On Data Awareness of Data Journalists in the Big Data Era and Its Cultivation: Postprint

Authors: Liu Yongxi

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

Abstract

In the era of big data, the proliferation of information has catalyzed the emergence of data journalism through journalists' mining, analysis, and innovative presentation of massive datasets. However, in practice, journalists encounter substantial challenges regarding source gatekeeping, journalistic production, presentation modalities, and value concepts due to their deficiency in data consciousness. To facilitate further development of data journalism, journalists should promptly cultivate rational data consciousness: they must enhance their capabilities in data processing and new technology application to elevate data literacy, while simultaneously acknowledging the artificiality and complexity of data to circumvent data pitfalls.

Full Text

On the Data Awareness of Data Journalists and Its Cultivation in the Big Data Era

Liu Yongxi
Kaifeng Daily, Kaifeng, Henan 475000

Abstract: In the big data era, the surge of information has propelled the rise of data journalism through journalists' mining, analysis, and innovative presentation of massive datasets. However, in data journalism practice, journalists face enormous challenges in source gatekeeping, gathering and dissemination, presentation methods, and value concepts due to their lack of data awareness. To further advance data journalism, journalists should promptly establish rational data awareness: they must strengthen their data processing and new technology application capabilities to improve data literacy, while also recognizing the artificial and complex nature of data to avoid falling into data traps.

Keywords: big data; data journalism; gatekeeping; humanistic literacy; gathering and writing; visualization

CLC Number: G647
Document Code: A
Article ID: 1671-0134(2021)01-105-03
DOI: 10.19483/j.cnki.11-4653/n.2021.01.032

Citation Format: Liu Yongxi. On the Data Awareness of Data Journalists and Its Cultivation in the Big Data Era [J]. China Media Technology, 2021(01): 105-107.

1. Challenges Facing Data Journalists

Data journalism is a form of visual news based on data analysis and computer technology. It uses data in news narratives to present content that would be difficult to convey through text alone, or discovers problems through data analysis to unearth news stories [1]. As an emerging interdisciplinary and cross-domain news reporting method in the big data era, data journalism currently focuses on topics in politics (mainly electoral activities), finance, energy, and environment—areas closely related to public affairs and national livelihood that are not easily understood or explained through traditional text or graphics [2].

As early as 2009, numerous internationally renowned media outlets including *The Guardian*, *The New York Times*, *The Washington Post*, BBC, and CNN had already launched data journalism-related operations. In 2012, with support from Google News Lab and the Knight Foundation, multiple media organizations jointly established the world' s first award to honor outstanding works in data journalism: the Data Journalism Awards (DJA). In contrast to the flourishing research on data journalism in foreign academic and industry circles, China' s theoretical exploration and practical application of data journalism started relatively late. In 2011, China' s three major web portals—Sohu, Sina, and NetEase—launched columns such as "Digital Way," "Graphic News," and "Data Reading" [3], marking the beginning of data journalism practice in China. Shortly thereafter, Xinhua News Agency, *People' s Daily*, and *Yi Du* also began exploring data journalism.

Today, data has become a carrier of social changes and human thinking and emotions, containing numerous news clues that serve as a source for journalists to obtain information and create news value [4]. This has brought about tremendous transformations in journalists' work patterns and value concepts, endowing their work with new connotations: using visualization as the primary presentation method and making data analysis results the core driver of reporting logic. Journalists' roles have also become more diverse: they are responsible not only for information collection, design, and production but also for participating in content promotion. In other words, journalists' work now extends from initial topic selection through to final user feedback throughout the entire process.

1.1 Source Gatekeeping in Data Journalism

In data journalism practice, changes in news sources have presented two major challenges for journalists' information gathering and writing. First, data sources are extensive and information is complex, making authenticity difficult to verify and authority hard to establish. Second, although current data sources available to Chinese journalists include data collected and released by government agencies, social organizations, and enterprises, user data from various media platforms, geographic information from mobile terminals, and status data from the Internet of Things [5], the anonymous nature of the internet means that individuals, organizations, and institutions differ in their motivations, sense of responsibility, and accuracy when providing data, resulting in varying degrees of credibility: some data is highly reliable, some is false and distorted, and some is maliciously fabricated [6]. Additionally, databases indiscriminately include vast amounts of unanalyzed and unverified data. Furthermore, news control limits data openness, relevant laws remain incomplete, and journalists have few channels for obtaining data, which restricts the scope of data information collection. Finally, inadequate personal information protection mechanisms mean journalists sometimes obtain personal data through improper channels, infringing on public privacy to some extent.

1.2 News Gathering, Writing, and Dissemination

The depth of journalists' data collection, investigation, and analysis forms the foundation for determining the quality of news reporting and is key to assessing whether journalists can identify main threads and problems within complex social phenomena. Currently, journalists' deep mining and processing of data remain at a relatively low level. Even when articles appear to contain numerous charts and data, they fail to achieve in-depth reporting based on data. Second, Chinese journalists have not yet developed systematic data awareness and are unfamiliar or even ignorant of restructuring methods such as orderly integration, arrangement, and presentation of multi-dimensional data. This leads to strong subjective coloring in data analysis, mistaking random results for regular patterns and confusing correlation with causation [7], resulting in biased reporting that compromises objectivity and neutrality. For instance, in DT Finance' s 2017 article *A Survival Guide to Wudaokou: What' s It Like to Hang Out at the "Center of the Universe" ?*, the journalist merely captured Weibo check-in comments from around Wudaokou and hastily conducted text and word frequency analysis on this data. The conclusion that "overall, people living near Wudaokou seem relatively happy" [8] clearly lacks causal logic: users checking in near Wudaokou do not necessarily reside there, and more importantly, a momentary mood cannot represent one' s entire life. This one-sided viewpoint inevitably sparked endless public debate.

1.3 Interactive Presentation Methods

Visualized interaction represents a major highlight of data journalism: by transforming originally disordered, chaotic, and obscure data into logical, interactive news reports, it not only stimulates readers' interest but also enhances their reading experience. In reality, however, journalists often overlook the applicability of data journalism and abuse visualization forms to explain news, increasing the difficulty for the public to understand it. Moreover, journalists' screening, filtering, and analysis of complex and cumbersome data remain superficial, leaving news works lacking in appeal and interactivity and affecting dissemination effectiveness. For example, in NetEase' s 2018 article *2018' s First Family Portrait of Dog People: Wishing You a Prosperous Year of the Dog*, the journalist collected 87 illustrations focusing on dogs. Such complicated and monotonous images not only affect readers' experience but also interfere with their comprehension of the theme.

1.4 Journalistic Value Concepts

Data journalism' s emphasis on and reliance on data technology has created an imbalance between humanity and technology for journalists. In traditional journalism' s conceptual system, people are the subject of news, the purpose of news is to serve people, and news narratives revolve around individuals' existence, lives, and stories. However, in current data journalism practice in China, journalists overemphasize discovering social problems from data relationships to conduct relatively macro-level social discussions, while inadequately explaining individual emotional experiences and value cognition. This behavior—emphasizing the universality of data journalism while neglecting individuality, focusing on visualization technology while ignoring narrative perspective—seriously affects the public' s cognitive experience and modes of identification [9].

2. Cultivating Journalists' Data Awareness in Data Journalism

In the big data era, journalists can neither ignore data' s increasingly important role in news reporting nor pursue data at the expense of long-accumulated humanistic literacy and insight into human affairs. Therefore, to promote better development of data journalism, journalists should promptly establish correct data awareness and continuously improve themselves in four aspects: source gatekeeping, gathering and dissemination, presentation methods, and value concepts.

2.1 Enhancing the Authority and Credibility of Data Sources

In today' s media society, whether journalists can clearly label sources after thorough investigation is crucial for helping readers grasp key information in a short time. First, journalists should conduct detailed investigations of data sources and clearly label them. If data is too cumbersome, lengthy, and disordered, they

can collaborate with journalists across multiple platforms to jointly construct a data source traceability mechanism. When source verification is difficult for journalists alone, they can also provide links or documents to data sources within the news, mobilizing the public to question and verify data authenticity and reliability. Statistics show that among the 2017 Data Journalism Award-winning works, 89% clearly labeled data sources, with 46% providing links to original data [10]. Second, given the reality that most of China' s current data comes from third parties such as government departments, enterprises, and research institutions with imperfect data disclosure systems, journalists should attract users to create data on their platforms through reasonable and standardized methods while opening and sharing data with users. Finally, journalists must use data reasonably and legally to avoid infringing on public data privacy.

2.2 Improving the Scientific and Professional Quality of Data Journalism Gathering and Writing

The big data environment has changed journalists' workflow and functional positioning. Journalists must possess not only basic professional knowledge and topic planning abilities but also fully leverage their keen social perception to increase the dissemination and influence of data news reporting. First, journalists can identify news value and plan topics from daily life based on their unique insights into public needs and social issues. Simultaneously, they can use data technology to explore correlations behind data and build bridges between individuals and society. For example, DT Finance' s data journalism project "Subway 1 Kilometer" used subway stations as benchmarks to collect and analyze various types of big data within a 1-kilometer radius, including real estate agencies (residential data), Amap (traffic data, public transport data, and Amap indices), Dianping (merchant POI data), and cloud housing data research centers. After cross-analyzing different data, they established a comprehensive evaluation system for "Subway 1 Kilometer" that includes service facilities, dining and entertainment, transportation popularity, housing prices, office functions, and multiple other indicators to observe the ecological trajectory of population flow, commercial changes, and economic development in various urban areas of Chinese cities. The entire report is both humane and interesting while possessing profound social value. Second, since social media can not only disseminate news quickly and widely but also serve as a source for data journalism production, journalists should value and reasonably utilize social media' s data aggregation and dissemination functions: on the one hand, by capturing and analyzing information from social platform users; on the other hand, by encouraging social platform users to participate in news tip provision and content dissemination. For instance, *The Guardian* journalists solicited opinions from global users based on presidential election results and, using the feedback data, produced the report *Obama Reelected US President: Global Public Attitudes*. Finally, journalists must master basic data analysis and mining methods, using data technology and logical thinking to interpret social issues without letting technology become a bottleneck in the transition from traditional to data

journalism.

2.3 Enhancing Visualization and Simplification in Data Presentation

In today' s mediated society, whether journalists' data processing results are concise, aesthetically pleasing, interesting, and vivid, and whether they align with current fragmented and timely reading habits, are key to producing good visualized data journalism. First, conceptually, journalists must clearly recognize that data journalism is not news that simply uses data, nor news composed of data; it has specific operational applicability, and not all news is suitable for data journalism. Data only transforms into information through use and presents its meaning through combination with stories. In news work, journalists should select appropriate, clear, and accurate forms to collect, mine, and analyze data based on news type, content, and audience to help the public better understand the core message. Second, in preliminary work, journalists should emphasize quantitative research methods and approaches in data gathering and writing, striving to find or select data information suitable for the news topic and type that can effectively interact with users and enhance reading experience. For example, the article *All Key Information About the Wuhan Defense Battle Is in This Map* clearly presented the paired assistance to Wuhan from 29 provinces and cities through proper data use: using a map of Wuhan as the background, 29 arrows representing medical assistance from each province, and highlighting "Wuhan" with a heart-shaped marker to demonstrate national unity of "when one place suffers, support comes from all directions" in facing disasters, while metaphorically expressing the government and people' s determination to win the epidemic battle. This news report' s visualization and simplification are worthy of reference.

2.4 Upholding Humanistic and Public Values in Data Journalism

Data' s inherent meaning is limited. The purpose of data journalists using data is not only to predict the future or discover hidden social patterns but also to explore relationships and meanings between people behind the data. Therefore, in data journalism gathering and production, journalists must strengthen critical thinking and spirit, improve their humanistic literacy, and demonstrate public concern and compassion. For example, Southern Weekend' s feature article *China' s Environmental Patients* tells the story of residents living near pollution sources who suffer from "public nuisance diseases" but lack legal protection. Using the individual Zhao Guangming as a breakthrough, the report demonstrates that this news event is not one person' s tragedy but a social tragedy: over 38 years, 400 cancer patients died at the Sulfur Factory Hospital, 90% of whom suffered from arsenic poisoning [11]. The data usage here is not merely to present facts but to examine deep social issues from individual lives, sparking heated public discussion about environmental protection and vulnerable groups' survival—achieving dissemination effects difficult for traditional text reporting to attain. Data journalism is not just simple fact presentation;

excellent data journalism can also provide profound viewpoints, build bridges from individuals to society, and reveal the life value implied in data. By analyzing complex and disordered data, journalists can deeply explore correlations behind data; using clear, readable, and visualized presentation methods to explain complex and changing situations achieves effects that pure text reporting cannot. Meanwhile, journalists must view data rationally. The vividness and depth of news stories can never be replaced by data, and long-term lightweight reading will affect the public' s critical thinking ability [12], thereby hindering social development. Therefore, data journalists must, on the one hand, value data application in news, accelerate the establishment of correct data awareness, accurately and rapidly capture data information with news value, and actively learn visual design tools and cultivate statistical analysis abilities to continuously improve their data literacy. On the other hand, data journalists must also guard against "data-only-ism," uphold humanistic spirit and journalistic ideals, highlight social value and individual life connotations behind data, and help people find solutions to complex problems in an uncertain world.

References

[1] Wu Xiaokun. Data Journalism: Theoretical Inheritance, Conceptual Applicability, and Definitional Dimensions [J]. *Journalism & Communication*, 2017(10): 120-126.

[2] Li Xiguang, Zhang Xiaoya. Journalism in the Big Data Era [J]. *Journalism & Communication*, 2013(01): 7-11.

[3] Yuan Man, Qiang Yuexin. Review and Prospect of Data Journalism Research in China [J]. *Journal of Zhengzhou University*, 2016(2): 130-135.

[4] Li Xiangfu, Li Guangqing. Thoughts on Data Journalism and Visualization in the 5G Era [J]. *Urban Party Newspaper Research*, 2020(09): 87-88+91.

[5][10] Wen Weihua, Li Bing. Data News Reporting in the Big Data Era—A Case Study of *The Guardian* [J]. *Modern Communication*, 2013(05): 139-142.

[6] Ding Baiquan. Data Journalism: Value and Limitations [J]. *Editorial Friend*, 2014(07): 84-86.

[7] Shen Hao, Tan He, Wen Lei. Data Literacy of Data Journalists [J]. *Youth Journalist*, 2014(21): 17-18.

[8] Luo Jie, Deng Shaogen. Problems and Countermeasures in Data Journalism Production and Dissemination [J]. *Youth Journalist*, 2017(22): 41-42.

[9] Liu Tao. Understanding the Concept of Data Journalism: Visualization Practice Criticism and Reflection on Humanistic Concepts of Data Journalism [J]. *News & Writing*, 2019(04): 65-71.

[11] Yu Xuan, Xiao Tao. Visual Narrative of Data Journalism in *People' s Daily* WeChat Official Account—A Case Study of COVID-19 Epidemic Reports [J].

Journalism Research & Development, 2020(16): 235-236.

[12] Zhang Junhui. From "Digitalization" to "Datafication" : Deconstruction and Reconstruction of Data Journalism Narrative Models [J]. *China Publishing*, 2016(08): 39-43.

About the Author: Liu Yongxi (1976-), male, from Kaifeng, Henan, graduate degree, intermediate journalist, research direction: journalism practice.

(Executive Editor: Yang Hu)

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