# "We need to find a revenue model": Data journalists' perceptions on the challenges of practicing data journalism in India

Geeta Kashyap*, Harikrishnan Bhaskaran**, Harsh Mishra**

*DAV University, Jalandhar, India
**Central University of Himachal Pradesh, Dharamshala, India

## Abstract

Data journalism is emerging as a new subspecialty in Indian journalism. Indian data journalism practice is in nascent stage and newsrooms have lately started recognizing it as a new practice. There were specific issues and challenges faced by journalists in western news rooms as they adopted the practice. To generate a comparative perspective, the study carried out semi-structured interviews with Indian data journalists about their perceptions on the challenges of practicing data journalism. The results show that there are external and internal bottlenecks in practicing data journalism in India. While irregular data releases, use of unfriendly formats by governmental authorities is a major concern; reluctance of media organizations to invest in data teams is identified as an internal challenge. Indian data journalists are skeptical about the audience's participatory potential, which has restricted them from relying on practices like crowd-sourcing.

Keywords: data journalism, data journalists, data policy, Indian media, India, Journalism Studies

## Introduction

Data journalism, a new subspecialty in journalism practice, has now seems to be normalised as a wide-spread practice in newsrooms globally. In India also, the new practice has found reflections, in the form of organizations developing data desks and rolling out data sections in their news products. However, a new set of professional practice, data journalism in this context, faces certain challenges when it gets adopted. This study seeks to examine such challenges, as perceived by the data journalists in India, to understand the adoption of data journalism in Indian media industry. Such an examination also helps to develop comparative perspectives about the adoption of a new practice across different media industries.

In India, digital news organizations such as IndiaSpend, How India Lives, Livemint, Health Analytics India, Factly, Newslaundary and legacy news organizations such as The Hindu, Hindustan Times, and Indian Express are producing data-intensive news stories. The Hindu was amongst the first legacy news organizations in India to introduce 'Data' as a separate section on its news website. Data journalism was termed as a slow start in India and is predicted to gather pace in near future (Oberoi, 2017). It became evident also when news organizations started investing in data journalism by introducing data teams in newsrooms. The Hindu's 'Data' section shows publication from April 2015 onwards. After a year, Hindustan Times launched its 'Interactive' section, carrying data news stories with dominating, interactive visualizations. Journalists from Indian Express were part of the international, data-intensive investigative project 'Panama Papers' and 'Paradise Papers', a landmark in collaborative investigative journalism, which developed deep insights from the huge cache of leaked data on investments in tax havens. BBC endeavors to capture audiences through data journalism in

India by inviting technical companies to develop tools for utilizing open data for storytelling (BBC, n.d.)[1]. News organizations are also investing to generate primary data to build databases for their news products. For instance, IndiaSpend team has installed air quality sensors developed in-house to monitor air pollution levels and generate open data about air quality in different cities of India under the project 'Breathe'. 'Breathe' secured 'Honourable mention' in Data Journalism Awards 2016. In 2016, Hindustan Times bagged three awards at the 25th Malofiej International Infographics Awards (Kakade, 2017)[2]. Apparently, data journalism in India is spreading as well as also bringing international recognition. Hate crime watch, a project by Factchecker has won 'The AP award for best data journalism team portfolio (small newsroom)' at Data Journalism awards in 2019.

Studies about data journalism practices in Indian newsrooms are largely absent in academic research about data journalism (Ausserhofer et.al, 2017). Rajasekar (2014) noted that Indian media is quite unresponsive towards the data journalism revolution and recommended that it should recognize the scope of data in journalism. Oberoi (2017)[3] observed that the growth of data journalism is slow in India. Global literature on data journalism has explored diverse aspects of the practice and discussed various perspectives of data journalism. Though, such studies acknowledged the roadblocks and challenges in the development of data journalism (for example Weber and Rall, 2012; Appelgren and Nygren, 2014), they did not specifically attempt to understand the challenges and limitations experienced by data journalists. The study aims to explore the issues and challenges faced by journalists while covering data stories in India.

**Review of Literature**

There have been many attempts to understand the phenomenon of data journalism as a news practice in different nations (Fink and Anderson, 2014; Knight, 2015; Borges-Rey, 2016; Lim, 2018). Many of these tried to define data journalism as a practice or a set of processes. For instance, Howard (2014) defined data journalism as a process which involves "gathering, cleaning, organizing, analyzing, visualizing and publishing" data in order to carry out "acts of journalism". Some other studies focused on the form of the content being produced. For instance, Knight (2015) defined a data story as a news story which has a numerical peg as opposed to an anecdotal one and which carries "substantial" data or visualization elements. Baack (2011) ascribed data journalism as a 'new style of news reporting'. Initial studies on data journalism were mostly to investigate about 'what' and 'how' of the practice and later scholars started studying data journalism beyond simply understanding the workflow (Ausserhofer 2017).

[1] BBC News Labs (n.d.). Growing Global Audiences Through Data Journalism in India.
[2] Kakade, S. (2017). How the Hindustan Times is introducing Indian readers to interactive stories. (in Storybench)
[3] Oberoi, S. (2017). Baby steps: A slow start but data journalism in India is gathering pace:266-272.

*Studies on data journalism practice: some trends*

Past studies examined different aspects of data journalism, like the impact of data-driven approach in newsrooms and emergence of new professional roles. Initial studies focused on different emerging forms of journalism like computer-assisted reporting, computational journalism and interactive journalism. Several scholars have tried to distinguish these emerging practices with that of data journalism, trying to conceptualize its contours. Parasie and Dagrial (2012) studied how newsrooms in Chicago hired programmers for producing data-driven news stories. According to them, in mid 2000s, the role of programmer-journalists turned out to be important when data analysis for investigative news stories become very vital. The study highlighted how the inclusion of programmer-journalists in newsrooms and contemporary journalism in context of public good brought an epistemological change in journalism. Felle (2016) argued that data journalism has a role which is similar to that of traditional journalism which reinstates the idea of journalism as the 'fourth estate'. The study observed that digital data reporting has strengthened the role of journalism in society and it is expanding the scope of investigative reporting. Examining the accountability and responsibility of data journalists towards society, the study suggested that digital data reporting is a significant resource to preserve the role of media in society as a watchdog. Tabray, Provost and Trottier (2016) have studied data journalism in context to Quebec newsrooms and identified new job profiles and actors in newsrooms involved in producing data-based stories. Veglis and Bratsas (2017) explained data journalism taxonomy based on data journalism projects. Stalph (2017) classified data journalism and observed how gradually data stories have changed into 'in-depth, long form, investigative and visually sophisticated projects.'

The award-winning data stories have attracted the attention of academic researchers a lot, making it a repetitive element in the literature about data journalism. Young, Hermida and Fulda (2015) analyzed projects submitted for data journalism awards by different organizations in Canada to understand the constituents of great data journalism stories. They highlighted the visual elements as well as other factors which constitute a good data story. Ojo and Heravi (2017) studied award-winning stories submitted to Global Editor's Network Data Journalism Awards to identify recurring elements in advanced data journalism projects. Similarly, Loosen, Reimer and De Silva-Schmidt (2017) analysed award-winning stories with an objective to understand actors involved in producing data journalism stories, topics covered under data journalism and the visual elements used in producing data stories. It showed that most of the award-winning data stories originated from large news organization's online departments. The study also found that the teams which produce data stories have diversified expertise and job profiles like writers, programmers and graphic designers. Majority of the nominations and the winners of data journalism awards are from the West, especially from countries like United States and United Kingdom. Because of this trend, it can be seen that the academic literature on data journalism also focus more on newsrooms in the West. Studies about data journalism practices in India are very scarce. Bajpai (2013) documented different initiatives in India where journalists, data enthusiasts and programmers form communities such as Hacks/Hackers and also organize meets and conferences which were significant in the development of data journalism in the country.

*Adopting data journalism practices: the global experience*

A news organization, especially a legacy one, faces several challenges when adopting innovations. For instance, Westlund & Krumsvik (2014) found that legacy newsrooms face the challenge of enabling collaborations between departments while trying to adopt an innovation, especially since the enthusiasm to integrate innovations is not evenly distributed across different departments in the organization. A perceived lack of interest from the management in integrating innovations in the newsroom is another challenge in this regard. Academic literature on data journalism shows that such concerns on adoption of innovations in newsrooms are relevant in case of data journalism practices also.

Rogers et al (2017)[4] show that lack of interest from management in promoting the practice is a major challenge. This reluctance takes the shape of editorial hesitancy in acknowledging specific job roles within news rooms in UK (Borges-Rey 2017) and US (Fink and Anderson, 2014). Such indifference from top management has resulted in lack of division of labour in newsrooms and shortage of financial resources allocated for data journalism projects as well (De Maeyer et al, 2015). What could be the reason for such a lukewarm response from the top management in news organisations? Aitamurto et al (2011) and several others argue that the absence of a viable revenue model centered on data journalism products as a major obstacle in this regard.  Data journalism projects are mostly associated with impact but less with revenue (Aitamurto et al, 2011). However, the relationship between web traffic and ad revenue has some influence on considering popular data stories are viable news products (Aitamurto et al, 2011) often forcing data journalist to look for stories in data that can be popular (Fink and Anderson, 2014). Wright & Doyle (2019) point out that the time and resources associated with the data journalism production often make them financially unviable, especially in the eyes of the top management. Moreover, news organizations have to compete with other entrants in the ecosystem of data brokering markets to explore possibilities of finding a viable revenue option outside the traditional news business (Aitamurto et al, 2011). These kind of restraints force organizations, especially the small ones to bring down the size of data journalism teams, often at the expense of expertise like legal proficiency (Fink and Anderson, 2014).

Other major challenges reported by past studies include skill shortage and issues of collaboration are also related to the issue of indifference from organizations to invest more in data journalism.  For instance, lack of expertise and skills needed for producing data stories are reported as a major challenge for its adoption (Halevy & McGregor, 2012; Borges-Rey, 2016). The work-around is collaboration with other departments within the organizations or with other professionals like statisticians, programmers and designers (Borges-Rey, 2017; De Maeyer et al, 2015; Appelgren and Salaverria, 2018). However, such regular collaborations are not affordable for news organizations (Borges-Rey 2017). Another option to fill such gaps of expertise is depending on generic tools (Borges-Rey, 2017). While such tools are limited in their scope in terms of the

---

[4] Rogers, S., Schwabish, J., & Bowers, D. (2017). Data journalism in 2017: The current state and challenges facing the field today. Google News Lab

interactive options and sophistication (Loosen, 2019), organizations are reluctant to use such third party tools since they have limited branding options (Rogers et al, 2017) as opposed to in-house tools.

Data journalists face several challenges at an individual level as well. Adopting innovations in newsrooms often require journalists to reorient the existing professional norms and values (Boyles & Meyer, 2016). For instance, in case of data journalists in UK, Borges-Rey (2016) found that the struggle to strike a balance between scientific rigour needed to deal with datasets and the story-telling norms of journalism is a challenge journalists face at personal level. Similarly, data journalists working in sports beat might be under editorial pressure to use expensive data but still struggle to connect the insights with the context of the game event being reported (Horky & Pelka, 2016).

Several internal factors in a newsroom also adds to these challenges. For instance, the effort to get data journalism practice acknowledged as a serious form of journalism inside the newsroom is tremendous (Borges-Rey, 2016; Fink and Anderson, 2014), especially in the absence of newsroom elders who have expertise in such innovations (Rogers et al, 2017). The incompatibility of existing digital infrastructure (like content management systems) with the third-party tools used by data journalists is another newsroom obstacle (Borges-Rey, 2016). Such third party generic tools are often relied upon as a work-around for lack of expertise in coding. However, the gap between the creative potential of the data journalist and the limited scope of such tools is an issue data journalists have to negotiate frequently (Boyles & Meyer, 2016). The struggle to keep updated with the changing tools and programming languages is reported as a major challenge for data journalists by several scholars (Borges-Rey, 2016; Halevy & McGregor, 2012; Wright & Doyle, 2019). Such fast-paced learning is more laborious than routine news work, which makes it a difficult feat to achieve in case of working journalists. This has resulted in data journalists showing low-confidence in their data products (Boyles & Meyer, 2016). Journalists' fear of numbers, skill shortage and lack of training in accessing data, data cleaning and visualization production are some of the major challenges of data journalism adoption in news organizations (De Maeyer et al, 2015; Appelgren & Nygren, 2014; Rogers et al, 2017; Halevy & McGregor, 2012; Weber and Rall, 2012).

Factors outside the news organization also add to these challenges. For instance, data journalists tend to rely on government data more (Knight, 2015; Rogers et al, 2017) but government datasets are plagued by several issues. Karlsen and Stavelin (2015) observed that journalists working with data struggle with investigating public records. It is usually very difficult to find out what kind of data exists and to access it in digital format. There is a lack of trained public servants as they do not possess enough knowledge about data export and digital transfer techniques. While getting access to these datasets in machine –readable format is a major challenge (Aitamurto et al, 2011), additional challenges include irregular, incomplete and out-dated data releases (Cruz & Carmona, 2019; De Maeyer et al, 2015) and the prolonged bureaucratic struggle involved in getting access to these datasets (Borges-Rey, 2017; Fink and Anderson, 2014; Appelgren and Nygren, 2014; Karlsen and Stavelin, 2015). Lack of granularity of data (De Maeyer et al, 2015), datasets with errors and incomplete metadata (Aitamurto et al, 2011) are issues often related to datasets released by government in many countries. In some countries, datasets related to certain pressing issues like land-conflict (Sambhav &

Aliwal, 2019) are almost absent or incomplete, forcing data journalists to restructure them on their own (Cruz & Carmona 2019).

A review of data journalism literature by Ausserhofer et al (2017) shows that academic attention about data journalism practices are skewed geographically with a lot of attention given to Western news rooms. For instance, more than half of the studies (22 out of 40 studies, 55 percent) reviewed by Ausserhofer et al (2017) were about data journalism practice in the US or UK, while studies about Asian news organizations were absent. This trend is now changing with some recent studies examining the practice in China (Lowrey & Hou, 2018) and in India (Kashyap & Bhaskaran, forthcoming). However, even such attempts do not focus on challenges faced by data journalists in Asian newsrooms.

This study seeks to address this gap in the literature by trying to understand the Indian situation in practicing data journalism. For this purpose, the study tries to answer the following Research Questions:

> RQ1: What are the main challenges in the development of data journalism practices in a developing country like India?

To put the findings of the first research question in context, the study seeks to compare it with the insights shared by similar issues reported by the existing academic literature. To understand if the issues faced by data journalists in India are similar to that faced by data journalists in other countries another research question was formulated:

> RQ2: How do the challenges faced by data journalists in India differ from that of their counterparts in the Western countries?

**Methodology**

Indian journalists who are practicing data journalism or the journalists who identify themselves as data journalists were interviewed and were asked to share their experiences. Data journalists were identified initially from bylines in the data stories published in different news organizations. Later, a snowball sampling approach was employed where the initial respondents were asked to suggest data journalists from their contacts to find more interview respondents. Hendricks & Blanken (1992) points out that snowball sampling has the potential to reach to a special population where it is difficult to locate the members of the special population. It is an efficient way to find contact and suits if the nature of the study is explorative, qualitative and descriptive. The method is criticized for being vague and unsystematic and it is difficult to generalize the results on the basis of a small sample. Despite its weaknesses, it is usually used in studies which seeks data and results from a special population (Biernacki and Waldorf, 1981). Eleven journalists (working with Indian digital media start-ups or English language, mainstream media organisations in India), were interviewed using semi-structured interviews (See Table 1). In Indian media houses, it is difficult to find big data teams. The

usual data team has two to three people who deal with the production of data stories. Journalists interviewed as part of the study are either part of data teams or in some cases, work in traditional journalistic roles in an Indian news organization but practice data journalism as well. IndiaSpend is the only organization which claims to be a dedicated data journalism initiative in India. Therefore, journalists who are doing freelancing for IndiaSpend were also approached as part of finding respondents for the study. Potential respondents were also located from the professional social networking site LinkedIn by using the keyword 'data journalist'. Among the professionals in the search results, those who have mentioned themselves as data journalists working in Indian news organizations were approached. Interviews were conducted telephonically, face to face and through emails depending on the convenience of the participants. Interviews were recorded, transcribed and analyzed qualitatively to identify recurring themes. An initial review of interview transcripts suggested some common themes. Later these themes were developed by a second reading and coding of the interview transcripts by two coders. A constant comparative method (Strauss & Corbin, 1990) was used to ensure that the themes identified are valid and reliable.  The discussion is categorized under varied themes, by giving a representative quote for the theme.

**Table 1:** List of interview respondents

| Name | Organization | Designation | Gender |
|------|--------------|-------------|--------|
| DJ1 | Hindustan Times | Data Journalist | Male |
| DJ2 | Indiaspend.com, Scroll.in, factchecker.in | Freelance data journalist | Male |
| DJ3 | Indian Express | Senior Editor | Male |
| DJ4 | Hindustan Times | Data Journalist | Female |
| DJ5 | Hindustan Times | Data Journalist | Male |
| DJ6 | The Hindu | Senior Reporter/Data Visualiser | Male |
| DJ7 | Indiaspend.com | Program Manager/ Data Journalist | Female |
| DJ8 | How India Lives | Founder/ Data Journalist | Male |
| DJ9 | BBC, New Delhi | Data Journalist | Male |
| DJ10 | Rediff.com India Limited | Data Journalist | Female |
| DJ11 | The Hindu | Chief of Bureau | Male |

**Analysis**

Journalists practicing data journalism in India perceive various roadblocks which need to be tackled to take forward the practice of data journalism in Indian news rooms. On a broader perspective, Indian journalists' experiences depict a gloomy picture of the status of the practice. Indian data journalists face difficulties while deriving a story out of numbers. Non-existent, inaccessible or unreliable data are some of the problems faced by the journalists. Journalists admit that there is lot of scope in practicing Data Journalism in India but they feel constrained in many ways.

*Data sources, format and accessibility*

One of the problems mentioned by almost every journalist who were interviewed as part of the study was the non-availability of updated data and concerns related to the format of data available. A data journalist working with a national newspaper shared the following opinion:

> "The data portal of Indian government, data.gov.in, has outdated data……... Moreover, data literacy on part of the government is another big challenge as there is no one who understands the language of data. Data is uploaded in PDFs, scanned and JPEG format, which takes a lot of time to clean and process, before we actually start analysis." [DJ1]

Another journalist said:

> "In India, there is no data, it is quite frustrating. It is very hard to find data. What is more frustrating, there is data but government does not want to share it. There is lot of stone-walling happening from the government, not to release data. And the fact is that the depth of data has been reduced, as compared to the past. Yes, the websites of the government have improved a lot but the depth of data is no longer there. So, you feel constrained, the kind of story you want to do, you are not able to do so." [D7]

She also mentioned that there is no regularity in publishing data on government websites. "For example, NCRB data comes with two-year lag. It's been three years, but still the data related to accidents and suicides has not been released" ,she said.
Journalists shared the opinion that open data requires lot of processing before it can be used for analysis and interpretation. Lack of categorization of data-sets, accessibility and quality of data are among the various issues of open data available in India.

> "The biggest challenge is that the data is in bad format. In India, based on my experience, I am confident to say that 80% of our time goes into cleaning the data. Some of the data is in the form of scanned PDF. So how to make it machine-readable is a tough challenge. When we try to convert certain data coming in PDF to Excel, the alignment is lost. Some of the data might not be updated. So, cleaning the data is the biggest challenge", said one journalist who is the team member of digital start-up, How India Lives. [DJ8]

*Lack of data policy*

Journalists are unhappy with the government's response to issues of data availability, regular data updates and the formats of data release. Government policy regarding open data is also uncertain. There is lot of ambiguity regarding the licensing of the data as government data is under copyright in India. Also, many of the provisions of licensing open data are discouraging. India is a developing nation and 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 (Kodali, 2017)[5].

Right To Information (RTI) Act (a freedom for information legislation which allows citizens to demand and access government data), which is termed as a weapon for Indian citizens to maintain accountability and transparency in government departments is not much fruitful for the journalists. Journalists said that there are many loopholes in the whole process of getting data from the government departments through RTI requests. It takes 30 days and sometimes more to get data and, in many cases, converting the dataset into a data story, after receiving, is a bleak possibility.

Then there is always a trust issue with government data. Reporters complained that it is difficult to believe in the data issued by the Indian government. "Sometimes we find mistakes in the data released by the government and pointing out this to the government authorities may offend them and they may even stop talking to us", said one journalist. Journalists also expressed disappointment as they are not finding good opportunities to work with primary data and have to mostly rely on secondary data.

> "I don't see a future for data journalism in India in the present form. The government can lie through its teeth and we cannot disprove it. Because slowly all the existing data sources are either becoming unavailable or inaccessible or poorly reported", said a journalist working with The Hindu. [DJ6]

---

[5] Kodali, S. (2017). Not Open or Accountable: The Government Open Data Use License Is Flawed. (In The Wire)

*Lack of resources inside newsrooms, absence of revenue model*

Moreover, news organizations are not very enthusiastic about investing in data teams. Covering data stories takes time and it is not similar to what is produced in newsrooms as routine. The investment is more and the end-product is sometimes not a part of news-cycle. It is difficult to justify the stories covered once or twice in a month when you have tight budgets.

One of the senior journalists in India working with The Hindu thinks that choosing business model for online journalism is one of the biggest challenges which also affect the growth of data journalism-based news websites.

> "A larger challenge, I think, is about data journalism practice online. Organizations in India are not able to decide to what extent they have to invest in something like this because creating a good data journalism team requires good investment. So, at a time when you are struggling to find a revenue model for online, how far you can go? I think that is a basic challenge. Technically, it [data journalism practice] is always a challenge because it is a complicated, technical subject. For a news organization, it does not look at things so technically. So, they are not able to decide how to prioritize it or where to fit it in" he said.  [DJ11]

Journalists repeatedly mentioned about the lack of resources in newsrooms which affect their efficiency and frequency of producing data stories. Mostly, newsrooms lack in-house tools for producing data visualizations and journalists had to depend on free tools available. Some respondents said that free tools sometime restricted the scope of the stories they were producing.  While learning technical skills is basic challenge for creating visualization, Indian data journalists who want to produce visualizations face other problems too. One journalist said that absence of shape files for creating maps makes the task more difficult.

Another journalist gave a vivid account of the challenge as follows:

> "Preparing visualization on Bihar floods was a challenge. To create maps, shape-files are required, but there were no shape files available. We had to go to different webpages of Bihar government to find the image of the blocks of a particular district. Our colleague extracted those images then he lined up that and then traced them, to make those shape files. At the end you see one map, but to build that map, that person took almost seven days". [DJ4]

*Audiences and Crowdsourcing*

Journalists in India also feel lack of support and participation from Indian audience.

> "In India data journalism is in an infant stage. There is a mindset among the common people that only journalist should do all these things but it's very wrong. Every individual who knows data analysis and has a passion for writing should practice data journalism. There is lack of awareness also about this profession. Many people do not even know there is a profession called data journalism," said one journalist working with a digital start-up. [DJ10]

Indian Data journalists are not relying on crowdsourcing as a substantial data source as it is used by data journalists in other parts of the world (Borges-Rey 2017; Loosen, 2019). At present there are not many data stories which used crowd as a data-resource. Political atmosphere and the pressure groups greatly influence media reporting on different issues. 'Hate Tracker', a crowd-sourced project about hate crimes in India started by Hindustan Times in 2017, was pulled down later by the organization without providing any reasons.

A journalist from Hindustan Times who was interviewed as part of the study, told about the project that they were creating a database named 'Hate Tracker', to track hate-based crime/violence occurred since September 2015. Detailing about the project, the journalist said that reporters daily updated the database with reports on hate crimes based on caste or religion. Public could also participate by reporting the cases where they have faced or they have come across a hate-based crime. The information provided by the public is then verified by the news organization before adding it to the database. The project was partly crowd-sourced. Despite its initial success, it was eventually shut down.

Indian data journalists think that crowdsourcing possess a great potential as a data source if used sanely. Data and reports provided by crowd require verification before consequential news reporting. According to a journalist working with Hindustan Times, New Delhi:

> "In our project we have asked people to provide reports only, not like if someone had abused you then also you report to us. It has to be if you have been abused and some paper has reported the incident then that will be included in our database. So, this is purely news report-based database. In this case, crowdsourcing has been used in a different angle. And crowd-sourcing does not have such potential where we can talk in absolute numbers. Like, we cannot just say that out of 100 people 50 say yes and 50 say No. It's not a poll.  But it's an excellent resource to reach out. Once you find sources, do reporting. Do not trust people blindly. Ask people to submit proper documentation. Double-check stories. In India, nobody is doing it, but it's not like we cannot do it in India. Only thing is that we should cross-check them.  Crowd-sourcing has the immense potential. In India, nobody is using that." [DJ4]

Journalists interviewed shared their opinion about crowdsourcing and why it is problematic at this stage to work on crowd-source based data journalism projects in India. Another journalist working with Hindustan Times, New Delhi said, "Right now, we do not crowd-source. There is an issue of quality control. But there are benefits also, like you can get data from the public about any issue." One journalist said, "Crowd may have its own ideology and that ideology may affect the kind of data they are going to provide".

Besides all these challenges, journalists are quite optimistic about the growth of data journalism in India. One journalist shared the following opinion:

> "In these years what change I have noticed is the increasing acceptance of Data Journalism in India and I see it increasing in a short of period of time. So yes, lot of people come to this field, many organizations have started hiring. In terms of acceptance, it is really great and we have crossed that stage where we actually have to sell data journalism to media houses. Like any other form of journalism such as sports journalism, digital journalism etc. data journalism will be the form of journalism and more and more organizations will spend money in this aspect." [DJ8]

## Discussion

Based on the response, the challenges faced by data journalists in India can be ascertained. The major roadblocks/challenges in the development of data journalism practices in India seem to be related to availability of data other than skills and resources. There is reluctance from government authorities to provide datasets. Open data is perceived as less reliable and journalists are skeptical to depend on data released by the government. Journalists are forgoing the opportunity of working with primary data due to lack of resources and internal support in newsrooms. This has mainly restricted their data journalism practice to press-release style reporting of the  data aggregates provided by government agencies. However, to put it in context, it shall be understood in comparison with the global experience in this regard.

The problems and challenges of Indian data journalism practice are similar to what has been identified by scholars in their studies about data journalism practice elsewhere. For instance, reluctance from the government officials to release datasets in time and in machine-readable format is a challenge faced by data journalists globally (Aitamurto et al, 2011). Similar issues have been reported by data journalists working in French-speaking Belgium (De Maeyer et al, 2015), Cuba (Cruz & Carmona, 2019), US (Fink and Anderson, 2014), UK (Borges-Rey, 2016) and in Quebec (Tabary et al. 2016). Lack of resources is a problem of small newsrooms in many countries (Borges-Rey (2015); Young, Hermida and Fulda (2017); Fink and Anderson (2014 ). Large newsrooms are resource-equipped and also have bigger data teams. However, in India, lack of resources and lack of skills is a problem with  small and large organizations.. Big data teams rarely exist and data journalists are perceived as multipurpose journalists as they have to do routine reporting also, along with producing data stories.

Similarly, the apathy from organizations in terms of promoting the practice is identified as a challenge by Indian data journalist in line with their counterparts elsewhere (Rogers et al, 2018). They are also aware about the challenge of finding a sustainable revenue model for the growth of data journalism in the country.

Certain aspects of data journalism production, like creating visualizations and crowdsourcing are being approached differently by Indian data journalists. Borges-Rey (2017) found that almost all data journalists in UK who were a part of the study tried crowd-sourcing at least once. Their concerns and challenges related to crowd-sourcing was mainly related to ensuring audience participation to a scale which is required to make it a success. However, Indian data journalists who were respondents used crowd-sourcing in rare instances only and were more skeptical about the quality of crowd-sourced data. Creating a visually appealing, interactive visualization is still not the supreme cause of the Indian data journalists. Though concerned about the lack of expertise in this regard, Indian data journalists believe that offering too much interactive options to the Indian audience is not necessary. Such an approach makes them less concerned about the challenges of producing sophisticated visualization with interactive options, out of line with the global academic opinion regarding the importance of good data visualizations Weber and Rall (2012).

Indian journalists also feel lack of support from Indian audience. They think there is no participation from the audience and there is no popular acceptance for data stories among the general Indian audience. For data journalists in India, how to incorporate crowdsourcing into news production and data journalism practice is still a big problem.


## Conclusion


The study intended to explore the challenges and roadblocks in the development of data journalism practices in India. It has identified certain problems and challenges Indian journalists encounter when they work on data stories. External forces as well as internal institutional forces influence the process of producing data stories. Data stories, unlike routine stories, are dependent on datasets, which are often acknowledged as the starting point of any data story. Without access to datasets, data journalism practice is not possible. Therefore, inaccessibility to data, government officials' hesitance to share data, format of data available, lack of digital data literacy among government officials, irregularity in updating data by the government are among the external challenges which affect the coverage of data stories in India.

Data journalists face certain internal challenges as well. Journalists acknowledged that they lack in skills required for creating visualizations and some journalists mentioned their newsrooms are not ready to invest in data-teams.  Journalists have to depend on open or free tools as news organizations are not ready to invest money in developing in-house tools. Data teams are small in number in case of legacy media organizations. In fact, the exclusive data journalism initiative teams are also small in number working in-house and they also rely on freelance journalists.  This apprehensive approach from news organizations in terms of investing  to develop expertise of the existing teams is an internal institutional force which impediment the growth and development of data journalism practice in India.

Journalists expressed concern about the present condition of data availability, accessibility and other factors. However, journalists also perceive that the practice has a good future in India, if few things could be improved. However, the current trends in Indian media industry show that the practice is getting more normalized with mainstream organizations limiting their sophisticated practice of data journalism to small teams while the general reporting integrates more data-journalistic practices like minimal interactive visualizations to supplement routine stories on web. This scenario could be problematic since it often puts forward a 'data for sake of data' approach where data aggregates provided by the government or other agencies, are used in a story with minimal interactive visualizations.

The study is based on a small sample of journalists from mainstream as well as digitally-native news organizations only. Future studies may include more journalists or even enrich the insights through focus group studies. In a country like India which is known for its diverse media practices, the challenges faced by vernacular media will be different than that of English media. Future studies may try to examine these challenges as well.

**References**

Aitamurto, T., Sirkkunen, E., & Lehtonen, P. (2011). Trends in data journalism. Next Media Hyperlocal. Espoo, Finland: VTT. 0-27

Ausserhofer, J., Gutounig, R., Oppermann, M., Matiasek, S., & Goldgruber, E. (2017). The datafication of data journalism scholarship: focal points, methods, and research propositions for the investigation of data-intensive newswork. Journalism, online first. https://doi.org/10.1177/1464884917700667

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

Appelgren, E., & Salaverría, R. (2018). The Promise of the Transparency Culture: A comparative study of access to public data in Spanish and Swedish newsrooms. *Journalism Practice*, 12(8), 986-996.

Baack, S. (2011). A new style of news reporting: Wikileaks and data-driven journalism. *Cyborg Subjects*, 1-10.

Bajpai, R. (2013). In the search of Big Data. In S. Winkelmann (Eds.). *Data Journalism in Asia (*pp. 21-31)*. Singapore: Konrad-Adenauer-Stiftung

Becker H (1963) *Outsiders: Studies in the Sociology of Deviance*. New York: The Free Press.

Biernacki, P., & Waldorf, D. (1981). Snowball sampling: Problems and techniques of chain referral sampling. *Sociological methods & research*, *10*(2), 141-163.

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

Borges-Rey, E. (2017). Towards an epistemology of data journalism in the devolved nations of the United Kingdom: Changes and continuities in materiality, performativity and reflexivity. Journalism, online first. https://doi.org/10.1177/1464884917693864

Boyles, J. L., & Meyer, E. (2016). Letting the data speak: Role perceptions of data journalists in fostering democratic conversation. *Digital Journalism*, 4(7), 944-954.

Cruz, Y., & Carmona, S. (2019). Reassembling Public Data in Cuba: How Journalists, Researchers and Students Collaborate When Information Is Missing, Outdated or Scarce. In J. Gray & L. Bounegru, The Data Journalism Handbook 2 (1st ed.). European Journalism Centre and Google News Initiative. Retrieved from https://datajournalism.com/read/handbook/two

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

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

Felle, T. (2016). Digital watchdogs? Data reporting and the news media's traditional 'fourth estate' function. *Journalism*, *17*(1), 85-96.

Halevy, A. Y., & McGregor, S. (2012). Data Management for Journalism. *IEEE Data Eng. Bull.*, 35(3), 7-15.

Hendricks, V. M., Blanken, P. (1992) Snowball sampling: Theoretical and practical considerations. In Hendricks, V. M., Blanken, P., Adriaans, N. (Eds.), Snowball sampling: A pilot study on cocaine use (pp. 17–35). Rotterdam, the Netherlands: IVO. Howard, A. (2014). The art and science of data-driven journalism. New York: Tow Center for Digital Journalism. Retrieved from http://towcenter.org/research/the-art-and-science-of-data-driven-journalism/.

Karlsen, J., & Stavelin, E. (2014). Computational journalism in Norwegian newsrooms. *Journalism practice*, 8(1), 34-48.

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

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

Lim, J. (2019). Representation of data journalism practices in the South Korean and US television news. *International Communication Gazette*, *81*(1), 89-113.

Loosen, W. (2019). Data-Driven Gold-Standards: What the Field Values as Award-Worthy Data Journalism and How Journalism Co-Evolves with the Datafication of Society. In J. Gray & L. Bounegru, The Data Journalism Handbook 2 (1st ed.). European Journalism Centre and Google News Initiative. Retrieved from https://datajournalism.com/read/handbook/two/situating-data-journalism/data-driven-gold-standards-what-the-field-values-as-award-worthy-data-journalism-and-how-journalism-co-evolves-with-the-datafication-of-society.

Loosen, W., Reimer, J., & De Silva-Schmidt, F. (2017). Data-driven reporting: An on-going (r) evolution? An analysis of projects nominated for the Data Journalism Awards 2013–2016. *Journalism,* online first. https://doi.org/10.1177/1464884917735691.

Lowrey, W., & Hou, J. (2018). All forest, no trees? Data journalism and the construction of abstract categories. Journalism, online first. https://doi.org/10.1177/1464884918767577.

Ojo, A., & Heravi, B. (2018). Patterns in Award Winning Data Storytelling: Story Types, Enabling Tools and Competences. *Digital Journalism*, *6*(6), 693-718.

Shrivastava, K., & Paliwal, A. (2019). Documenting Land Conflicts Across India. In J. Gray & L. Bounegru, The Data Journalism Handbook 2 (1st ed.). European Journalism Centre and Google News Initiative. Retrieved from https://datajournalism.com/read/handbook/two/assembling-data/documenting-land-conflicts-across-india

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

Strauss, A. & Corbin, J. (1990). Basics of Qualitative Research: Grounded Theory Procedures and Techniques. Newbury Park, CA: Sage Publications.

Tabary, C., Provost, A. M., & Trottier, A. (2016). Data journalism's actors, practices and skills: A case study from Quebec. *Journalism*, 17(1), 66-84.

Veglis, A., & Bratsas, C. (2017). Towards a taxonomy of data journalism. *Journal of Media Critiques*, *3*(11), 109-121.

Weber, W., & Rall, H. (2012). Data visualization in online journalism and its implications for the production process. In 2012 16th International Conference on Information Visualisation (pp. 349-356). IEEE.

Westlund, O., & Krumsvik, A. H. (2014). Perceptions of intra-organizational collaboration and media workers' interests in media innovations. *The Journal of Media Innovations*, 1(2), 52-74.

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

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