Kajian Jurnalisme
Volume 06 No. 01 July 2022
https://doi.org/10.24198/jkj.v6i1.37167

Verification and Data Processing of Open Data for Data-Driven News Stories in *Lokadata.id* and *Katadata.co.id*

Amalia Nurul Muthmainnah, Maulana Arief, Nurul Fitriyani
Faculty of Social and Political Sciences, Universitas 17 Agustus 1945 Surabaya
anmuthmainnah@untag-sby.ac.id

Abstract

With the existence of open data, journalists have additional investigative tools to report on issues in the public interest and to hold the government to account. However, it remains outside the mainstream of journalism due to the limitation to access and work with data. There are indeed a handful of studies that examine the implementation of open data portals in Indonesia but rarely give attention to how data journalists are using these data and their perceptions about the quality of data provided by the government. This study is intended to see the process of how data journalists in Indonesia, particularly in Lokadata.id and Katadata.co.id, turn data into news stories. Through research that uses a descriptive qualitative approach and collects data through semi-structured in-depth interviews, we found that data provided by the authority sources (government) are still the main sources for data journalists in Indonesia. Even though the quality of data is getting better, most journalists in this study argued that the government's open data has not yet been integrated and has no basic standard that is collectively used. Therefore, the verification process and data processing in each media are important. The most common ways that journalists use to verify the data are by exploring the raw data, contacting the institution that provided the data, and considering the reputation of the data sources. However, there is a tendency of data journalists to be passively relying on official sources and 'hide behind' the perceived credibility of their sources.

Keywords: data; journalism; media; online; verification

Abstrak

Dengan adanya keterbukaan data (open data), jurnalis memiliki alat investigasi tambahan untuk melaporkan masalah-masalah yang menjadi kepentingan umum, sekaligus untuk meminta pertanggungjawaban pemerintah. Sayangnya, pemanfaatan open data masih berada di luar arus utama jurnalisme karena jurnalis sering menghadapi hambatan dalam mengakses dan mengolah data. Beberapa studi terdahulu memang telah mengkaji implementasi portal data terbuka di Indonesia, namun jarang yang memberikan perhatian pada bagaimana jurnalis menggunakan data ini dan persepsi mereka mengenai kualitas data yang disediakan oleh pemerintah. Penelitian ini lantas bertujuan untuk melihat proses yang dijalani jurnalis Indonesia, khususnya di Lokadata.id dan Katadata.co.id, dalam mengolah data terbuka sebagai berita. Hasil dari penelitian yang menggunakan pendekatan kualitatif deskriptif dengan metode pengumpulan data melalui wawancara mendalam semi-terstruktur ini menunjukkan bahwa data dari sumber otoritas (pemerintah) masih menjadi sumber utama jurnalis data dalam menyusun berita berbasis data. Meski kualitasnya semakin membaik dari tahun ke tahun, mayoritas jurnalis dalam penelitian ini menyebutkan bahwa data terbuka pemerintah Indonesia masih kurang terintegrasi dan tidak memiliki standar baku yang digunakan bersama. Maka, jurnalis data perlu berhati-hati dalam proses gatekeeping untuk menjaga kualitas berita yang dihasilkan. Secara umum, upaya verifikasi yang dilakukan jurnalis adalah menelusuri data mentah (raw data), menghubungi lembaga penyedia data, dan mempertimbangkan reputasi lembaga penyedia data. Namun, terdapat kecenderungan jurnalis data bergantung secara pasif pada sumber resmi dan bersembunyi di balik kredibilitas sumber data mereka.

***Kata kunci**: daring; data; jurnalisme; media; verifikasi*

Correspondence: Amalia Nurul Muthmainnah, S.I.Kom., MA., Faculty of Social and Political Sciences Universitas 17 Agustus 1945 Surabaya, Jl. Semolowaru No. 45 Menur Pumpungan Surabaya 60118, Email: anmuthmainnah@untag-sby.ac.id

Submitted: December 2021, **Revised**: April 2022, **Accepted**: April 2022, **Published**: July 2022
ISSN: 2549-0559 (cetak), ISSN: 2549-1946 (online), Website: http://jurnal.unpad.ac.id/kajian-jurnalisme

Volume 06 No. 01 July 2022
https://doi.org/10.24198/jkj.v6i1.37167

INTRODUCTION

The development of data journalism has raised the debate about the quality and openness of data provided by public institutions. Journalism has always had the main role to act as a watchdog on a de facto democracy. The data journalism community, especially with the advancement of digital technology, has played an important part in this regard, bringing stories to light that would otherwise have remained hidden (Felle, 2016). With open data, journalists have an additional investigative tool to report on issues in the public interest and to hold the government into account (Porlezza & Splendore, 2019). Through strong data analysis, journalism stories will be more detailed, interesting, and credible, as data journalism can sharpen the news angle before the stories are published (Imawan, 2018). Nevertheless, open data remains outside the mainstream of journalism due to the limitation to access and processing of data (Stoneman, 2015).

The open data initiative in Indonesia began in 2010 due to the mandate of UU No. 14 Tahun 2008 concerning Public Information Disclosure. One of the implementations can be seen in the establishment of the Portal Satu Data Indonesia (https://data.go.id/) which was launched in 2019. However, in reality, open data initiatives are often nothing more than "data dumping" (Stagars, 2016). Several studies show that open data portals in the world, especially in Asia, are implemented without the right policy basis, where the depth and quality of datasets are still inadequate (Stuart, 2011).

There are indeed a handful of studies discussing the implementation of open data portals in many countries, including Indonesia. However, only a few paid attention to how data journalists use this data and their perception of the quality of data provided by the government. Whereas, the implementation of data journalism is in line with the importance of applying the principle of open data (Imawan, 2018). Especially with the fact that the availability of data does not automatically improve democracy. The good news is that the media and journalists in Indonesia respond to the trend of data journalism quite well. This can be seen in the increasing number of media companies that produce data-driven news stories. There are even several media that focus on data journalism products. As shown in figure 1, based on a research report published by the Aliansi Jurnalis Independen (AJI) Indonesia, the implementation of data journalism started in 2001.

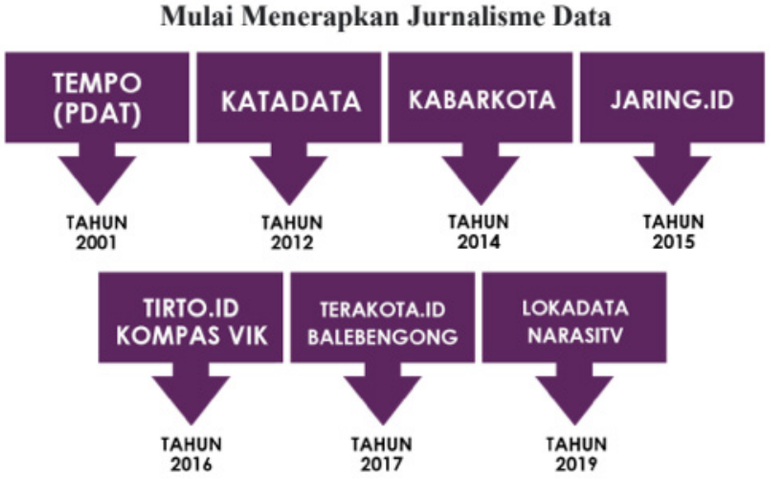

Figure 1 Media companies in Indonesia that implement data journalism
Source: (Danayanti, Wardhana, Marsiela, & Galuh, 2021)

From the listed media in figure 1, our research focuses on digital native media that has consistently used a data journalism approach in the news production process since the very beginning. Digital native media is the type of media that only operates online, and has a specific implementation of news values and a distinctive approach to covering news (Salaverría, 2020). As a pure player with a digital nature, these media outlets are different from legacy brands in terms of business model, distribution strategies, corporate organization, and editorial priorities (Vara-Miguel, 2020). Based on figure 1, the example of legacy media is Tempo and Kompas, who at the beginning were print media companies and later expanded their companies by also creating online media Tempo.co and Kompas.id/Kompas.com. Meanwhile, the rest are digital native media. Among others, we decided to specifically study Lokadata and Katadata. Although in figure 1 it is written that Lokadata was only established in 2019, it is worth mentioning that Lokadata is replacing Beritagar which was established in 2015. Therefore, both of these media are the early adopters of data journalism in their organization. Both of these media also have fairly good popularity with a huge number of audiences.

This research explores the implementation of open data initiatives in Indonesia from the perspective of data journalists. This study seeks to identify the perceptions of data journalists in Indonesia towards the implementation of open data by the government, the ease and obstacles faced by data journalists when accessing open data provided by the Indonesian government, as well as the process carried out by data journalists in verifying and processing open data as news material. Since previous research has shown that the implementation of open data by the Indonesian government is still not optimal, this research also explores the efforts made by data journalists when open data provided by the government is considered inadequate.

Data journalism itself is not new to the world of journalism. For a long period of time, journalists have used data as news material. The term data journalism became popular thanks to The Guardian's Datablog in 2009 (Stalph & Borges-Rey, 2018). However, the practice of data journalism itself is said to have originated from what is known as Computer-Assisted Reporting (CAR), which is a technique where surveys, content analysis, and statistics are used to validate journalists questions (Coddington, 2015). One of the journalists who first realized that databases were a key technology in journalistic practice was Philip Meyer, whom he combined social science, journalism, and technology to produce news (Parasie & Dagiral, 2013). In 1967, Meyer reported on the civil rights prosecutions in Detroit, one of the most brutal demonstrations in US history, in which 43 people died and more than 7000 were arrested. Through a statistical survey, Meyer managed to reveal that college graduates are just as vulnerable to triggering riots as those who only graduated from high school (Knight, 2015). What Meyer did show was how data can provide context to an event so that the public can understand an issue more deeply. Thus, it can be concluded that data journalism can encourage better reporting quality by using statistical analysis techniques which are able to provide deeper insight into a report and highlight relevant data (Danayanti et al., 2021).

Technological developments bring a huge impact on the advancement of data journalism which is increasingly being implemented in the newsroom. It allows the use of digital devices to browse data and produce visual content that attracts readers attention. Data journalism can provide a personal story approach to the audience because an event can be developed by journalists through interesting visualizations such as graphs and maps (Danayanti et al., 2021). This visualization can encourage the audience to have different cognitive experiences that enable them to broaden their horizons and understanding of information (Segel & Heer, 2010). It is certainly relevant in this era of information overload, where there is an abundance of fake news circulating and often an event is only reported at the surface level. Reports based

Volume 06 No. 01 July 2022
https://doi.org/10.24198/jkj.v6i1.37167

on data journalism are expected to be able to play a significant role in changing norms in society, increasing public trust in the media, and being a solution to the increasingly worrying problem of fake news (Sanusi, 2018).

What distinguishes the use of data in journalism between the conventional era and the current digital era is that data is no longer only complementary to news stories. Now, data has become the main source and the initial process for journalists in compiling news. Data is the spearhead of the reporting process from the planning to evaluation stages. Journalists no longer need to always go to the field and interview sources. With data, an in-depth and quality report can be produced by mass media companies. David McCandless, a data journalist from the UK, said that the use of data can shift the role of journalists from being considered '*the first person to report a story*' to being '*the first person to report what happened*'. Therefore, data journalism is closely related to investigative news. One of the most popular data journalism works that were done jointly by journalists from various countries is the Panama Papers, which was released in 2016. Tempo is one of the media from Indonesia involved in this global-scale investigative project.

A study that aims to find out why and how media digitization is used in data journalism in Tempo revealed eight reasons why Tempo applies data journalism in the investigative process, namely: (1) Realizing and maximizing the vision of a fully digital newsroom; (2) Strengthening Tempo's investigative differentiation; (3) Empowering media through studying and seeking innovations; (4) Fighting the phenomenon of hoax through changes in news production patterns; (5) Help find phenomena or patterns that occur in the community; (6) Making investigations more objective; (7) Contains issues that are more relevant to the public and are not driven by the interests of certain groups; and (8) the business prospects and practices of data journalism that will continue to grow (Asprilla & Maharani, 2019). In contrast to this research, our study seeks to dig up information not only on Tempo journalists but on journalists at online media companies in Indonesia, which from the very beginning have specialized in producing data-driven news stories. We also do not focus on the implementation of data journalism in the newsroom but pay more attention to how data journalists make use of open data provided by the Indonesian government. This is considering that previous research shows that the availability of data in Indonesia is still sporadic and valid data is difficult to obtain.

RESEARCH METHOD

This research uses a descriptive qualitative approach with a case study method on two online media that implement data journalism, i.e. Lokadata.id and Katadata.co.id. Both are based and located in Indonesia. Of course, other media also use data as news material, but Lokadata and Katadata are digital native media that have implemented data journalism since their establishment as media companies. This is the main consideration why the data team and the editors in these two media become informants of this research. Another reason why we focus only on online media is the principle of data journalism which is considered to conflict with online journalism. The process of collecting and analyzing data takes quite a long time, while online journalism focuses on speed (Widiantara, 2021). It is certainly interesting to examine how data journalists in online media deal with this contradiction.

The data for this research was collected through in-depth semi-structured interviews, where we prepared a list of questions before interviewing the informants, but this list is flexible and open to following the answers from the informants. In each of the selected media, we interviewed 5 (five) informants. The duration of the interview is one to two hours. Interviews were conducted from August to September 2021. At Lokadata, informants were part of the

Data & Research Team consisting of the Chief Data Officer, Head of Data Logistics, Data Scientist, and Data Analyst. While at Katadata, the informants were part of the data journalism team consisting of the Managing Editor, Editor, Reporter, and Data Engineer. The following table shows the code and positions of the informants. We will use this code throughout the article when quoting their statements.

Table 1 List of informants position and code

No	Media Company	Code of Informant	Informant's Position
1	Katadata	Informant 1	Managing Editor
2		Informant 2	Editor
3		Informant 3	Reporter
4		Informant 4	Reporter
5		Informant 5	Data Engineer
6	Lokadata	Informant 6	Chief Data Officer
7		Informant 7	Head of Data Logistics
8		Informant 8	Senior Data Analyst
9		Informant 9	Data Analyst & Content
10		Informant 10	Data Scientist

Source: Researcher's Interview Result, 2021

Considering the COVID-19 pandemic situation and the imposition of mobility restrictions since the end of June 2021, the interviews in this study were conducted online using the Zoom Meeting application. Interviews were recorded using the record feature in the application. Each recorded interview was listened to and then transcribed. The interview transcripts were then carefully summarized into several topics. These topics are included in specific categories using simple coding techniques as part of the data reduction phase. The data that had been categorized was analyzed which led to conclusions being drawn. To ensure the validity of the data, this study used a triangulation technique. The four aspects of triangulation are (1) Data triangulation; (2) Investigator Triangulation; (3) Theory Triangulation; and (4) Triangulation Methodology (Zamili, 2015). For the first aspect, triangulation is carried out by interviewing more than one informant and carried out at different times. For the investigator aspect, two people became interviewers in this study to minimize the subjectivity of the researcher. Of course, theoretical and methodological triangulation is also carried out where we compare the results of research interviews with relevant theories and previous studies.

RESULTS AND DISCUSSION

After conducting interviews with ten people involved in the production process of data-driven news stories at Lokadata and Katadata, we acquired the details on how data journalists accessed, processing, and verified the data before the news stories were published in these two media. The discussion of our findings is divided into three sub-chapters below.

The Sources of Data for News Stories

As mentioned in the introduction, data is no longer just a complement to the news text that is visually displayed. Data is now the main source, even the only source used by journalists in producing news. When data becomes the main source of news, it is certainly important to examine what data sources are used by journalists and how journalists access and process the

Volume 06 No. 01 July 2022
https://doi.org/10.24198/jkj.v6i1.37167

data. The results of our interview with the informants are in line with previous research entitled "Menilai Penerapan Jurnalisme Data dan Investigasi Berbasis Data di Indonesia (Assessing the Practice of Data Journalism and Data-Based Investigation in Indonesia)" (Danayanti et al., 2021).

In general, journalists use data owned by the government as the main source of their news reporting. Lokadata in particular mostly uses research data from Badan Pusat Statistik (BPS). The tendency to refer to BPS data compared to data from other government agencies is based on several reasons. One of the reasons given by Lokadata's Data Analyst is the excellent reputation of BPS as a research institution (Informant 9, Interview. 2021, August 13). Informant 9 mentioned that BPS received an award as one of the best research institutions in Asia. This reputation can be seen in the transparency of the data collection methods carried out by BPS. Informant 9 added that it was easier to validate BPS data because it clearly stated the number of respondents and the data collection process. The data obtained is also still in raw format, making it easier for journalists to process it. This is something rarely done by other government agencies.

Other government agencies often only share data in the form of publications, which means the data has been processed so that journalists do not have access to the raw data. When in fact, the existence of raw data that is shared openly is very important so that various parties have the opportunity to interpret the data and generate their knowledge from the data (Baack, 2015). Without raw data, the information submitted to the public is only based on the interpretation of one party, in this case, the government's interpretation. This situation can affect how journalists contextualize and trust the data provided by the government. This is something that needs government attention as the main data source for journalists because open data is not only "making data public, but also the making of public data" (Gray & Lämmerhirt, 2019).

Furthermore, the data with publication format is considered impractical for journalists. The Head of Lokadata's Data Logistics Unit stated, "Several government agencies share their data in PDF form. This of course makes the data cannot be directly processed by us. We have to change the format first and perform data cleaning" (Informant 7, Interview. 2021, August 16). Informant 1 mentioned a similar issue. His team often had to retype the data obtained from government agencies to then reprocess it through the various applications they use, such as Microsoft Excel, R, or Python (Informant 1, Interview. 2021, August 20). It is important to note that the use of third-party applications to process this data can also bring its dilemma because there is a risk that there will be a change in the model in the application which can have an impact on the loss of datasets used by data journalists (De-Lima-Santos, Schapals, & Bruns, 2021). The availability of raw data is crucial as it not only makes the data processing easier, but it is also easier for journalists to do verification if they feel there is something bizarre with the data, which the Lokadata team calls '*outliers*'. If they find this, they will re-check the raw data to see what aspects or components might be the cause of the discrepancy.

In addition to the format and transparency of data, Lokadata tends to use BPS data because the scope of its research is generally much larger than other government agencies. Data Scientist in Lokadata instantiated, "The BPS research involves millions of respondents. Meanwhile, other agencies are usually still on a scale of thousands of respondents" (Informant 10, Interview. 2021, August 16). While Lokadata uses a lot of data produced by BPS, the Katadata team uses more diverse data sources. However, these two media both use government data as the main source of news because the government is considered the party that has authority over various issues.

Referring to the previous research results, the main obstacle experienced by data journalists is the limited data sources to be processed into news (Danayanti et al., 2021). We asked our informants if they face a similar problem and what is their solution to solve this. Our finding shows that when experiencing this kind of obstacle, what journalists usually do is look for data from other sources. For instance, data released by private institutions, non-profit institutions, survey institutions, or scientific journals. If it turns out that the data needed for a particular issue cannot be found, the editorial team decides to cancel the discussion on the issue or discuss the issue from another news angle. As stated by Informant 10, generally the determination of the topic to be covered was adjusted to the availability of data held by the media. When proposing a topic at a meeting, the data team first looks at what data they have. Then they look for interesting things to report from the data. As a Data Scientist whose job is to process raw data (data that still uses machine language), Informant 10 also plays a role in the process of framing an issue. Before providing the data to the Data Analyst, Informant 10 usually has determined what variables are interesting to be appointed as news stories. The data journalism workflow is indeed possible in two ways. Either the dataset provides information on the news that has been designed in the newsroom or the dataset serves as the starting point for the whole news (Widiantara, 2021).

On the other hand, based on the answer from the informants of this study, we conclude that the availability of data in Indonesia is quite good. However, not all data is freely accessible to the public. Lokadata itself collaborates with several government agencies, one of which is BPS, where they can gain more access to certain data that needs to be paid for. Another effort that can be made to obtain data is by conducting independent research. However, both Lokadata and Katadata stated that this is very rarely done. This is because independent research requires a large amount of time, energy, and money. What these two media generally do is take advantage of the results of commercial research conducted by their agencies for their clients.

In this digital era, the online media business model has indeed shifted. The media can no longer completely depend on the traditional advertising model as their main and only source of income. They need to innovate through multiple revenue streams, namely *Advertising* (Programmatic ads, Native ads, Sponsored content), *Vendor* (Merchandise, Partnerships, Subscription), and *Services* (Events, Benefit clubs, Crowdfunding, Education projects, Reports, and data compilation) (De-Lima-Santos & Mesquita, 2021a). In data journalism-based media such as Lokadata and Katadata, their main business model is through the *Services* component where they do data-based projects in collaboration with certain clients. These clients ask data journalism media companies to conduct research or process data based on their needs. If those needs turn out to coincide with a topic that can be raised as news, then the editorial team can use client research data to become their source of data. This of course depends on the form of the cooperation contract between the media company and its clients. Not all research data for client projects can be used as news material. In Lokadata, what they do is utilize research data that is not used by the client or not part of the client's main interest. Meanwhile, in Katadata, what the editorial team often does is ask for help from the research team called Katadata Insight Center when working on a client's research project (Informant 1, Interview. 2021, August 20). Thus, when the Katadata Insight Center research team went into the field, there were several additional questions from the data journalism team beyond the client's needs that were included in their research.

Furthermore, when being asked about the Indonesian government's implementation of open data, almost all of the informants gave the same answer. In general, the informants of this research considered that Indonesian government data was still less integrated between

Volume 06 No. 01 July 2022
https://doi.org/10.24198/jkj.v6i1.37167

one agency and another. For instance, the data from agency A is different from agency B, even though they both discuss the same matter. Informant 1 from Katadata gave an example of the differences in farmer production data in Indonesia by Bulog and the Ministry of Trade. According to Informant 7 from Lokadata, this often happens because of differences in methods or area codes used by these agencies. There is still no common standard that is shared between government agencies in Indonesia. Another problem, according to Informant 10, is the different times on updating the data by an agency. Sometimes the central and regional levels of a government agency do not update their data at the same time. The central level may have updated the data this year, while at the regional level the data for this year has not yet been published. This is why it seems that there is a difference between data at the central level and data in the regions. In cases like this, journalists have a responsibility to verify the data and clarify this issue to their audiences.

Verification Process of Data-Driven News Stories

As described in the previous sub-chapter, one of the problems often experienced by data journalists is the difference in data between one agency and another. This data discrepancy is certainly a big problem because it can confuse the public, and even give the public wrong information. Therefore, it is important to know how the strategy of data journalists in overcoming this problem. The first thing that is generally done by our informants is to check the research method of the data, considering that sometimes the difference in data is due to different data collection methods. This kind of situation is where data journalists come into play. Open data can be accessed by anyone, but not everyone bothers to understand the data. So, this is where data journalists play a role in finding, translating, and interpreting statistical figures (Cushion, Lewis, & Callaghan, 2017).

This is also a problem of the Indonesian government that needs to be addressed immediately. Not all agencies transparently show '*behind the scenes*' of the data that they published. As mentioned by Informant 9, several government agencies only provide data that is already in the form of publications. This is different from what is done by BPS, governments of other countries, or international institutions which are more transparent in explaining the process of obtaining the data. Despite the promises mandated by the law, the Indonesian government is indeed still facing challenges in its open data practices. Research findings on challenges and critical success factors of Satu Data Indonesia implementation have shown the problems are in four tenets: (1) Data Planning: The priority data has not yet been determined; (2) Data Collection: There is still the difficulty of integrating data; (3) Data Inspection: Noncompliance with data standards, non-uniform metadata, and the absence of data reference code; (4) Data Dissemination: The non-readiness of digital infrastructures in terms of integrated data center and the development of an application that can be used by all central and local government agencies (Islami, 2021). There is also the challenge of lack of data competence by those operating the system about open data. Therefore, to increase the integrity of their data in the eyes of data journalists, the steps that could be taken by the Indonesian government are increasing transparency where citizens and news workers can scrutinize the data, developing services by third parties that can benefit citizens and companies, and establishing new services that spur the economy (De-Lima-Santos & Mesquita, 2021b).

Another strategy that data journalists do to verify data is to confirm directly with the institutions who published the data. Informant 7 explained that this confirmation can be done in two ways. Namely, go directly to the institution or contact them via email. The first way is more effective and efficient because it is faster for them to meet directly with the parties

responsible for the data. However, during this pandemic, it is not possible to go directly to the institution's offices. So, most of the time, they have to wait for an email reply from the agency that published the data. This takes longer because it usually requires approval from not just one person. When working from home, the responsible people are not in the same location, which makes the process more time-consuming.

Speaking of data verification, of course, it is not supposed to be only done when different data is found between government agencies. Ideally, journalists should always double-check any data they obtain, before publishing the data to the public. However, from the results of interviews with our informants, journalists tend not to feel the need for certain verification when there is no discrepancy in the data. Journalists feel that any data published by the agency must have gone through a separate verification process. "I do not think someone can falsify data and then publish it. Before being published, of course, the agency has gone through a long process of data processing" (Informant 1, Interview. 2021, August 20). Moreover, our informants also feel that the responsibility for the validity of the data still lies with the data provider agency. As data journalists, their job is only to continue the publication of data that has been carried out by the agency. They believe that by including the source of data in each news, the responsibility for the validity of the data lies with the agency that issued the data. This finding is in line with previous research that showed how journalists seldom verify the numbers they use, mainly because they see it as outside their role to do such work and because they '*hide behind*' the perceived credibility of their sources (Lawson, 2021). His research suggested that "journalists perceive their role to be limited to the assessment of trustworthy sources rather than the direct interrogation of the number itself" (p.14).

Thus, journalists were found to develop the practice of gathering '*evidence of evidence*' where instead of verifying the data, they prefer to determine which people and institutions they believed they could trust. Our informants mentioned if there is a dead-end when finding different data between agencies, one of the strategies taken by them is explaining to the public that there are differences. Another strategy is to choose one of the data that is considered more valid, generally seen from the history and reputation of the agency that issued the data. Relying on the source credibility which is usually based on the source's '*officialness*' or organizational role is not something new in journalism practice, this is journalists' '*autopilot*' mode that functions more than two-thirds of the time (Barnoy & Reich, 2022). Journalists need to realize that their reliance on the official source means that alternative views are being overlooked. This reflexive source-based trust rooted in official authority leaves them vulnerable to being misled and leaves out valuable alternative or marginalized voices as well. Journalists are becoming more passive, mere processors of one-sided information or bland copy dictated by sources (O'Neill & O'Connor, 2008), when in fact the development of data journalism is a great opportunity for journalists to become more critical and do investigative works.

How Data Turns into News Stories

News is never as raw and factual as desired because the news is not the actual event. News is not a reflection of reality, the work of news is to '*reconstruct*' the actual incident (Santana, 2017). Therefore, the real job of the media is to sell news, not to sell events (Abrar, 2014). This makes the news stories published by the media news that is considered attractive and has a selling value. In a media company, the party that determines whether or not an event is reported is the gatekeeper through the gatekeeping process. The position of gatekeeper in a media is generally occupied by journalists who have the position of the editor or higher, for example, the managing editor, deputy editor in chief, or the editor in chief. This is because the parties with

Volume 06 No. 01 July 2022
https://doi.org/10.24198/jkj.v6i1.37167

these positions, especially the editor-in-chief as the highest leadership, will be the ones held accountable formally and informally for all messages published by a media company (Romli, 2017). Thus, it is important to know the workflow applied to media companies, in this case, the work routines at Lokadata and Katadata. This workflow can show the topic selection process carried out by the media and how the data is processed into news stories by journalists.

In these two media, the topics that are being published as news are generally suggestions from various individuals in the media company which are then approved by the editor. At Katadata, reporters list news ideas every day. Meanwhile, at Lokadata, ideas for news topics often originate from data scientists. But, it is also possible for the editor to determine what issues need to be reported from the very beginning.

> "Every morning, we do a listing where we propose a topic to write about, and then the editor will see if the listing is okay or not, whether it is a hot topic that is being discussed or not. Or perhaps there is another hotter topic, the editor will propose something better. The editor is the gatekeeper, so in case there are topics that are a bit sensitive, which seem like it is better not to talk about it.. well, it is usually the editor who tells you to find another topic. For example, it is too sensitive or does not match the value of Katadata. But from my experience, what I propose is being used more, I mean compared to what the editor proposed." (Informant 3, Interview. 2021, September 29)

Something similar but not entirely the same also happened in Lokadata. At Lokadata, the editorial team and the data team are in different divisions. Informant 10 explained that the topics raised as news stories were an agreement between the two divisions in which the final process was through the approval of the editorial team who has the editorial board and is led by the editor-in-chief. As a data scientist, Informant 10 often has compiled hypotheses on the data he found before discussing it with the team. "After the editorial team gives their approval, then I will process the data to find out if my initial hypothesis is the same as my findings" (Informant 10, Interview. 2021, August 16).

Both Lokadata and Katadata also have the same tendency in choosing news topics that generally revolve around economic and business issues. This tendency lies at the individual level, organizational level, and extra media level (Shoemaker & Reese, 1996). At the individual level, the majority of the informants in this research had educational or occupational backgrounds in economics before working at their current media companies. Meanwhile, at the organizational level, the two media companies where the informant's work are focused on economic and business issues. This seems to be influenced by the commercial clients of both media.

As mentioned in the previous section, media with a data journalism approach has a business model that is no longer dependent on traditional advertising. Lokadata and Katadata have clients who use their services in terms of data processing and data analysis. Generally, these clients are agencies or organizations in the economic and business fields. Therefore, these media also publish news that is not only interesting to the public in general but also that is considered relevant by the client. According to Informant 7, the news published on their website is a showcase that can show the quality of Lokadata's performance and has the potential to invite new clients to work with them. It is not a problem when media operate commercially as they are also a business organization, but ideally, there should be a clear line (firewall) between the business division and the editorial division in a media company. The purpose of this firewall is to build the objectivity of news stories and maintain a journalistic code of ethics (Jamroji & Nasrullah, 2020). At Lokadata, clients are managed by the data team which is different from the editorial team. However, in general, the editorial team relies on the data team for data accessing and processing. The editorial team only needs to compile the news stories from the data provided by the data team. Katadata has a different system where they

have an editorial team dedicated to data journalism. Meanwhile, the media business and clients are managed by the Katadata Insight Center. Both media have separate divisions but often work together in obtaining data, especially when conducting independent research for clients, as we have explained in the previous section about data as news sources.

This might create a blurred line on the firewall, as is seen in other research on online media in Indonesia, where media is practicing market-driven journalism that sees the audience as costumer instead of citizens (Manik, Mulyani, & Kusmayadi, 2021). The blurred firewall could be seen in how advertising contents are polished as news stories. This is something common in traditional media but turns out, it is still happening in online media companies, including Katadata. Even Informant 3 stated that she did not realize it until we pointed it out. "It is rarely discussed in the newsroom. Maybe because it is one of the ways Katadata makes money, so maybe the editorial team is a bit loose on this. As long as the news can still be beneficial for the public interest, the difference is very thin between editorial news and advertising contents" (Informant 3, Interview. 2021, September 29). This, added to the fact that news stories becoming a showcase for window dressing, shows how commodification works in the media organization. Data-driven journalism media even take this commodification further where the audience is not only the product, but data-driven news stories are also the product to attract more clients.

The extra media-level also has an important influence on the selection of issues to be published as news. This relates to the availability and quality of data sources. Lokadata's Chief Data Officer said that data related to the economy generally has better quality. "Ministries and institutions, everything related to economic issues is good because they are required to provide the report, such as the food security index or the economic index every month. That is the good one. But those related to social issues are a big homework. However, from our experience, the government data is indeed getting better day by day" (Informant 6, Interview, August 23, 2021).

Our research shows that there are significant differences between traditional journalism and data journalism. In traditional journalism, the topics published as news are the hot or trending issues of the day. News production is carried out continuously and intensely, so it is not surprising that journalists in online media are now targeted to produce ten to twenty stories per day. This situation does not occur in data journalism media such as Lokadata and Katadata. These two media of course also follow the trends that develop in society, but the news they produce is in-depth news, not only at the surface level of an event. There are three categories of use of statistical data in reporting: (1) The data is only used as a simple reference; (2) Data is presented clearly but with little or limited context; and (3) The data is presented by providing context to the data (Cushion et al., 2017). The ideal condition is of course the third one, but the production of in-depth news hugely relies on the adequacy of data sources.

The availability of data is what makes Lokadata and Katadata able to raise an issue by offering different perspectives. With data, the context of an issue can be explained better. Informant 6 gave an example of the news regarding the imported products of basic ingredients such as onions, chilies, and salt. News stories about imported products often create negative sentiment in the public towards the government because the public thinks that we have local farmers who produce these products, so why import them? He explained that by using the data, it can be seen that although our farmers produce a lot of onions, the irrigation technology used is still conventional with a limited harvest period each year. Another example is the issue of imported salt, even though Indonesia has a long coastline, the salt used in the kitchen does not come from there because salt from mines costs less to produce. Informant 6 regrets that some

data journalists cannot contextualize news like this and only focus on visualizing interesting data using infographics.

All in all, our research findings indicate that the gatekeeping process in a data journalism-based media company is different from traditional journalism, as their process takes a longer time. However, the parties involved are generally the same. The gatekeeper role is still carried out by individuals with the position of editor. In addition, what distinguishes it from traditional journalism is the need for additional technical skills, especially in using data processing applications. Based on the interviews with all informants of this research, generally, they have this ability by self-taught or by forming informal training activities regularly with their co-workers. This might be sufficient for now, but media organizations should consider giving formal education about data processing as this is a skill that not everyone can master, at least not right away. Especially with the fact that the level of experience is correlated to a positive attitude toward data journalism as well as to working with numbers and statistics (Appelgren & Nygren, 2014), which leads to a better quality of data-driven news stories produced by their journalists.

CONCLUSION

In general, journalists mostly use the data provided by the government to then be processed into news stories. Several informants stated that they were more inclined to government agency data than private agency data due to the larger scope of research by government agencies. The main problem with government agency data, according to data journalists, is the lack of data integration between agencies. If it is difficult to access the data of a government agency, another effort is to carry out commercial cooperation, which means media companies are paying to obtain data. Sometimes media companies can also do research independently, but considering the time, effort, and money, media companies rarely do it. These data journalism media generally take advantage of the results of research carried out by other divisions in their company who work on data projects for clients.

The effort on data verification carried out by data journalists is by checking the raw data if it is available. If they find discrepancies in the data, the journalists confirm it to the relevant agency that published the data. However, there is a tendency of data journalists to be passively relying on official sources and '*hide behind*' the perceived credibility of their sources. The gatekeeping process in data journalism media is also not much different from traditional journalism media. The gatekeeper position still lies in the individual with the position of editor. But, the news stories produced by data journalism media are generally more in-depth with a longer processing time. It will be interesting if future research studies other media that also apply data journalism. In particular, data journalism implementation in legacy media, because media originating from conventional media may have a different culture and routine compared to digital native media.

REFERENCES

Abrar, A. (2014). *Mengarungi hubungan DPR dan pemerintah dengan jurnalisme*. Yogyakarta: Gadjah Mada University Press.

Appelgren, E., & Nygren, G. (2014). Data journalism in Sweden. *Digital Journalism*, *2*(3), 394–405. https://doi.org/10.1080/21670811.2014.884344

Asprilla, A., & Maharani, N. (2019). Jurnalisme data dalam digitalisasi jurnalisme investigasi Tempo. *Jurnal Kajian Jurnalisme*, *2*(2), 212–224. https://doi.org/10.24198/jkj.v2i2.21362

Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates

notions of democracy, participation, and journalism. *Big Data & Society*, *2*(2), 1–11. https://doi.org/10.1177/2053951715594634

Barnoy, A., & Reich, Z. (2022). Trusting others: A pareto distribution of source and message credibility among news reporters. *Communication Research*, *49*(2), 196–220. https://doi.org/10.1177/0093650220911814

Coddington, M. (2015). Clarifying journalism's quantitative turn. *Digital Journalism*, *3*(3), 331–348. https://doi.org/10.1080/21670811.2014.976400

Cushion, S., Lewis, J., & Callaghan, R. (2017). Data journalism, impartiality and statistical claims. *Journalism Practice*, *11*(10), 1198–1215. https://doi.org/10.1080/17512786.2016.1256789

Danayanti, E., Wardhana, B., Marsiela, A., & Galuh, F. (2021). *Menilai penerapan jurnalisme data dan investigasi berbasis data di Indonesia*. Jakarta. Retrieved from https://aji.or.id/upload/article_doc/Laporan_Menilai_Penerapan_Jurnalisme_Data_dan_Investigasi_(8).pdf

De-Lima-Santos, M.-F., & Mesquita, L. (2021a). In a search for sustainability: Digitalization and its influence on business models in Latin America. In M. F. Salaverría, Ramón & de-Lima-Santos (Ed.), *Journalism, Data and Technology in Latin America* (pp. 55–96). New York: Springer. https://doi.org/10.1007/978-3-030-65860-1_3

De-Lima-Santos, M.-F., & Mesquita, L. (2021b). The strategic value of data journalism. In M.-F. Salaverría, Ramón & de-Lima-Santos (Ed.), *Journalism, Data and Technology in Latin America* (pp. 97–136). New York: Springer. https://doi.org/10.1007/978-3-030-65860-1_4

De-Lima-Santos, M.-F., Schapals, A. K., & Bruns, A. (2021). Out-of-the-box versus in-house tools: How are they affecting data journalism in Australia? *Media International Australia*, *181*(1), 152–166. https://doi.org/10.1177/1329878X20961569

Felle, T. (2016). Digital watchdogs? Data reporting and the news media's traditional 'fourth estate' function. *Journalism*, *17*(1), 85–96. https://doi.org/10.1177/1464884915593246

Gray, J., & Lämmerhirt, D. (2019). Making data public? The open data index as participatory device. In *Theory on Demand: Good Data*. Amsterdam: Institute of Network Cultures.

Imawan, K. (2018). Jurnalisme data resistensi ruang siberdi era post truth. *Jurnal Signal*, *6*(2), 1–18. Retrieved from https://jurnal.ugj.ac.id/index.php/Signal/article/view/1316

Islami, M. (2021). Implementasi satu data Indonesia: Tantangan dan Critical Success Factors (CSFs). *Jurnal Komunika*, *10*(1), 13–23. https://doi.org/10.31504/komunika.v10i1.3750

Jamroji, & Nasrullah. (2020). Breaking down "firewall" among public relations and journalism. *Jurnal Interaksi : Jurnal Ilmu Komunikasi*, *4*(1), 86–97. https://doi.org/10.30596%2Finteraksi.v4i1.3957

Knight, M. (2015). Data journalism in the UK: A preliminary analysis of form and content. *Journal of Media Practice*, *16*(1), 55–72. https://doi.org/10.1080/14682753.2015.1015801

Lawson, B. T. (2021). Hiding behind databases, institutions and actors: How journalists use statistics in reporting humanitarian crises. *Journalism Practice*, 1–21. https://doi.org/10.1080/17512786.2021.1930106

Manik, C., Mulyani, H. S., & Kusmayadi, I. M. (2021). Penerapan pagar api pada Female Daily Network. *Jurnal Kajian Jurnalisme*, *4*(2), 177–191. https://doi.org/10.24198/jkj.v4i2.27156

O'Neill, D., & O'Connor, C. (2008). The passive journalist. *Journalism Practice*, *2*(3), 487–500. https://doi.org/10.1080/17512780802281248

Parasie, S., & Dagiral, E. (2013). Data-driven journalism and the public good: "Computer-assisted-reporters" and "programmer-journalists" in Chicago. *New Media & Society*, *15*(6), 853–871. https://doi.org/10.1177/1461444812463345

Porlezza, C., & Splendore, S. (2019). From open journalism to closed data: Data journalism in Italy. *Digital Journalism*, *7*(9), 1230–1252. https://doi.org/10.1080/21670811.2019.1657778

Romli, K. (2017). *Komunikasi massa*. Jakarta: Grasindo.

Salaverría, R. (2020). Exploring digital native news media. *Media and Communication*, *8*(2), 1–4. https://doi.org/10.17645/mac.v8i2.3044

Santana, K. (2017). *Jurnalisme kontemporer* (2nd ed.). Jakarta: Yayasan Pustaka Obor Indonesia.

Sanusi, H. (2018). Jurnalisme data: Transformasi dan tantangan era digital. *Jurnal Dakwah Tabligh*, *19*(1), 20–43. https://doi.org/10.24252/jdt.v19i1.5909

Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 1139–1148. https://doi.org/10.1109/TVCG.2010.179

Shoemaker, P. J., & Reese, S. D. (1996). *Mediating the message: Theories of influences on mass media content*. London: Longman. https://doi.org/10.1177/1326365X14540245

Stagars, M. (2016). *Open data in Southeast Asia: Towards economic prosperity, government transparency, and citizen participation in the ASEAN*. London: Palgrave Macmillan.

Stalph, F., & Borges-Rey, E. (2018). Data journalism sustainability. *Digital Journalism*, *6*(8), 1078–1089. https://doi.org/10.1080/21670811.2018.1503060

Stoneman, J. (2015). *Does open data need journalism?* Oxford. Retrieved from https://reutersinstitute.politics.ox.ac.uk/our-research/does-open-data-need-journalism

Stuart, D. (2011). *Facilitating access to the web of data: A guide for librarians*. London: Facet Publishing.

Vara-Miguel, A. (2020). Cross-National similarities and differences between legacy and digital-born news media audiences. *Media and Communication*, *8*(2), 16–27. https://doi.org/10.17645/mac.v8i2.2733

Widiantara, I. K. A. (2021). Tren dan fenomena jurnalisme data pada media online di Indonesia. *Danapati: Jurnal Ilmu Komunikasi*, *1*(2), 118–125. Retrieved from https://jurnal.ekadanta.org/index.php/danapati/article/view/134

Zamili, M. (2015). Menghindar dari bias: Praktik triangulasi dan kesahihan riset kualitatif. *LISAN AL-HAL: Jurnal Pengembangan Pemikiran Dan Kebudayaan*, *9*(2), 283–304. https://doi.org/10.35316/lisanalhal.v9i2.97