

Post-Print Journalistic Consciousness in the Data Journalism Production Process

Authors: Li Xuan, Wang Mingming

Date: 2023-10-08T00:00:00+00:00

Abstract

With the advancement of Internet technology, big data has demonstrated significant value across various industries. In the journalism sector, the mining and analysis of news elements through big data tools have given rise to a reporting paradigm distinct from traditional news media, namely data journalism. Data journalism relies on Internet big data throughout its production process. This paper, through an analysis of the generation process of data journalism, clarifies the functional mechanisms of each stage in this process, thereby facilitating data journalism in better leveraging its distinctive characteristics to report facts to audiences.

Full Text

1. Data Journalism

Data journalism represents the product of big data's infiltration into the news industry. It is characterized by two most prominent features: (1) data-driven support and intuitive presentation methods. The data support can be further divided into two aspects: first, the value of data journalism can be assessed by capturing and analyzing data from databases, making value judgments based on audience information needs; second, the content of data journalism reporting is saturated with data, using data to present the theme and angle of news reports, and reporting facts through analysis of data correlations. (2) The presentation method of data journalism is visualized, intuitive, and concise, representing its most obvious departure from traditional news reporting methods.

2.1 Topic Selection

Topic selection constitutes the first step in data journalism. When selecting topics for data journalism, two key factors must be considered: first, news value,

and second, whether the content is suitable for data journalism presentation methods.

Data collection and organization require considerable time, making data journalism potentially less timely than traditional news. Therefore, the timeliness requirements of different news types must be fully considered during topic planning. Data journalism reveals the logical relationships behind data, often involving the reuse and deep excavation of existing news data to discover underlying relationships. This undoubtedly determines that topic selection must emphasize significance, making the news data worthy of excavation. Proximity refers to the relevance of the topic to the audience. Before planning a topic, the proximity between the topic and the audience, and the degree of this proximity, should be considered. Big data can provide accurate rankings of this relevance.

Data journalism's reporting characteristic lies in presenting the interrelationships among data to the audience through visualization techniques. These presentation types include images and text, audio, video, etc., all serving as visualization methods for data journalism.

Data journalism is primarily applicable to topics with large volumes of data where certain relationships exist between the data, and where revealing and deeply exploring these relationships holds news value.

2.2 Data Collection and Screening

[Note: The original text contained corrupted characters at the beginning of this section which have been omitted.]

Data journalism production is based on a team that must include four roles: journalist/editor, data analyst, graphic designer, and programmer. The journalist/editor is responsible for content organization, the data analyst handles data collection and screening, the graphic designer manages graphic design and optimization, and the programmer is primarily responsible for presenting graphics through code. In practice, however, one person often fulfills multiple roles.

3.1 Data Quality

In data journalism, data acquisition occurs through obtaining shared data from government agencies or from non-governmental organizations and enterprises. News media rarely collect data themselves due to the massive engineering effort required and the limited and closed channels for data acquisition. Since data obtained from others is inherently limited, and data journalism is data-driven and cannot exist without data, data acquisition has become an obstacle to the development of data journalism.

Many current data journalism works contain large amounts of data. Professional practice involves annotating data sources, yet some news media fail to disclose their data sources. This is because the data may originate from other media

reports—that is, from third parties. When third parties lack authority, the news becomes less persuasive. This practice damages media credibility and reduces audience trust. Moreover, logical analysis in data journalism, deepened and presented through erroneous data, may likely lead to the spread of fake news among audiences. The authenticity of data sources requires verification, yet the inability to verify sources represents a current malady in data journalism.

3.2 Data Analysis and Presentation

Data journalism differs from news data; the data in data journalism is massive. Although it possesses relative “comprehensiveness” compared to traditional sampling surveys, such large volumes inevitably lead to some data inaccuracies. Therefore, analysis based on this data is difficult to achieve precision and carries a certain degree of ambiguity.

Data presentation amplifies the agenda-setting function of news. Due to the powerful impact of data on audiences, combined with visualized presentation methods, the logical relationships between data reported in data journalism become deeply entrenched among audiences. For example, after CCTV’s “Data Says Spring Festival Migration” aired, audiences formed a vivid perception of Spring Festival travel and became less receptive to news presenting contrary perspectives. This invisibly deepened and solidified the audience’s cognitive schema regarding China’s Spring Festival travel. Due to the characteristics of data journalism, if media exploit this feature to conduct purposeful agenda-setting for audiences, the consequences are undoubtedly significant. In this regard, data journalism presentation is not conducive to the objective shaping of the pseudo-environment by news.

3.3 The Enhanced Gatekeeper Role

In data journalism, data exists objectively, and data analysis is achieved through tools. Data journalism presents the correlations between data derived from analysis. All of this appears to dilute the gatekeeper’s role, but in reality, the gatekeeper’s role in data journalism is not diminished but rather enhanced. The gatekeeper appears and performs their duties during the topic selection process. Once topic selection goes wrong, reporting will lead to incorrect values and create adverse public opinion. Simultaneously, during data screening and cleaning processes, data selection also reflects the gatekeeper’s role. If data selection is erroneous, flooding valid information capacity with false information, the relative relationships among data will produce corresponding deviations, causing data journalism reporting to lead incorrect social public opinion and harm the normal development mechanism of society.

Conversely, in data journalism, if the gatekeeper slightly manipulates the data, it will create significant social impact. Once ill-intentioned individuals assume the gatekeeper role, the consequences are unimaginable.

4. Countermeasures

For participants in the data journalism production process, it is necessary to strengthen theoretical study and improve journalistic professionalism while also learning multiple skills. As data journalism develops, journalists who only know how to gather and edit may not adapt to the technically demanding production of data journalism. Therefore, journalists must continuously improve their skills. Enhancing journalistic professionalism is an eternal requirement for news practitioners. In the field of data journalism, all personnel involved in news production should improve their professional qualities. Only when individuals with high professional qualities participate in the news production process can they better fulfill the gatekeeper role, achieve balanced reporting in data journalism, and pursue the value and significance of data journalism as journalism itself.

Open data sharing represents an important approach to solving data source problems and expanding data journalism reporting angles. Once data is openly shared, data distortion and errors will be minimized, and large amounts of data can enable news media to excavate different correlations behind the data, forming rich reporting themes. This can also prevent the “monologue” phenomenon where certain media monopolize data, and other media cannot verify whether reports based on exclusive data are reasonable. Although this is not yet possible at the current stage, it represents a hope for the healthy development of data journalism.

When selecting appropriate visualization presentation methods for data journalism, audience comprehension levels must be fully considered. Although data journalism possesses intuitive visualization, reports on professional fields must be matched with corresponding accessible explanatory text; otherwise, non-professional audiences will lose interest in reading such information, and the advantage of intuitive presentation disappears. Even for data journalism reports on non-professional fields, narrative storytelling should be fully integrated to prevent dry, lifeless data reporting. Data journalism presentation methods that incorporate storytelling both cater to audience comprehension and give data journalism vibrant tension.

Data journalism has injected fresh blood into the development of journalism and promoted the era theme of media convergence. Data-driven journalism makes news stories more objective and intuitive, and for audiences, it also conforms to current fast-reading habits. However, data journalism is still developing through continuous exploration and has many drawbacks, presenting new research topics for both academia and the industry. (Author’s affiliation: Hebei University of Economics and Business)

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.