

[DOI: 10.20472/AHC.2019.005.009](https://doi.org/10.20472/AHC.2019.005.009)

**MONWIPA WONGRUJIRA**

Sukhothai Thammathirat Open University, Thailand

## **EXPLORING THE USE OF DATA-DRIVEN JOURNALISM IN THAI MASS MEDIA**

### **Abstract:**

Technology and mobile devices allow many news consumers become news senders—prosumer (i.e. being both consumers and producers of news and information at the same time). Anyone could be a reporter. Also, there are tons of news and information flow around us every day. The differences between media reporting stories and information running around social media are the quality of news and information. If the media do only report ‘who what when where why how,’ they did not accomplish their task as a journalist. Data-driven journalism becomes significant in news reporting process. It needs not merely Big Data, but also analysis process and presentation. This paper intends to explore the use of data journalism among the mass media in Thailand. Whereas social media become more and more popular and drawing attention among Thai news consumers, the professional media need to differentiate their news reporting to focus on in-depth or investigative reporting. How the professional media apply data-driven journalism; to what extent did they use data for reporting a story; and what are the obstacles affecting their application of data journalism. Factors affecting the use of data-driven journalism included: data sources (incomplete, unstructured, and difficult to access), data compiling, time consuming and limitation of technology for data analysis and presentation.

### **Keywords:**

Data journalism, professionalism, news reporting, mass media, social media, Thailand

## 1 Introduction

Technology has changed our daily life drastically. Digital economics and digital society are approaching. These lead to the change of communication behaviors. Consumers need conveniences, speed and mobility. Internet and social media allow consumers create and distribute news and information by themselves. Public rely less on traditional media (i.e. newspaper and broadcasting) than internet and social media. They are also becoming “prosumers” or consumers who also generate and publish content similar to producers. These led to vast variety of news and information around social media. However, quantity does not led to quality. Increasing vast amount of data leads to not only overloaded news and information, but also fake news and misinformation.

Roles of professional media do not merely mirroring what is going on in the society, but also monitoring and educating the society. Technology nowadays assists the media to gather and analyze the data. Actually, the media has been using data for reporting news for a long time, but with a limitation of data and analysis tools. When online media become popular among users and the media, both structured and unstructured data become “Big data,” which allow the media to analysis and visualize them in their report.

The aim of this article is to explore the use of data-driven journalism by Thai press, both traditional and online media. The following topic included explanation of data-driven journalism, its application in Thai mass media, and some factors obstructed the use of data-driven journalism.

## 2 Data-driven journalism

Data-driven journalism or data journalism is defined as a process of news reporting that mainly focuses on information database. The process starts from information search and data collection for analysis. Next, journalist can re-organize, categorize and revise information, to make sure that receivers easily understand the information reported by journalist. In fact, Information Technology and Communication systems are used for Data Journalism processing to increase the effectiveness of news reporting. Further benefit is information can be analyzed in sights with more detailed, which probably have not been unveiled before.

What makes data journalism different to the rest of journalism? Perhaps it is the new possibilities that open up when you combine the traditional ‘nose for news’ and ability to tell a compelling story, with the sheer scale and range of digital information now available. Data journalism can help a journalist tell a complex story through engaging infographics (Bradshaw, 2012)

News stories are flowing in as they happen, from multiple sources, eye-witnesses, blogs and what has happened is filtered through a vast network of social connections, being ranked, commented. This is why data journalism is so important. Gathering, filtering and visualizing what is happening beyond what the eye can see has a growing value. Becoming knowledgeable in searching, cleaning, and visualizing data is transformative for the profession of information gathering, too. Journalists who master this will experience that building articles on facts and insights is a relief. (Lorenz, 2012)

When information was scarce, most of our efforts were devoted to hunting and gathering. Now that information is abundant, processing is more important. We process at two levels: (1) analysis

to bring sense and structure out of the never-ending flow of data and (2) presentation to get what's important and relevant into the consumer's head. Like science, data journalism discloses its methods and presents its findings in a way that can be verified by replication (Meyer, 2012).

The first example of data journalism at the Guardian dates from 1821. It is a leaked table of schools in Manchester listing the number of students who attended it and the costs per school. According to Rogers this helped to show for the first time the real number of students receiving free education, which was much higher than what official numbers showed (Roger, 2017). The way that media organizations like Guardian and the New York Times handled the large amounts of data released by Wikileaks is one of the major steps that brought the term into prominence. Around that time the term started to enter into more widespread usage, alongside 'computer-assisted reporting', to describe how journalists were using data to improve their coverage and to augment in-depth investigations into a given topic (Bounegru, 2012).

The latest case study of using data journalism was the Panama Papers. It led ICIJ (International Consortium of Investigative Journalists) to win both Pulitzer price award in 2017 and Data Journalism award in 2018. The documents show the myriad ways in which the rich can exploit secretive offshore tax regimes. Twelve national leaders are among 143 politicians, their families and close associates from around the world known to have been using offshore tax havens (Harding, 2016).

But none of those stories would have appeared without a lot of work preparing the data. The Panama Papers comprised some 2.6TB of data and 11.5 million documents. That document data is called unstructured since it does not come in the neat rows-and-columns of traditional databases. ICIJ the ICIJ tech staff had to take the leaked information and rebuild the original SQL database structure it came from. Technology allowed them to analyze the huge data set and became series of news stories with collaboration of 370 journalists from 78 countries (Darrow, 2017).

Paul Bradshaw (2011) explained the 5C stages of data journalism: comply, clean, context, combine and communicate.

**Comply:** Data journalism begins in one of two ways: either you have a question that needs data, or a dataset that needs questioning. Whichever it is, the compilation of data is what defines it as an act of data journalism. Compiling data can take various forms, such as data from organizations (both government and private sectors), databases hidden behind online forms, or by collecting the data yourself through observation, surveys, online forms or crowd sourcing.

**Clean:** Cleaning typically takes two forms: removing human error; and converting the data into a format that is consistent with other data you are using. For example, datasets will often include some or all of the following: duplicate entries; empty entries; the use of default values to save time or where no information was held; incorrect formatting (e.g. words instead of numbers); corrupted entries or entries with HTML code; multiple names for the same thing (e.g. BBC and B.B.C. and British Broadcasting Corporation); and missing data. There are simple ways to clean up data in Excel or Google Docs such as find and replace, sorting to find unusually high, low, or empty entries, and using filters so that only duplicate entries (i.e. those where a piece of data occurs more than once) are shown.

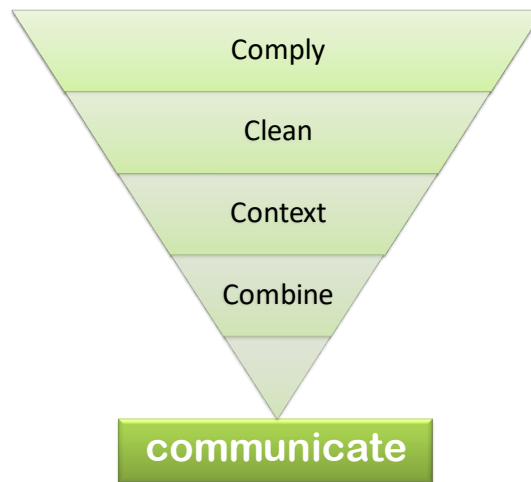


Figure 1: The inverted pyramid of data journalism (Bradshaw, 2011)

**Context:** Like any source, data cannot always be trusted. It comes with its own histories, biases, and objectives. So like any source, journalists need to ask questions of it: who gathered it, when, and for what purpose? How was it gathered? (Methodology). What exactly do they mean by that? Journalists also need to understand jargon, such as codes that represent categories, classifications or locations, and specialist terminology. All the above will most likely lead to compile further data. For example, knowing the number of crimes in a city is interesting, but only becomes meaningful when a journalist contextualizes that alongside the population, or the numbers of police, or the levels of crime 5 years ago, or perceptions of crime, or levels of unemployment, and so on. At this stage, statistical literacy is needed; or journalists might request help from data scientists.

**Combine:** Good stories can be found in a single dataset, but often a news report will need to combine two together. A multiple-source story is much preferred than a single-source one. A more mundane combination is to combine two or more datasets with a common data point. That might be a politician's name, for example, or a school, or a location. This often means ensuring that the particular data point is formatted in the same name across each dataset.

**Communicate:** In data journalism the all-too-obvious thing to do at this point is to visualize the results – on a map, in a chart, an infographic, or an animation. But there's a lot more here to consider – from the classic narrative, to news apps, case studies and personalization.

### 3 Data-driven journalism in Thai mass media

Whereas social media become more and more popular and drawing attention among Thai news consumers, the professional media need to differentiate their news reporting to focus on in-depth or investigative reporting. How the professional media apply data-driven journalism; to what extent did they use data for reporting a story; and what are the obstacles affecting their application of data journalism.

There were three levels of usage of data-driven journalism. The first level is a simple reveal of data or information from the source, e.g. statistics or survey report by government organizations or private sectors. These kinds of coverage need the least technique of data analysis. For

example, the following data derived from Stock Exchange of Thailand revealing the publishing business in 2017 comparing to previous years.

**GLOOMY PROSPECTS** Unit: million baht

	2013	'14	'15	'16	'17 (Q2)
<b>Nation Multimedia Group</b>					
Revenue	3,060.67	2,951.23	3,165.36	2,183.14	975.80
Net profit	251.72	64.46	35.58	-1,102.85	-519.13
<b>Bangkok Post</b>					
Revenue	2,474.94	2,281.25	2,211.59	1,865.15	665.30
Net profit	128.89	-168.11	-251.76	-215.33	-119.64
<b>Matichon</b>					
Revenue	1,608.89	1,226.35	1,076.48	983.14	468.89
Net profit	109.21	-50.85	-103.34	-63.67	7.12
<b>Amarin</b>					
Revenue	2,088.63	1,913.97	2,003.86	1,945.01	940.26
Net profit	285.08	-85.18	-417.15	-628.12	-170.81
<b>Siam Sport Syndicate</b>					
Revenue	2,716.01	2,647.17	2,076.31	1,104.06	487.71
Net profit	12.22	-133.86	-33.89	-358.04	-129.68

Source: Stock Exchange of Thailand

BANGKOK POST GRAPHICS

Figure 2: Revenue of the publishing business from 2013 to 2017

Source: <https://www.bangkokpost.com/business/news/1311347/still-life-left-in-struggling-newspaper-giants>

The second level of data-driven journalism need more afford and data analytical skill, i.e. comparison of data across time or items in a news story. For example, comparison of elected PM between different political parties from the previous elections, differences of tourism numbers after releasing new visa exemption. For example, the following infographic of Prachatai, an online news agency, comply the political parties' policy targeting the elderly people.

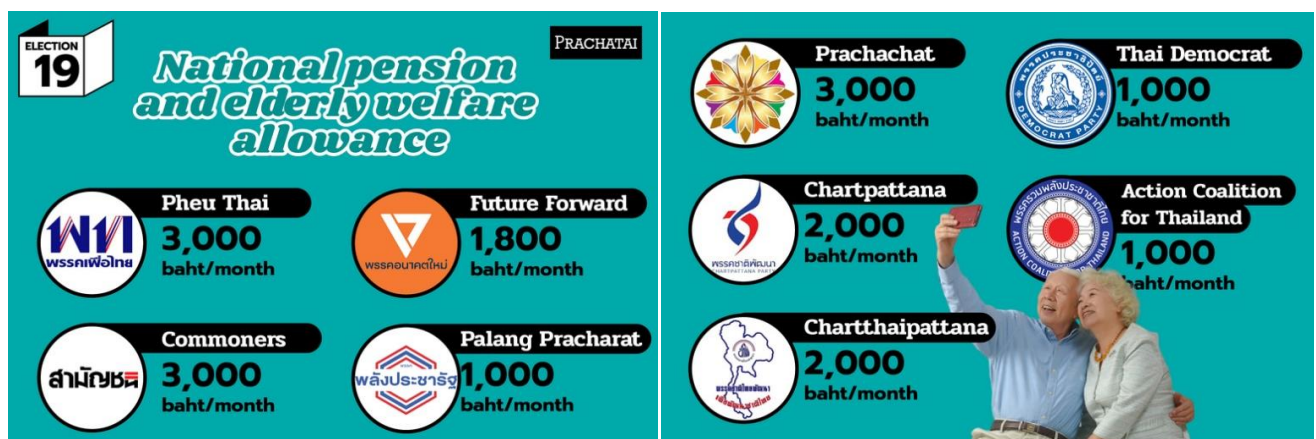


Figure 1: Differences of political parties' policy on elderly

Source: <https://prachatai.com/english/node/7980>

The third level is fully use of data journalism through the process of data gathering, cleaning, filtering, analyzing and visualizing data. The third level involves data scientists, computer programmers and visual designers. A well-known case was Panama Papers which brought International Consortium of Investigative Journalists (ICIJ) to win the Pulitzer Prize in 2017 (Harding, 2016). Data usage among Thai mass media emphasized on the first and second level of data journalism, whereas the third level of data-driven journalism is not commonly used, mostly in online news media rather than traditional one.

ThaiPublica, an online news agency, is an example of applying data-driven journalism. An investigative report was about the lottery quota distributed by the government lottery office. There had been unconfirmed information about distribution quota of lottery to some distributors, which could lead to corruption and non-transparency. ThaiPublica officially requested the data from the government lottery office according to the Official Information Act. B.E.2540. The data provided by the office were in pdf file (i.e. scanned documents). The news agency had to clean that unstructured data, by creating and manually reentering all the data into Excel files. Still, there were missing and incomplete data. The process of 5Cs (comply, clean, context, combine and communicate) had took more than six months. To visualize the data, ThaiPublica received support from Boomee Lab, a start-up company, to design and present the stories into visual; such as infographic, chart, and interactive game and animation. The lottery coverage by ThaiPublica caused the government lottery office to reconsider and revise all the quota and lottery distribution process. The following example was an interactive game for users to try buying a lottery on the game and the result would show where the money will go from buyers' pocket to the lottery office and other stakeholders.

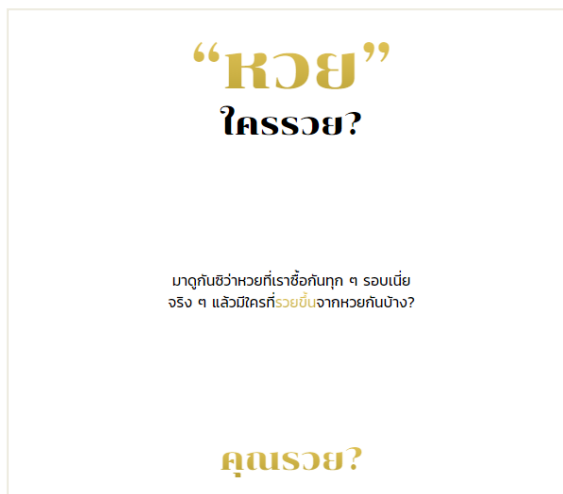




Figure 3: Lottery game related to the news coverage of lottery quota

Source: <https://data.boonmeelab.com/lottery/>

#### 4 Factors affecting the use of data-driven journalism

As we can see, data journalism needed human and technology resources and time consuming. There were some obstacles of applying data-driven journalism among Thai journalists.

1. Data Sources: Although data-driven journalism needs both quantity and quality of data, most open data sources were incomplete and unstructured, especially government data source. Even the Thai government has policy about e-government for many years, still the official open data sources are not completed and have lots of missing data. Many of them are still in paper forms, not online forms. Some of them are created with different format leading to difficulty of understanding and analysis. Some state agencies had believed of government data needed to keep secured and secret, so revealing them might cause problems later.

2. Data Compiling: Because of the previous factors, data complying stage, then, needed more manual work. After data complying, instead of processing the data thru cleaning and analyzing stage, the news staff had to manually clean and structure the data before using machine analysis. Another factor was accessibility to data sources. Sometimes it needed personal relationship.

3. Time: As mentioned before, the process of accessing and complying data were time consuming. Therefore, using data-driven journalism in a news report needed time. While social media allow speed and timeliness, professional media should concern the accuracy. Data-driven journalism reports still not frequently apply. But its application assists an increasing credibility and accuracy of the news coverage.

4. Technology for data analysis: To apply data-driven journalism at the third level, news room needed experts from both data science and data visualization. Many newsroom staff still lack of these expertise. Technology for data analysis was another investment factor, which managerial level started to concern. Therefore, not only improve the computer hardware and

software, but improve the skill of people ware (e.g. news staff, reporters, rewriters) in data analysis and reading data is a must.

## 5 Conclusion and suggestion

Data-driven journalism has applied among Thai mass media to what extent. It has been used in simple style; i.e. reporting data directly revealing from the sources (such as research report, statistics report). It has also been used in a more complex style; complying variety of data from different sources, cleaning the data and analyze them, and finally visual presentation. The complex use of data journalism need team work, data analytical skill, and time consuming. Thus, complicated data-driven journalism in Thai mass media is still not widely applied. However, to enhance the media credibility and the accuracy of the news coverage, data-driven journalism would be a vital option. In the middle of the information ocean, news consumers still need expertise to guide them.

## 6 References

- Bounegru, Liliana (2012). Data journalism in perspective. In *Data Journalism Handbook*. Retrieved from <https://datajournalismhandbook.org/handbook/one/introduction/data-journalism-in-perspective>
- Bradshaw, Paul. (2011). The inverted pyramid of data journalism. Retrieved from <https://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/>
- Bradshaw, Paul. (2012). What is data journalism? In *Data Journalism Handbook*. Retrieved from <https://datajournalismhandbook.org/handbook/one/introduction/what-is-data-journalism>
- Darrow, Barb. (2017). How Tech Made the Pulitzer Prize-Winning Panama Papers Coverage Possible. *Fortune*. Retrieved from <http://fortune.com/2017/05/30/panama-papers-data-tools/>
- Gray, J., Bounegru, L., & Chambers, L. (eds.) (2012). *The Data Journalism Handbook: How Journalists Can Use Data to Improve the News*. Retrieved from <https://datajournalismhandbook.org/handbook/one>
- Harding, Luke. (2016). *What are the Panama Papers? A guide to history's biggest data leak*. Retrieved from <http://www.theguardian.com/news/2016/apr/03/what-you-need-to-know-about-the-panama-papers>
- Lorenz, Mirko. (2012). Why journalists should use data. In *Data Journalism Handbook*. Retrieved from <https://datajournalismhandbook.org/handbook/one/introduction/why-journalists-should-use-data>
- Meyer, Philip (2012). Why is data journalism important? In *Data Journalism Handbook*. Retrieved from <https://datajournalismhandbook.org/handbook/one/introduction/why-is-data-journalism-important>
- Rogers, Simon. (2017). *The state of data journalism in 2017*. Retrieved from <https://www.blog.google/outreach-initiatives/google-news-initiative/data-journalism-2017/>
- Suchiva, Nanat. (2017, August 23). Still life left in struggling newspaper giants. *Bangkok Post*. Retrieved from <https://www.bangkokpost.com/business/news/1311347/still-life-left-in-struggling-newspaper-giants>.
- Welfare state policies in the 2019 general election. (2019). *Prachatai online*. Retrieved from <https://prachatai.com/english/node/7980>

## Website



<https://data.boonmeelab.com/lottery/>

<http://thaipublica.org/investigations/lottery/>