

Post-print Discussion on the Development of Data Journalism Visualization in the Era of Big Data

Authors: Xiao Zhuoming, Luo Qiyuan, Wu Xian

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

Abstract

With the proliferation of intelligent devices and the rapid advancement of Internet technology, the concept of big data has emerged. Empowered by robust technological support, the journalism and media industry mines meaningful information from fragmented and complex structured and unstructured data, integrates and analyzes it to uncover its profound value, thereby giving rise to an interdisciplinary and cross-domain novel form of news production—data journalism. Data journalism enriches investigative methodologies and narrative techniques, enhances the authenticity, objectivity, and readability of online news, yet simultaneously confronts threats and challenges. This paper attempts to conduct a multi-level analysis, explore issues encountered in the visualization development of data journalism, and propose targeted recommendations, aiming to contribute modestly to the advancement of data journalism.

Full Text

Preamble

Title: Exploring the Development of Data Journalism Visualization in the Era of Big Data

Abstract: With the proliferation of intelligent devices and rapid advancement of internet technology, the concept of big data has emerged. Supported by powerful technologies, the news media industry now mines meaningful information from massive volumes of structured and unstructured data, integrating and analyzing it to uncover deep value. This has given rise to an interdisciplinary, cross-domain news production model—data journalism. Data journalism enriches investigative methods and narrative techniques, enhancing the authenticity, objectivity, and readability of online news, yet it also faces threats and challenges. This paper attempts to analyze the problems encountered in the

development of data journalism visualization from multiple perspectives and proposes targeted recommendations, aiming to contribute modestly to the advancement of data journalism.

Keywords: big data; data journalism; visualization

Authors: Xiao Zhuoming, Luo Qiyuan, Wu Xian

1. Overview of Data Journalism Visualization

Data visualization emerged alongside statistics and advanced significantly with the advent of computers. As IT, internet, and mobile internet technologies have flourished, human behaviors and interactions are intelligently recorded, rapidly processed, and expressed, generating vast quantities of structured and unstructured data through internet and smart hardware—giving rise to the concept of big data. With the popularization of cloud computing, distributed processing, storage technologies, and sensing technologies, big data has found extensive application and exploration across various industries. This concept is quietly transforming journalism, with the most representative outcome being the birth and development of data journalism.

In today's news market, single-dimensional interpretation of events can no longer satisfy public demand; audiences increasingly prefer multi-dimensional data analysis. This requires journalists to interpret news events from multiple angles and perspectives, deeply exploring the relational chains behind the data—precisely why data journalism visualization has gained traction and widespread adoption. Data journalism visualization enriches investigative methods and narrative techniques, enhances news authenticity and objectivity, accelerates the transformation of traditional media roles, and drives social development through trend prediction and public opinion guidance. Consequently, its application is becoming increasingly widespread, though numerous problems have emerged during its development.

Data journalism practice began with outstanding foreign media outlets such as *The Guardian*, BBC Radio, *The Wall Street Journal*, and *The New York Times*, which started exploring data journalism in the early 21st century. Through collaborative online efforts, media practitioners jointly authored the renowned *Data Journalism Handbook*, which offers high practical value and utility. As data journalism has evolved, increasing numbers of institutions, journalists, freelancers, and enthusiasts have established dedicated websites to showcase their data journalism works across numerous fields. Compared to mainstream media, data journalism started relatively late in China. However, with the introduction of data journalism production concepts and the natural trend of news digitization in the big data era, this field has become a battleground for new media transformation. Platforms such as NetEase's "Data News," Sohu's "Digital Path," Caixin's "By the Numbers," and Sina's "Graphic World" all strive to present

facts through data, providing lightweight reading experiences.

Leveraging the convenience of self-media platforms and the development of H5 technology, traditional media has also invested heavily in data journalism visualization in recent years, with remarkable data visualization reports frequently appearing on WeChat public accounts and Toutiao accounts. Excellent visualization news can spread rapidly through self-media platforms, extending the influence of traditional media to a broader audience. Publications such as *The Beijing News*, *Southern Metropolis Daily*, *The Paper*, *Read*, and *South Reviews* have even established dedicated data news sections in their print editions. It is evident that data journalism visualization has moved from the periphery to center stage.

2. Practical Problems in Data Journalism Visualization

Early experiments combining data visualization with news reporting pioneered the field and have achieved significant results to date. However, data journalism visualization still faces substantial challenges, including data ambiguity, limited transparency, potential crime or social crises, uneven data literacy among editorial teams, and low penetration of data visualization tools.

2.1 Limited Open Data Resources and Low Data Accuracy

Data sources for data journalism are diverse, including information publicly released by organizations and enterprises, data from international research institutions, user behavior data from social media, and data accumulated by media websites. However, much of this data is not legally accessible to the public. Although China has continuously promoted information disclosure in recent years, most disclosed data comes in the form of reports and statements rather than raw data, preventing deeper analysis, processing, and mining of source data. Additionally, as various industries increasingly recognize the value of data and strengthen confidentiality measures, the information they proactively disclose is selective. Low levels of data openness represent a major obstacle to data journalism development—the openness and sharing of data are urgent issues that must be addressed for the field to grow.

While big data offers powerful statistical and integration capabilities, any human-collected data inevitably contains some degree of ambiguity. Data journalism works are essentially based on existing research findings, representing secondary value mining of historical data. If large-scale data is distorted, the resulting analysis will significantly diminish in guiding value and may even lead to unpredictable consequences. The “truthfulness” of data determines the quality of journalism, making it essential to screen and verify data during collection to avoid distortion caused by basic errors.

2.2 Privacy Security Risks and Data Monopoly Leading to Social Instability

Big data provides tremendous convenience for monitoring and predicting people's lives, yet personal privacy is simultaneously exposed to invisible third-party surveillance. Internet service providers—whether e-commerce platforms, search engines, or social networking sites—mine and analyze user behavior for commercial gain, inevitably threatening ordinary citizens' privacy. If data falls into the hands of unscrupulous businesses, it will inevitably trigger a frenzied commercial storm, with merchants potentially resorting to extreme measures to promote products based on consumer preferences and travel patterns. Without effective ethical mechanisms to protect privacy rights, social problems can easily arise, trampling on human values such as fairness, freedom, equality, and dignity.

Accompanying the development of the data industry are various conflicts and contradictions, with data monopoly being a primary concern. Data monopoly refers to important data being controlled by a few entities and distributed and enjoyed unreasonably. In the big data era, businesses can control or monopolize commodity prices based on data, creating economic monopolies. If data is maliciously resold, it could significantly increase social instability.

2.3 Shallow Application of Data Journalism Visualization

Although data journalism visualization has achieved some development driven by new and traditional media, and some excellent works have emerged, its overall application in China remains superficial, manifesting in several aspects.

2.3.1 Monotonous Visualization Forms and Lack of Depth Due to cost and speed advantages over dynamic graphics, most domestic data journalism works adopt static forms, using line charts, pie charts, bar graphs, and other information graphics to supplement textual information. The few dynamic interactive works primarily employ timelines and slideshow formats with similar visual elements, resulting in formulaic news production. Overall, these works suffer from low recognizability, insufficient dynamism, lack of innovation, and limited interactivity with readers.

Data itself holds no value; only through analytical interpretation and deep mining can new value and functionality be created. However, most domestic data journalism currently serves merely as bait to attract readers—information graphics that explain new phenomena or clarify certain structures without comprehensively narrating the full story of news events or organically integrating with other content. The communication of data journalism's inherent meaning has not perfectly aligned with depth and profundity.

2.3.2 Scarcity of Predictive Visualization News Specific computer algorithms can automatically analyze large datasets to uncover hidden relationships and patterns. By collecting data and reasonably predicting future trends and

potential relationships, data journalism can fulfill its core mission. As Viktor Mayer-Schönberger, author of *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, stated: the essence of big data is prediction—not teaching machines to think like humans, but applying digital computation to massive datasets to predict the likelihood of events. However, predictive news currently constitutes a very small proportion of data journalism, which instead focuses primarily on explanatory illustrations. The field has not fully mined data value or studied relationships behind the data. Using existing data to predict future trends and directions represents the top priority in data journalism—a method to engage readers, a source of value, and a responsibility to the audience.

2.4 Lack of Professional Talent and Lagging Team Building

Traditional news reporting can be completed independently by a single journalist or freelancer. However, excellent data journalism works generally require teamwork due to complex production processes and significant time requirements, including collaborative efforts in topic selection, data collection and organization, data analysis, and visualization. Currently, China's data journalism is relatively nascent, with teams mostly composed of traditional news editors who lack a learning process in data journalism, resulting in low overall specialization. Due to insufficient data mining and analysis capabilities, these teams struggle to identify valuable leads and produce largely homogeneous news presentations. Quality data journalism requires producers with sophisticated data literacy, keen data intuition, and analytical skills—the ability to identify valuable data within massive databases, excavate stories behind the data, develop worthwhile news topics, establish data models based on databases, monitor similar information, and produce predictive news reports that further enhance journalistic value.

2.5 Low Penetration of Data Visualization Tools

Due to the specialized nature of data analysis and visualization, data visualization tools have not achieved high penetration in the media industry. Tools such as Tableau, Plotly, R ggplot2, Raw, and Infogram are already unfamiliar to traditional media practitioners, not to mention more fundamental professional visualization programming languages like R and Python. Because of low tool penetration, many existing visualization news works are completed by journalists and editors with assistance from designers. Deeper data analysis and more sophisticated data presentation may require additional specialized professionals such as programmers and data analysts, representing a considerable cost for traditional media.

3. Development Strategies for Data Journalism Visualization

Given these problems, how should traditional media address them to improve data journalism visualization?

3.1 Leverage Media Strengths to Expand Data Sources

Data is the key to news production. Obtaining massive, effective data and creating high-quality databases constitute the first step in promoting data journalism development.

3.1.1 Promote Data Openness and Sharing Relevant departments and enterprises hold vast amounts of data: tax authorities possess tax data, transportation departments hold traffic data, retail companies have user purchase data, internet enterprises control user online behavior data, and manufacturing enterprises generate substantial production data. However, most of this data remains protected and is not reasonably opened or shared. From a media perspective, progress can be made in “data openness and self-service, data quality management, legal data protection, data preservation, and data sharing and utilization” to advance the data openness process. Drawing on experiences from outstanding foreign media, we should promote information disclosure, build government affairs databases, address political sensitivities, improve data processing capabilities, dilute data concentration, and prevent data monopolies. Simultaneously, we must manage enterprise data effectively, create platforms for data transaction and circulation, establish regulatory departments, formulate cyber laws and regulations, and improve supervision mechanisms for gatekeeping behaviors to ensure legal and secure information use, providing more standardized and comprehensive databases for domestic data journalism development.

3.1.2 Integrate IoT Data and Leverage Social Media Resources Diversifying data sources requires fully mining and utilizing social media resources while effectively integrating IoT data to enhance data journalism dissemination effects. For instance, surveillance cameras throughout cities collect massive amounts of data daily, which could generate enormous value through effective analysis and operation. Social media networks, represented by the “two microblogs and one terminal” (Weibo, WeChat, and news apps), are thriving and often react more quickly than traditional media. They can provide richer content and receive more audience feedback. Leveraging social networks’ sharing advantages to collect user data enables finely tailored personalized services, extending data benefits from the national and social level to the individual level.

3.2 Utilize Multimedia Technology to Enhance Visualization Interactivity

We should leverage the advantages of different platforms to create news works suited to each medium, finding the most appropriate presentation method for every channel. In terms of visual forms, visualization types can expand to include data maps, timelines, word clouds, and relationship diagrams. Graphical applications extend and supplement the connotation of information graphics. In terms of media formats, we can employ audio/video, 3D technology, virtual reality, and facial recognition to enrich data journalism and enhance aesthetic appeal, transforming 原本的“flat” content into more “three-dimensional” experiences. This not only satisfies visual perception but also increases auditory and tactile immersion, strengthening news presence and dissemination power while reshaping reading experiences.

Incorporating feedback channels during the visualization process enhances the value of reader feedback. On one hand, communication between people and information allows reader comments to help news producers view content and information dissemination more objectively. On the other hand, interactions through likes and replies facilitate more active information flow and enhance effective information dissemination. Meanwhile, collision of viewpoints helps readers eliminate information uncertainty and understand news more rationally.

3.3 Improve Data Literacy Among Media Practitioners

3.3.1 Collaborate with Universities to Develop Interdisciplinary Courses and Reserve Talent Currently, majors in journalism and communication, computer technology, and art design are common in universities, but the field of data journalism visualization still faces certain shortcomings. Integrating the needs of traditional media with university education systems by adding interdisciplinary courses can promote the cultivation of multi-disciplinary talent and provide human resources for data journalism visualization. In curriculum design, horizontally integrating journalism and communication with economics, computer science, law, and art design not only makes full use of excellent university educational resources but also cultivates comprehensively developed talent. Combined with media practice, we should doubly emphasize cultivating students' data sensitivity, improving their ability to distinguish data authenticity, and fully equipping them with data mining, analysis, and visualization technologies to explore deep meanings behind data, laying a solid foundation for future data journalism work.

3.3.2 Transform Thinking Patterns and Introduce Professional Talent

The original intention of visualization is to intuitively present information and trends behind complex data. It should not pursue visualization for visualization's sake, as overly flashy designs and decorations can obscure data journalism content. Therefore, information graphic design should prioritize simplicity and clarity, clarifying data information and highlighting key points. Additionally,

news practitioners must dare to break free from mental constraints and embrace fresher approaches, such as having machines participate in routine tasks to liberate themselves from basic, heavy workloads and engage in more intelligent and creative work.

Emphasizing data journalism talent cultivation requires hiring professionals sensitive to both news and data, and establishing specialized data journalism visualization production teams. News organizations need to set up corresponding position training courses to enhance practitioners' professional qualities, help them quickly understand the current era's background and media environment, and familiarize them with data visualization tools and production processes, thereby improving data screening, analysis, and mining capabilities. This enables them to form their own set of insights, cognition, and behavioral logic to better provide information services with richer content to audiences.

3.4 Deeply Mine Data Value

3.4.1 Enhance Topic Selection Value and Conduct Multi-Dimensional Cross-Comparison Topic selection represents the “gatekeeper” role in communication studies. Through scientifically effective technologies to analyze and interpret data from a holistic perspective, we can increase the importance of topics, discover intrinsic connections between data, events, and people, help readers identify connections between events and themselves and understand impacts on their lives, and find their own connection points with news, thereby enhancing reader proximity to news.

When analyzing or interpreting events, collecting information from different dimensions and revealing connections behind data through data overlay and comparison can piece together the full picture of events as much as possible. Analyzing data from different dimensions can uncover deep underlying relationships between seemingly loosely related datasets.

3.4.2 Promote Data Crowdsourcing Models with Media and User Participation Relying on the power of social groups to find fragments of news event occurrence, development, and outcomes can stimulate public enthusiasm for participating in news reporting, producing a “1+1>2” effect that drives innovative development in data journalism production. Relying solely on individual journalists may yield unsatisfactory results and cannot satisfy readers' thirst for news information and their desire for involvement. Adopting crowdsourcing methods to collect data—inviting users to actively participate and share, uploading personally discovered data, and obtaining information through open real-time comment platforms and electronic questionnaires—both ensures data sources and reduces data collection costs. Therefore, data crowdsourcing represents one viable path for data journalism visualization exploration and development.

4. Development Trends in Data Journalism Visualization

As intelligent devices proliferate and data generation channels diversify, news presentation methods are becoming increasingly varied. Data analysis and news visualization will undoubtedly become mainstream trends in the media industry's future development. In particular, the popularization of mobile intelligent terminals will drive the continuous development and maturation of data journalism, with mobile data journalism works becoming a normal form of news reporting. The maturation of big data, artificial intelligence, cloud computing, and other technologies may subvert traditional media production methods and provide more convenient production tools for data journalism visualization.

Data analysis of people's behavior is becoming increasingly accurate, making it possible to tailor action plans for readers. Based on big data, producing universal data news, then user-centrally integrating customized information visualization into users' daily tasks and designing information visualization environments can promote user collaboration and sharing. This represents not only the secondary development and utilization of data but also an important means of integrating into the new media discourse field and consolidating and expanding news influence.

Data journalism production is a comprehensive process requiring multifaceted talent to collaborate effectively. Whether news editors or design workers, the cultivation and enhancement of data processing capabilities are urgently needed. In short, exploration of visualization journalism has just begun. With economic development and continuously improving public media literacy, data journalism can be expected to have even broader development prospects in the future.

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

Source: ChinaXiv — Machine translation. Verify with original.