

## **Data Journalism Concept and Practices in India: A case study of data journalism initiative in India by Indiaspend.com**

**Geeta Kashyap**<sup>\*</sup>  
**M. Rabindranath**<sup>\*\*</sup>

**Abstract:** Technology always affects human life in different ways and information is one of them. The easy access and abundance of information is possible because of digitization of information technology. The one and zero is the answer to almost everything. In the age of globalization and digitization, governments, organizations, and individuals across countries may store and potentially have access to unlimited amount of data which may be available in structured or unstructured format. The availability of the data in reusable format works a raw material for journalism. This has given birth to a new term Data Journalism. The present research paper focuses on the development of data journalism in India. India is quite late to join this data revolution thus; at present it becomes important to study the concept and practice of data journalism in Indian context.

**Key words:** Data Journalism, Quantitative, Information, Technology, Computational Journalism

### **Introduction**

The technology has drastically changed the way we use to look at information. The easy access and abundance of information is possible because of digitization of information technology. The one and zero is the answer to almost everything. In the age of

---

\* **Geeta Kashyap:** Research Scholar, Department of Journalism and Creative Writing, Central University of Himachal Pradesh

\*\* **Dr. M. Rabindranath:** Associate Professor, Department of Journalism and Creative Writing, Central University of Himachal Pradesh

globalization and digitization, governments, organizations, and individuals across countries may store and potentially have access to unlimited amount of data which may be available in structured or unstructured format. Virtually, every country today is in the midst of database revolution where from politicians to bureaucrats, academicians and journalists including general public are shaping a new platform for information collection and dissemination. The information and communication technology has provided a room for changing journalism practices. (Gynnild, 2014).

Whatever we do every day and everywhere it leads to the generation of some sort of data. However earlier it was difficult to capture, store and collect the data especially to sort data in some pattern to create any story. But, it has changed now and the availability of data in reusable format works as a raw material of journalism. The process of disseminating news is now redefined by mechanization. One of the most important questions for journalism's sustainability will be how individuals and organizations respond to this availability of data. (Bell, 2012)

However, in today's time telling a story with numbers is everywhere. Data journalism, is seemingly "everywhere," based on the industry buzz and accelerating scholarly interest. (Fink and Anderson (2014, 1). The phenomenon of data journalism is always related with the revolution of information and communication technology which had made possible to access large amount of data.

These data may be accessed by the public and are easily available for the analyses. News reporters and journalists are accessing these data and may utilize it according to their requirement. Journalists can do a very elaborative and interpretative story with numbers or they may merely include the figures in a news story. It depends on the journalists how they are going to use numbers in their story.

If we observe the current practices of journalism it can be said that these data-centric phenomena, if not transformed but on some accounts are going to influence some of the most fundamental aspects of news and its production and distribution by humans and machines. And when we talk about these changes, it has all become possible because of the growth of new mobile devices, tracking tools and cheap storage system. (Lewis, 2015).

Journalism appears to be taking, as Petre (2013) puts it, “a quantitative turn”, where numbers are at forefront. Data journalism is always used interchangeably with computational journalism and computer assisted reporting. However, there is thin line difference between all these terms. Computational journalism is the output of intersection between journalism and Computer and this has nothing to do with the availability of data and tools. (Karlsen & Stavelin, 2013). Jonathan Stray has given a working definition of computational journalism and he defines it as Computational journalism as the application of computer science to the problems of public information, knowledge, and belief, by practitioners who see their mission as outside of both commerce and government. (Stray, 2011)

Computer-assisted Reporting (CAR) is always tied to investigative reporting and sometimes it is referred that data-driven journalism seems to have taken up mantle of CAR only, in present journalism scenario. However, the definition by Stray broadens the concept of Data Journalism as compared to CAR. There is difference of opinion among scholars too about the CAR as a predecessor of data-driven journalism. The scholars tried to differentiate between CAR and Data journalism on many grounds by broadening its area with the use of data, visualization, graphics and presentation of a news story. CAR is more bound towards investigative stories only. Data journalism has also been closely associated with the use and proliferation of open data and open-source tools to analyze and display that data. (Gynnild, 2014). Data journalism is characterized by its participatory openness and cross-field hybridity, because now data is available for public and from various fields. Hybridity meaning something that is formed by combining two or more things and Data Journalism is simply the outcome of data and journalism. (Coddington, 2015)

The data journalism concept in India is still in its nascent stage. There are very few news organisations which have adopted data journalism as a separate section. There are many organisations which are still far away from this concept because Indian media seems busy with covering as much as stories they can just and put it like breaking news and corner news. Because, it takes time to present a news story in elaborative manner with good numbers.

“When data journalism as a form of journalism was gaining respectability, the Indian media was undergoing a transition away from serious informational and critical journalism to one with a greater focus on soft stories needing less reader or viewer attention,” argues Professor C.P. Chandrasekar of the Centre for Economic Studies and Planning at the Jawaharlal Nehru University, Delhi.

C. Rammanohar Reddy, the managing editor of Economic and Political Weekly (EPW) admits that the EPW does process a lot of information; he says the information is put out in traditional ways that he terms “old” data journalism. He says, “Data Journalism is a highly skilled job; the ability to visualize a story must be combined with the careful use of information. I do not see any Indian media outlet, print or online, doing so satisfactorily, though many are trying.”

There are many viewpoints on data journalism in India. Ravi Bajpai, multimedia editor of the Hindustan Times says that he is quite optimistic about the prospects for data journalism in India but he is quite critical of the quality of open data. He was earlier attached with Down to Earth, a science and technology publication and was using free data visualization tools. On open data he says, “Open data is hardly available. The government’s idea is limited to the Right to Information Act. That legislation is perhaps the best tools data journalists can exercise to access robust data sets.” Along with it he criticizes the data provided by government on one more ground that it could take a long time to access comprehensive government data on basic issues. The time loss proves costly and the story often does not see the light of day.

However, the data.gov.in is an appreciable step to provide a platform for open government data. It has a database for transport time tables, government budget, company register, National statistics, Agriculture, health performance etc. There are 100 departments including state and central about which you can get information at a glance. At a basic level it is a very good initiative but it can be expected that it can be more interactive in future. Indian journalists may delve into the website and can create a story based on data and visualization. (Rajshekhar, 2014)

Using data and facts in a story is a common practice in any newsroom. Data is available and it is accessed by different streams

for different purposes e.g. Environmental Science, Political Science, and Economics etc. How media is using data that is the pertinent problem to be studied. The interpretation of data in words and visualizing news story with numbers is the central idea of data journalism. It is becoming a common practice in media at global level. The large number of data can be accessed easily because of digital revolution. Data from different sources is available in public domain; however, it is not always accessible in simple language. Most of the data is available in complex nature and that required simplification before it can be produced before public in a news format. The studies done on data journalism say that for interpretation of data in understandable news content it requires special skills for a journalist. However, there are different tools also available which helps in converting data into news format with text, visuals, links and video e.g. Pageflow. It eliminates the need for journalists to code. The practice of data journalism has touched a new form of journalism and differentiating traditional journalism from the current practice.

In India there are very few news organizations which are very actually involved in the process of doing stories based on data.

*The Hindu*, *The Indian Express* and *The Time of India* are among the organisations which are doing a bit on data journalism. The Indian Express set the legacy with the involvement of its journalists in the first of its kind of data journalism projects popularly known as *Panama Papers*. If we go by the websites and prints of these newspapers, The Hindu newspaper's website segregates its data based stories by putting stories under the hyperlink 'Data'. Similarly Times of India also publishes one column 'STATOITICS', where text is low and graphs and visualization is high. It is published in the left corner and not a very good space is devoted to this section. However, it can be easily identified in comparison to other stories. On the other hand, Indian express is providing a good coverage of data based stories but it is not segregated under any special sub-section or hyperlink. In this series, news websites have different criteria. They provide recognizable format to the stories. Newlaundry.com works as an aggregator for the data based stories which provides data based stories from different sources like *indiapend.com*, *factly.in* etc. at one place. There are other blogs

also which are devoted to data journalism. Indiaspend.com is a website which is known as first data journalism initiative India, as it claims.

To know the trend of data journalism more extensively and minutely, a study of data journalism website indiaspend.com has been carried out.

In India, www.indiaspend.com is a website which claims to be the first data journalism initiative in India. According to the website, they utilise open data to analyse a range of issues. They are working on a broader objective including better governance, transparency and accountability in Indian government. It was founded in late 2011 by Govindraj Ethiraj, a television and print journalist and former Founder Editor-in-Chief of Bloomberg TV India, One of the objectives is also to make the use of data exciting and a first recourse for young India, rather than opinion. To do so it utilises various social media sources to reach young population. India spend has also taken one more initiative to cross-check the facts stated by various politicians and authorities so that public may also know the difference between the facts stated and actual facts.

### **Research Objectives**

The research paper focuses on the following research objectives:-

1. To identify the significance and status of data journalism in India.
2. To identify the various initiatives taken by mainstream media.
3. To study the initiative taken by indiaspend.com on data journalism.

### **Methodology:**

The method used here is exploratory research method. The exploratory research is concerned with discovering patterns in research data, and to explain/understand them. Research lays basic descriptive foundation. It may lead to generation of hypotheses. (Fitzgerald & Howcroft, 1998, p.160). There are countable articles and papers written on data journalism in India. Thus, the focus was the explorations and establishment of the concept of data journalism in Indian context. For the purpose of the study, analysis of the

websites and the prints of the mainstream media; The Hindu, The Indian Express and The Times of India, was carried out as these are known for their data journalism initiative in India. Content analysis of the [indiaspend.com](http://indiaspend.com) website was done, which claims to be the first data journalism initiative in India. To move further on this, European Journalism centre's initiative on data journalism website of 'Data-driven Journalism Initiative' has been employed as a benchmark for the study. This will not be a comparative study but this will help in knowing the foundation of data-driven journalism in Indian context. Content analysis of the [indiaspend.com](http://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 was 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 on regular basis.

## Case Study

### Parameters for website analysis

#### 1. Sections of the website

#### 2. Frequency of updating

1. **Sections of the website:** There are in total ten sections of the website which includes Home, About, Fact check, Events, States, Data Room, Vizonomics, Donate, Special Sections and Breathe. Home section displays the different stories, placing the latest on the top. 'About' gives information about the website. Fact check they have introduced later as an interesting section for the common public through which they can check where their politicians gone wrong while mentioning the figures. Events is the least updated section, however it displays the events about to happen in future related to Data journalism only. States sub-section is divided into six sections: North, North-east, East, South, West and Central India, which provides information about the states specifically but most of the section is not updated since 2013. Another link Data Room provides graphs and bars on the status of Indian states about health, education, household assets,

population and economic indicators etc. Vizomics includes the videos of different stories covered. Special sections started covering special stories different from the current happenings. They it introduced recently as the first story appeared in February. Each story carries a by-line. And then Breathe is the air monitoring network built where one can check the quality of the air of their respective state by clicking on the location.

2. **Frequency of Updating:** On an average there are 17-20 stories updated in one month. The home page consists of the latest news and it is updated daily in comparison to other section. If we look at the special section of the website it is not updated daily but yes the stories covered are very extensive in nature as they are the exclusive stories.

### **Parameters for Content Analysis of the News stories**

‘Hate-Speech Accused 3 Times More Successful In Elections’ and from special section ‘The Air we Breathe’ the news story ‘15 Days After Odd-Even, Delhi’s PM 2.5 Levels Rise 15%’ is selected for the analysis on the following criteria:

1. **Nature of the story**
2. **Arrangement of the story**
3. **Source of data**
4. **Number of graphs and tables**

**Headline:** ‘Hate-Speech Accused 3 Times More Successful In Elections’

**Date:** March 28, 2016

1. **Nature of the story:** The news story published is from political beat. The story is about the how hate speech accused got 3 times more successful in elections. The story covered is based on the data of last ten years, which draws a comparison between the years and the percentage of increase in the success of hate speech accused.
2. **Arrangement of the story:** The story consists of the numbers and percentage. Along with it, it carries the sections related to the hate speech and recent examples of hate speech.



3. **Source of data:** In this news story Indiaspend.com has used secondary source of data that is self-disclosed crime records of candidates, who have contested various elections in last 12 years.
4. **Number of Graphs and Tables:** There is no visual presentation of the news story. Therefore, the number of graphs and tables is zero.

**Observation:** The above parameters show that the story is based on the interpretation of the news story with the comparison of various data. Twelve years of the data has been converted into a news story, which draws a co-relation between hate speech accused and their success over a period of time. Otherwise, this would have been simple information. But putting an analysis into it and making a comparison of different time period has made this an exclusive story.

**Special Section heading:** ‘The Air we Breathe’ the news story  
News Headline: ‘15 Days After Odd-Even, Delhi’s PM 2.5 Levels Rise 15%’

**Date:**

1. **Nature of the story:** This news story is related to the environmental problem especially on air pollution. The story is covered in public concern and it also raises awareness on the odd-even formula of Delhi government whether it’s effective or not. Therefore, the story helps in creating awareness among public whether any policy is made applicable is really worth or not.
2. **Arrangement of the story:** The news story is arranged with text, graphs, maps and a video. One graph shows the comparison of the concentration of PM (particulate matter) in Delhi air before and after the implementation of the odd-even scheme. Then there is one video which discusses the story based on the Indiaspend coverage. Following that, there is bar graph which shows briefly growth of the PM and then there is hourly description of the increase level on different dates.

3. **Source of data:** The data used here was generated by the Indiaspend.com team itself. They have placed the Breathe air-quality monitoring devices.
4. **Number of graphs and tables:** There are in total 6 graphs, one map, one video and number of tables is zero.

**Observation:** This news story interprets the story with numbers, graphs and maps. It is a very good example of visualisation and presentation of any data-driven story as stories are depicted in a way public can understand it easily. Through numbers this story states the real picture of the odd-even scheme. It draws a comparison between the vehicles, which are more pollutant than other through graphs. For this news story Indiasend.com has used its own sources of data. They have employed their own monitoring devices and collected data over a different point of time.

### **Conclusion:**

The research study 'Data Journalism Practices and concepts in India' explored the phenomenon of data journalism in India. The aim of the research was 'to identify the significance and status of data journalism in India', 'to identify the various initiatives taken by Indian media organisations' and 'to study the initiative taken by indiaspend.com on data journalism'. In order to answer these questions the websites and prints of the media organisations were analysed. The aim was to know how they are reacting to the data-driven journalism. The initiative taken by these organisations establish the significance of the data-driven journalism because every newspaper is giving some or large space to the data based stories. The status is not much good if we compare it with the global media organisations. The Guardian, New York Times are internationally acclaimed for their data-driven stories. In India, only Indian Express is able to achieve such fame and that also very recently through the Panama Papers investigations. The second objective was to take a look on the data journalism initiative by mainstream media. The Hindu is publishing stories under 'Data' hyperlink but not updated every day, it was analysed by checking its web page for the month of March. The Times of India is giving 'STATOITICS', which gives numbers and their little interpretation.

The Indian Express far ahead and giving good investigative stories in their newspaper and the involvement in the Panama Papers project will certainly raise the standards of data journalism in India. The Third objective was to study the initiative taken by Indiaspend.com. The website is publishing good stories on data journalism. At present, they are using two types of sources of data i.e. self generated data and the data generated from other sources like government open data etc.

The Data driven initiative website of European Journalism centre is doing quite elaborative stories. So if we look at both initiatives one at National level and another at International level; Indiaspend initiative is also appreciable in comparison to international. The website is updated every day with the stories from different beats. Data Journalism Initiative is involved in different projects too. Therefore, it will take time for Indian media to match up with the international standards. In India, journalists are facing other problems too like availability of open data and most of the times it takes long time to access comprehensive government data even on basic issues.

Indian media needs to shift its preferences from breaking news to the exploration of new ideas of journalism. Our news channels and newspapers are still busy with old patterns of news stories where selling a story is more important than telling a story. This new concept adaptation may lead to the new phase of journalism where public can really read and listen what is actually valuable for them.

In India, data journalism is still in its infancy but its significance and status is expected to be more concrete in future.

## References

1. Coddington, M. (2015). Clarifying journalism's quantitative turn: a typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism*, 3(3), 331-348.
2. Emily Bell (October 2012) Journalism by numbers [http://www.cjr.org/cover\\_story/journalism\\_by\\_numbers.php](http://www.cjr.org/cover_story/journalism_by_numbers.php)
3. Fink, Katherine, and C. W. Anderson. 2014. "Data Journalism in the United States: Beyond the 'Usual Suspects'." *Journalism Studies*. doi:10.1080/1461670X.2014.939852.
4. Fitzgerald, B., & Howcroft, D. (1998). Competing dichotomies in IS research and possible strategies for resolution (pp. 1-17). ICIS'98 Proceedings of the

- International Conference on Information Systems. Atlanta, GA: Association for Information Systems Atlanta.
5. Gray, J., Chambers, L., & Bounegru, L. (2012). *The data journalism handbook*. " O'Reilly Media, Inc. "
  6. Gynnild, A. (2014). Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets. *Journalism, 15*(6), 713-730.
  7. Indiaspend (January 9, 2016). 15 Days After Odd-Even, Delhi's PM 2.5 Levels Rise 15%. The Indiaspend.com. Retrieved from <http://www.indiaspend.com/cover-story/15-days-after-odd-even-delhis-pm-2-5-levels-rise-15-67571>
  8. Joakim Karlens & Eirik Stavelin (2014) Computational Journalism in Norwegian Newsrooms, *Journalism Practice*, 8:1, 34-48, DOI: 10.1080/17512786.2013.813190
  9. K. Manoj (March 28, 2016). Hate-Speech Accused 3 Times More Successful In Elections. The Indiaspend.com. Retrieved from <http://www.indiaspend.com/cover-story/hate-speech-accused-3-times-more-successful-in-elections-42521>
  10. Lewis, S. C. (2015). Journalism in an Era of Big Data: Cases, Concepts, and Critiques. *Digital Journalism, 3*(3), 321-330.
  11. Petre, Caitlin. 2013. "A Quantitative Turn in Journalism?" Tow Center for Digital Journalism, October 30. <http://towcenter.org/blog/a-quantitative-turn-in-journalism/>.
  12. Rajshekhhar, Priya. (July 21, 2014). *India's Media- Missing Data Journalism Revolution*. Retrieved from <http://gijn.org/2014/07/21/indias-media-missing-the-data-journalism-revolution/>
  13. Stray, Jonathan. 2011. "A Computational Journalism Reading List". Jonathanstray.com, January 31. <http://jonathanstray.com/a-computational-journalism-reading-list>
  14. [www.datajournio.in](http://www.datajournio.in)
  15. [www.datadrivenjournalism.net](http://www.datadrivenjournalism.net)
  16. [www.factly.in](http://www.factly.in)
  17. [www.indianexpress.com](http://www.indianexpress.com)
  18. [www.newslandry.com](http://www.newslandry.com)
  19. [www.thehindu.com](http://www.thehindu.com)
  20. [www.timesofindia.indiatimes.com](http://www.timesofindia.indiatimes.com)