

Mapping/Locating Data Journalism Research in India

Communication & Journalism Research

10 (1) pp 69-80

©The Author (s) 2021

Reprints and Permissions:

masscomhod@uoc.ac.in

ISSN 2348 – 5663

Sikha N*

Research Scholar, Journalism & Mass Communication, University of Calicut, Kerala, India

Muhammadali Nelliullathil

Head, Journalism & Mass Communication, University of Calicut, Kerala, India

Abstract

Data Journalism has gradually appeared over the last decade, driven by the availability of enormous data in digital form. Currently, a considerable amount of data journalism projects are being produced all over the world. In recent years data journalism (DJ) has drawn significant attention both in academia and in media and as a result, this has generated a sizeable body of research literature focusing on various communication premises and approaches as well as social contexts. India is not an exemption to this trend. This article explores the existing research trends in data journalism in India. To better understand its current state in the country, the researchers observed and reviewed the published research literature and works that studied data journalism and related practices empirically. Analyzing these studies allows to map the development of the literature and identify gaps in data journalism research in the country. This article offers some suggestions for future research in the field of data-driven journalism.

Keywords

Data Journalism, Data Visualization, Algorithmic Journalism, Data Journalism Education, Data Ethics and Privacy, Data-driven Stories.

Introduction

With the rise of ICTs, our modalities for living and communication, settling on choices and understanding the truth are being changed. In like manner, the ordinary rationale of news coverage has also impressively been changed by inculcating computational speculation into the newsrooms. Having no ideal definition for what comprises data in journalism, tied with various labels (such as data journalism (DJ), data-driven journalism, database journalism, algorithmic and automated journalism, computational journalism, computer-assisted reporting (CAR), data visualization etc.) for data-related journalistic practices, understanding data's place inside journalism is challenging. Simon Rogers who built up the *Guardian's* Data Blog in 2009, contends that the newspaper has been doing data journalism since 1821 when its first edition conveyed a table of information about Manchester Schools (Rogers, 2013).

* Correspondence: Email: sikhaarts@gmail.com

The data journalism is a fairly new journalism wonder that started to appear in industry publications after 2006 however the idea has not been investigated scholastically that much. The practice of making data journalism has been vague even in recent times. One of the first vivid descriptions was portrayed in *Data Journalism Handbook* (Gray, Bounegru & Chambers, 2012), which was published as a collective effort where practitioners and researchers across the world contributed perspectives in the form of a workshop. It was a novel step and a sort of brainstorming effort, most suitable to identify the areas of focus in an emerging field with a serious note. Data journalism literature has concentrated on issues like the conflicts in Afghanistan and Iraq, Olympics, horse-meat trafficking, immigration, tobacco smuggling, crime in big cities, pollution in certain areas, and traffic jams (Bounegru et al 2012)

It is possible to date the beginning of computer-assisted reporting and data journalism to 1952 when Philip Meyer at *The Detroit Free Press* used a supercomputer to take a gander at an assessment of Detroit residents for the purpose of understanding and clearing up the serious riots that detonated in the city. The early years of the 21st century saw the Global Investigative Journalism Network begin to have a critical influence in the development, beginning with its first meeting in 2001 in Copenhagen that offered a strong computer-assisted reporting track and hands-on training. Considering all the elements included in data journalism, from computer-assisted reporting to data visualization, there is a final definition accepted by Veglis and Bratsas (2017): “Data journalism can be defined as the process of extracting useful information from data, writing articles based on the information and embedding visualizations (interactive in some cases) in the articles that help the readers understand the significance of the story or allow them to pinpoint data that relate to them.”

By the twentieth century, there was a perceptible move in the manner in which publishers utilize data. Rather than simply plotting trends, individuals also collected data to make predictions. For example, in 1952, Navy mathematician Grace Murray Hopper and a team of programmers used voting statistics from earlier elections to inside one percentage point. This enthusiasm to anticipate elections was carried into the 21st century by *Five Thirty- Eight* founder Nate Silver, who precisely anticipated 49 out of 50 states during Obama's first presidential bid. Toward the beginning of the 21st century, an upheaval in mobile computing, increases in online connectivity, access, and speed; and explosion in data creation in a general sense changed the landscape for computer-assisted reporting.

As opposed to the European and American instances, the influxes of data journalism practice has developed from numerous Asian nations. Afghanistan's biggest free news Pajhwok Afghan News is the nation's sole news source to utilize data to inform the public. It uses simple data visualization with a strong narrative to appeal to an audience with low data and digital literacy and is regularly referred to by advocates for instance of how data journalism should be possible significantly under incredibly troublesome conditions.

Data journalism is in a growing stage in India. There are as of now not many sites that are operating in this area, and news organizations still have far to go to arrive at the global benchmark. Fortunately, it has gotten and grabbed the eye and caught the attention of the Indian government, news organizations, analytics companies, and data visualization companies. DJ can be accustomed to realize about accountability and transparency among people, particularly in a democratic system. Be that as it may, this refinement about DJ is gradually drawing in open consideration as it requires different literacy levels. Then again, DJ is quick turning into a significant piece of Indian newsrooms as the industry internalizes global trends in mass communication particularly those in multimedia and digital reporting.

Research Background

Being an emerging area of study there are chances of uncertainty in clarifying terms and concepts in the data journalism discourses. Some scholars seek to clarify the meaning of these terms about to somewhat different practices of using data and computers in journalism while others position the entire conceptualization either in the domain of journalism or digital studies. However, the inter-disciplinarily approach seems to be more suitable to have a better understanding. Supporting this argument, Philip Hammond (2015) in his work *From Computer-Assisted to Data-Driven: Journalism and Big Data* argues that while journalism is for sure changing, digital data and computer technology are less fundamental to understanding this than is frequently expected. If one wanted to sum up this change in terms of a shift in terminology, then the contrast between journalism which is 'computer-assisted' and that which is 'data-driven' would be a good indicator: computers and data, rather than the human subject, have come to be understood as the active party in the relationship. Rather than digital data technologies simply causing changes in reporting, nonetheless, it is more that both Big Data and (data) journalism are increasingly understood in ways that are consonant with broader shifts in way we consider the human subject and his/her capacity to think about and follow up on the world as an item. These more extensive changes are additionally obvious in what is by all accounts a

developing agreement in the academic commitment with data journalism (Hammond, 2015).

Megan Knight in a preliminary survey of data-based stories presented in the national news in the UK lays the basis for examination and typology of the structures and organizations of data journalism as a media practice. The analysis shows that superficial data journalism is being rehearsed by UK national papers and is constrained in scope and format. No proof was found of a promise to data projects among the news outlets analysed, and just one occasion of plan of action to the Freedom of Information Act was seen. Most data presented were superficial, and sourced from conventional outlets. Data journalism is practiced as much for its visual intrigue with respect to its insightful characteristics, and the general effect, particularly in the newspaper position is as a lot of brightening as informative. This contention helps us to presume that epistemological lucidity essentially relies on the goal and nature of the act of DJ in a particular setting. (Knight, 2015)

Using semi-structured interviews with Australian data journalists, Scott Wright (2018) analyses how journalists view their role in news organizations; the structure of Australian data journalism inside and beyond Australian newsrooms; and how their practice has changed and identifies the elements that drive this change. This study finds that there has been a decrease in bigger tasks and in exploratory dashboards, just as in the active data journalists in Australia. Be that as it may, there stays a small core of data journalists close by certain moves to standardize data into everyday reporting. This proposes instead of purely a story of decline, how data journalism is institutionalised has begun to develop to adapt with these changes indicating its mainstream trends. (Wright, 2018)

Through Scenario Network Mapping, Florian Stalph and Eddy Borges outline the possible futures of data journalism practice by deciding its supportability in the present atmosphere of news coverage severity. Results propose three potential situations; Skill set Perspective, which recommends that data journalism will soon be regarded as fundamental for each journalist. Secondly, as a genre, data journalism will remain a niche storytelling format but will ultimately find its way into smaller newsrooms due to decreasing limitations. Lastly, because of money related and staff constraints, data journalism will be relinquished by the prevailing mainstream media, who will outsource data analysis. Inside this specific circumstance, it is not yet clear whether data journalism can keep enhancing to stay focused in the continually advancing environment of the present news creation. (Florian Stalph & Eddy Borges Ray, 2018)

Research Objectives

Against the background of the researches discussed above, this study focuses on mapping out the major research works conducted on data journalism practices and related aspects in India. This article attempts to map the research works on data journalism conducted in India. Therefore, this study pursues the following research objectives

1. To analyse the nature and framework of data journalism studies conducted in India
2. To trace out the research gaps identified and to provide propositions for future research on data-intensive journalism in India

Each of these objectives is materialized in the analysis section of this article

Method

In India, various books and reports are in circulation to provide practical knowledge and processes regarding data journalism, but very few researches cover the right aspects of the topic. Being a developing and innovative stream, genuine examinations ought to be led to propose different possibilities of data journalism in the country.

This article maps out trends in data journalism research conducted so far in India. The researcher selected fourteen papers published in research journals and were analyzed to identify and describe themes in data journalism research covered by these publications over the years. A systematic review of the selected articles has been carried out to find out the nature and framework of the selected studies. Analysis of titles, research objectives and methodologies applied allows the researcher to classify the selected papers under the following heads: Digitization and Convergence, Data Visualisation, Data Ethics, and Privacy, Big data sets and Indian politics, Data Journalism Initiatives & Platforms, Newsroom challenges with Data and Data Journalism Education. The research gaps in the field were also identified.

Digitization and Convergence

Data journalism in India has been driven and catalyzed by digitization initiatives by different platforms and the growing popularity of social media. Convergence is more than merely technological change rather it has distinctive cultural dimensions. The shift caused by convergence shapes the

“relationship between existing technologies, industries, markets, genres and audiences” (Jenkins, 2004).

The research article *Convergence of technologies and journalists: Translation of journalistic practices through ANT perspective* investigates how the convergent media and practices modify the journalistic system of news and production. Further, this research paper centres on how certain actors in the newsroom become an irreplaceable piece of the news generation procedure. It is evident from the investigation that news technologies are at work in and around the newsroom. The attendant practices of these technologies indicate a great deal of *translation* or *alteration* of the journalistic network in terms of how journalists do their work and how the resources of the organisation and the knowledge produced are coordinated. Since this paper is focused on exploring only the *translation* of the journalistic network, further focused research is expected to enquire how these new advancements are themselves getting adjusted or deciphered. (Kumar & Haneef, 2016)

Anoop Kumar & M. Shuaib Mohamed Haneef in another article *Digital Media Habitus, Agency and Structure: Convergence Practices among Journalists in Indian Newspapers* opines that intermingling procedure has corrected the 'rules of access' and 'rules of the game' for journalists. The article analysed assimilation of convergence practices into the media field by journalists, changes in journalistic values, changes in doxa of the field and agent's habitus through the lens of Pierre Bourdieu's field theory. The thematic analysis of the data reveals that journalists use appropriate new technologies to redefine the habitus and adapt to contemporary dominant work practices. (Kumar & Haneef, 2016).

Data Visualisation

As data become increasingly ubiquitous (Kitchin, 2014), so do data visualizations — that is, the visual representation of data and datasets which communicate precise information and values. Some DJ studies conducted in India focus on data visualisation mainly social media analytics and infographics. Social Media Analytics is all about collecting data produced from Social Media platforms like Facebook, Twitter, LinkedIn, WhatsApp, Wikipedia, YouTube, Pinterest, Instagram, Tumblr, Snapchat, Google+, WeChat, and many others.

Shweta Patnaik & Swati Sucharita Barik in their research article *Social Media Analytics using Visualization* discusses the work of visualization and analysis methods in data journalism. This paper gives an outline of various

phases in social media analytics and spotlights on different open-source tools useful for pre-processing, cleaning, and analyzing and visualization process. Researchers analyze various social media pages using visualisations, to know their popularity in the audience. Article detailed various data visualisation tools- Tableau, Trend Miner, TwitIE, and Twitter Zombie. Along with analytics and tools, the article explains the process included in data journalism like data mining, data identification, data cleaning, and extraction. (Barik, 2015)

Rajeev Ghode, Research Scholar, Department of Communication Studies, Pune University conducted a study on Infographics and visualisation in news presentation of dailies *Times of India and Indian Express explores* various parameters like importance, visual design, instructional approach, space and sections allotted, to use of Infographics with quantitative parameter like numbers of infographics used in both the newspapers' edition for six months. Overall, the study arrived at a conclusion that the Times of India has effectively used all levels of infographics over Indian Express. The quality of infographics is also appreciable in Times of India in comparison with Indian Express (Ghode, 2016).

But while speaking about DJ practice in India, only Indian Express achieves such a fame competing with international standards very recently through the Panama Papers investigations. Here the researcher focused on visualisation and infographics and even failed to address the data journalism perspective.

Data Ethics and Privacy

The ethical implications of collection, storage, and use of big data in the Indian context need to be investigated. An article titled *Privacy in India in the age of big data* discussed the issue of privacy in the present big data age, and how big data gathered through ICT tools and social media platforms can be used against citizens. The possible ways to protect citizens' private data on the Internet in India were also addressed in this article. The study recommended that the State should have a set of clear guidelines on the collection, monitoring, storage, and owning of data, for authorities, tech companies and other stakeholders which are in an area of collecting user data. A strong privacy policy and security measures to protect the citizens from potential cybersecurity threats and misuse of power in the hands of the government and private parties are needed. Users' consent is necessary for collecting data, storage, and use. (Halder, 2015)

Big Data Sets and Indian Polity

The Influence Industry Digital Platforms, Technologies, and Data in the General Elections in India is a research article by Elonnai Hickok (2018) tries to find out different types of companies, platforms, and techniques involved in the generation, collection, and use of data in elections with a focus on the 2014 and 2019 elections towards mapping the data ecosystem for elections. It also explores government regulations around the use of digital platforms and data in Indian elections. For this purpose, news items, academic articles, policy and legislation, company websites and promotional material, campaigning and election material such as advertisements and voter roll in the context of the 2014 and upcoming 2019 elections in India were identified and studied. The study concluded that the use of emerging technologies and data influence the impact on the fairness, independence, and impartiality of elections. It suggested that a Data protection law is needed in India that extends to the public and private sector and recognizes the evolving nature of data and rights

Data Journalism Initiatives

In India, data-intensive journalism is only more than just to fly. Big media outlets are investing heavily, but the real action is happening among the non-profits and start-ups in the country. Such initiatives are aiming to break new ground in public-interest journalism. More individuals depending on mobile phones and tablets for data, it is inescapable that more stories will be told through sophisticated data and visuals.

The availability of the data in reusable format works as a raw material for data journalism. It becomes important to study the data journalism platforms and initiatives available in India. A study was conducted to study the initiative taken by *indiaspend.com* on data journalism. Content analysis of the *indiaspend.com* website as a whole was carried out and along with it, two stories are picked from different beats. Stories from the year 2016 from January to March were selected randomly one each from the section 'Latest news stories' and 'Special section'. The Special section consists of one broader theme under which various related stories are covered regularly.

Bringing data into journalism: the Indiaspend project also investigates the trajectory of *Indiaspend's* growth and reach through analytics. Also, discuss the prospects of data journalism and computer-assisted reporting through interviews and views of the founder of the organization and the editor. s an attempt to study how public discourses drove the establishment of *Indiaspend* and the use of computer-assisted tools in reporting through

graphic visualizations, that *Indiaspend* specializes in. (Chaitanya Mallapur, 2015)

Newsroom Challenges with Data

Digital technology and new practices like data journalism made a huge impact

in Indian media organisations. While looking at the expertise with data journalism in Indian newsrooms, there are various hindrances and challenges journalists find to adapt themselves with data and its various implications. The finest way to integrate data journalism in Indian newsrooms is to provide ample acquaintance and knowledge on data extraction and analysis tools to deal with. The data extraction process still remains an obstruction for media persons. They considered the data extraction and visualisation process is out of their discipline. So simple and uncomplicated programming tools or apps need to be introduced in order to decode and analyse the huge open-source datasets available in our country. (Lakshmi, A,2020). Journalists also face the scarcity of resources while incorporating data in newsrooms. Data stories held as higher standards for accuracy and corrections in a newsroom. Data journalism practice will bring more traffic to news audiences.

Data journalism would help to build a more democratic, reliable, accurate and informing news cycle India. It would help to keep the Indian audience informed more and better than ever before and the people in a position of power accountable than ever before. Data Journalism is the need of the hour in our country, ie; data stories should act as a shift from fake news to confirmable and trustworthy news. It would allow the journalists to abide with news reporting based on the facts and also allows the audiences to scrutinize those facts. Every journalist can do low-tech DIY (Do It Yourself) data journalism and then only data journalism can transform the Indian media industry (Narula, 2019).

Data Journalism has found reflections in India in the form of organizations possessing data desks and implementing data segments in their news. Journalists have apprehensions towards the inadequacy of updated data and also with the data formats available. Even the data available in Indian government's *india.gov.in* is outdated. And also they are concerned about the data literacy on the part of authorities, mainly in the case of open data sources. Data is uploaded in PDFs, scanned and JPEG format which make the data complicated to clean and process. Data policy of our country looks gloomy and uncertain (Kashyap et al., 2020).

India is one of the first countries to join open data movement by introducing National Data Sharing and Accessibility Policy (NDSAP) in 2012. However, the open data policy by Indian government does not mention anything about promoting data literacy and data skills within government departments or any provisions for a capacity building mechanism. (Kashyap et al., 2020). Journalists have to depend on open tools as news organizations are unwilling to develop in-house tools for fetching and processing data. Data teams are small in number in legacy media organizations and they rely on freelancers and other data journalism platforms for data-based stories.

Data Journalism Education

Indian data journalists are heavily dependent on data aggregates and start-ups provided by the govt authorities. This reveals the lack of data sourcing methods prevalent in our country. In India, journalists hardly rely on crowd sourcing for data crawling. Media can also effectively use RTI act to get data. While looking at data analysis and visualisation context, free tools and less interactive maps and visuals are used. Using of coding language instead of DIY tools will definitely improve the data stories' standard.

An involvement is need from the part of journalism educators in India with a pedagogic approach to impart the entire data dissemination process and for developing and analysing data driven stories. They can highly contribute for bringing better data literate professionals. This pedagogic policy will make the legacy media organisations equipped for practicing data journalism and the over dependence on the data aggregates can be reduced. Such an approach improves the industry standards as well as the quality of journalism education in India.

(Kashyap & Bhaskaran, 2020)

Research Gaps Identified and Suggestions for Future Research

As the data journalism is still in its infancy in our country, same is the case of research related to data journalism. Almost all the articles selected for this study are exploratory in nature. The study recommends more research works should happen regarding various aspects of Indian data journalism ecosystem. Conceptualisation is very crucial for every field to flourish. More studies may explore the changing nature and conceptualisation of data journalism practices in India. This will help to understand the methods and techniques of data journalism.

Ethnographic studies could help in shaping diverse insights towards data journalism. It could even bring theories or new business models in data

driven journalistic practice. India being a place with varied cultural characteristics, regional media organisations and their dealing with data stories could be studied.

Some studies selected here discussed about the challenges faced by journalists while dealing with data. But no epistemological explanations were provided for combating the challenges faced in newsrooms. More comparative studies related to data journalism practice in different media organisations (legacy media and DJ platforms) should be initiated for conceptualising theories examining the data journalism practice in India. No studies tried to explore the quality of the data driven stories published in India so far. A comparative analysis is possible with data driven stories appeared in different newspapers or with newspapers and data news portals. Text, visualisations and multimedia content of the story may be explored. Future studies may explore data skill and knowledge of journalists in all broad processes of data journalism- data scraping, filtering, visualising and presenting.

Best thing to integrate data journalism in newsroom can be done only when the journalists have adequate knowledge on data. Data journalists have conclusively great part to play as a public service agent that changes the deluge of public data into findings and meaningful stories that facilitate a democratic society. However, in view of the challenges to extensive acceptance of data journalism, it is necessary that interventions are cautiously and deliberately planned. As data journalism is well known and highly prevalent in India, there is plentiful scope for research in this area in future. Future studies may look at how Data Journalism practices and evolved over a period since its adoption by various media organizations in India. And the changing nature in the practice of data journalism by different media firms could be enquired.

References

- Rogers, S. (2013). *Facts are sacred: The power of data*. London: Faber and Faber.
- Gray, J., Chambers, L., & Bounegru, L. (2013). *The Data Journalism Handbook*. Beijing: O'Reilly.
- Bratsas, C., & Veglis, A. (2018). The Principles, Features and Techniques of Data Journalism. *Studies in Media and Communication*, 6(2), 36. doi: 10.11114/smc.v6i2.3208.
- Kumar, A., & Haneef, S. (2016). Convergence of technologies and journalists: Translation of journalistic practices through ANT perspective. *Estudos Em Comunicação*, (22), 105–122. doi: 10.20287/ec.n22.a06

- Wright, S., & Doyle, K. (2018). The Evolution of Data Journalism: A Case Study of Australia. *Journalism Studies*, 20(13), 1811–1827. doi: 10.1080/1461670x.2018.1539343
- Hammond, P. (2016). From computer-assisted to data-driven: Journalism and Big Data. *Journalism: Theory, Practice & Criticism*, 18(4), 408–424. doi: 10.1177/1464884915620205
- Knight, M. (2015). Data journalism in the UK: a preliminary analysis of form and content. *Journal of Media Practice*, 16(1), 55–72. doi: 10.1080/14682753.2015.1015801
- Stalph, F., & Borges-Rey, E. (2018). Data Journalism Sustainability. *Digital Journalism*, 6(8), 1078–1089. doi: 10.1080/21670811.2018.1503060
- Lakshmi, A. (2020). Study on Data Journalism in Tamilnadu & the Challenges. *International Journal of Innovative Science and Research Technology*, 5(6), 1105–1111. <https://doi.org/10.38124/ijisrt20jun715>
- Kashyap, G., & Bhaskaran, H. (2020). Teaching Data Journalism: Insights for Indian Journalism Education. *Asia Pacific Media Educator*, 30(1), 44–58. <https://doi.org/10.1177/1326365x20923200>
- Kashyap, G., Bhaskaran, H., & Mishra, H. (2020). We need to find a revenue model”: Data journalists’ perceptions on the challenges of practicing data journalism in India. *Observatorio (OBS) Journal*, 14(2), 121–136.
- Narula, D. S. (2019). India and Data Journalism Revolution. *Journal of Content, Community and Communication*, 10(9). <https://doi.org/10.31620/jccc.12.19/01>
- Shah, R. U. (2020). We Don’t Need More Data, We Need the Right Data. *Circulation*, 142(3), 197–198. <https://doi.org/10.1161/circulationaha.120.045968>
- Hickok, E. (2018). *The Influence Industry Digital Platforms, Technologies, and Data in the General Election in India*. ourdataourselves.tacticaltech.org.