281-1300.
