Cultivating a Data-Driven Mindset and Constructing a Big Data System for Press and Publishing: Postprint

Authors: Wu Cuijiao, Zhang Yu

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

Abstract

In fact, as early as the 1980s, the American futurist Alvin Toffler highly praised the concept of "big data" in his work *The Third Wave*, calling it "the magnificent movement in the Third Wave" and thus prematurely demonstrating to the world the inherent potential of big data. Entering the 21st century, big data technology has been widely applied globally, becoming the cornerstone of the information technology industry, while establishing a data-driven mindset and building industry-wide big data systems have become top priorities for the future of the industry. This paper primarily takes the news publishing industry as an example to explore specific approaches and implementation plans for establishing a big data mindset and building an industry-wide big data system within this sector, and offers relevant development recommendations.

Full Text

Preamble

Abstract: As early as the 1980s, American futurist Alvin Toffler praised the concept of "big data" in his book *The Third Wave*, calling it "the magnificent movement in the third wave" and demonstrating its inherent potential to the world. In the 21st century, big data technology has been widely adopted globally, becoming a treasure of the information technology industry. Consequently, cultivating data thinking and constructing industry-specific big data systems have become top priorities for future development. This paper uses the news publishing industry as a case study to explore specific approaches and construction plans for fostering big data thinking and building industry-wide big data systems, offering relevant development recommendations.

Keywords: big data system; data thinking; news publishing industry; construction approach; plan; recommendation

Classification Code: G641
Document Code: A
Article ID: 1671-0134 (2019) 12-066-03
DOI: 10.19483/j.cnki.11-4653/n.2019.12.018
Citation Format: Wu Cuijiao, Zhang Yu. Cultivating Data Thinking and Constructing a Big Data System for News Publishing [J]. *China Media Technology*, 2019(12): 66-68.

Authors: Wu Cuijiao, Zhang Yu

The Rise of Big Data as a National Strategy

In 2012, the White House Office of Science and Technology Policy first released the "Big Data Research and Development Initiative," establishing a high-level steering group dedicated to big data and elevating big data strategy to a national development priority. China followed suit, introducing the concept of "big data" in its 2014 Government Work Report. The 2015 "Several Opinions on Using Big Data to Strengthen Services and Supervision of Market Entities" clarified three key objectives: implementing government data resource policies, constructing public service big data projects, and optimizing critical big data technologies and related products. Over the past four years, China has launched numerous big data engineering projects and security initiatives. The country has now introduced the "Internet Plus Big Data" action plan, aiming to implement its cyberpower strategy and usher in a new chapter of big data construction for the new economic era.

1. Innovative Data Thinking in Big Data Systems

In the 21st century—a time of information deluge—the emergence of big data technology systems has disrupted development processes and transformed the rhythm of social change. Big data systems built around this technology aim to revolutionize human thinking about future development, a transformation particularly pronounced in the socially interactive news publishing industry. Today' s news publishing professionals must update their worldviews, deeply engage with their social activities and positions, acquire and utilize comprehensive big data information, integrate production and consumption processes, highlight the value of big data applications, continuously innovate data thinking within the framework of big data systems, and guide social production and life.

1.1 Three Innovative Data Thinking Approaches for the News Publishing Industry

A popular adage states, "Except for God, everyone must believe in and apply data, letting data speak." This principle appears in China' s "Big Data Development Action Outline," which advocates "using big data to speak, make decisions, manage, and innovate," establishing a complete, feature-rich big data system centered on big data technology. The core of this management system

naturally revolves around big data, including the data thinking and data culture it embodies. In the news publishing industry, big data must be elevated to the level of culture because it pursues rational thinking and scientific spirit, ensuring news reporting seeks truth from facts, correctly guides public opinion, and uses data to prove facts—the most fundamental principle of the industry. Austrian big data expert Viktor Mayer-Schönberger outlined three insightful and forward-thinking innovative approaches in his book *Big Data*: holistic thinking, diverse thinking, and correlational thinking. For the news publishing industry, however, these can be understood as value thinking, application thinking, and sharing thinking.

1.2 Value Thinking

For the news publishing industry, big data represents an intangible asset, and the sector has long adhered to the principle that data equals wealth. Value thinking constitutes the core of big data. In news publishing houses, this value thinking encompasses data management, data transactions, transaction behaviors, and user data. Traditional news publishers must recognize that users represent an invisible wealth that cannot be ignored. As early as 2014, when Facebook announced its $16 billion acquisition of WhatsApp, it proved this point: exchanging tangible assets for massive intangible assets (users) is worthwhile, representing Facebook' s path to success.

1.3 Application Thinking

In the 21st century, some claim that those who possess big data own the future, though this assertion merits discussion. What is clear is that big data helps enterprises occupy commanding heights and seize initiative, demonstrating that data' s value lies in its application rather than mere possession. The news publishing industry requires extensive social communication and must use data to drive all internal business processes and product development. Therefore, the industry needs to leverage data to drive operations and generate value through operations—one reason traditional news publishers seek transformation. Compared to production processes that provide fixed quantitative thinking, data offers fluid qualitative development thinking. Data support is essential for decision-making, print-run decisions, and topic selection planning. By building a comprehensive big data platform system, constructing a full-dimensional recording system, and optimizing data relationship architecture, publishers can deeply engage in data correlation and mining technologies, implement predictive and recommendation processes, effectively address defects in traditional news publishing, provide data-driven services for news products and content, solve persistent pain points, and ultimately achieve datafication of news publishing products and services.

1.4 Sharing Thinking

Many news publishers currently suffer from severe data silos caused by inability to share information among publishers, editorial offices, and authors, potentially trapping publishers in isolation crises. Additionally, traditional news publishers lack big data platform support, preventing industry collaboration and data sharing during business development. Based on the concept that "sharing is for use, not possession," weak internal data liquidity and inability to integrate with external data ultimately render data dead. Publishers must address two issues: first, reasonably balancing data sharing relationships and optimizing sharing efficiency; second, establishing sharing platforms based on interest-balancing mechanisms—key to traditional news publishers' future transformation and development path.

2. Real-World Development Challenges Facing the News Publishing Industry

Traditional news publishing faces numerous developmental problems that seriously hinder healthy industry progress. First, severe overcapacity and supply-demand mismatches objectively and significantly impact development, creating substantial risks and prompting the industry to pursue green development prospects. Since the big data concept emerged in 2014, traditional news publishers have yet to change their development patterns, with national news publication sales reaching only 78 billion yuan, inevitably creating an inverted development trend. Simultaneously, the "warlords rising" phenomenon in distribution markets has led to a severe lack of unified national distribution, making a unified, open, and orderly competitive national market an elusive goal that prevents China's publishing and media groups from growing stronger.

Furthermore, the publishing and distribution ecosystem currently exhibits severe 畸形 (distorted) development. Its consignment system leads to unconditional returns, long payment cycles, difficult fund recovery, and loss of credibility, complicating single-title accounting. Objectively speaking, the industry pursues extensive management models. Failure to fundamentally address these constraints may render big data technology systems, industry big data platforms, and credit information platforms passive. It is therefore necessary to implement comprehensive precision management across the entire industrial chain—from upstream to downstream—ensuring data-driven topic selection planning, providing printing decision content, deeply and precisely analyzing user and marketing behaviors, and building personalized development mechanisms to create unlimited possibilities for reshaping production, sales, and management models. News publishing enterprises must accurately grasp this point during transformation.

3.1 Big Data Content in the News Publishing Industry

Big data is key to the rapid development of the news publishing industry, meeting information-era development standards. However, applications of big data

and related technologies in news publishing remain subject to diverse opinions. Current big data content in publishing houses can be classified in numerous ways, primarily by business level. News publishers' big data content includes six categories: First, institutional data, including news publishing government agencies, industry associations, public institutions, publishing houses, newspaper offices, distribution groups, and corporate publishing revenue and marketing profits. Second, personnel data, including civil servants, editors, journalists, experts, and basic data for all industry practitioners. Third, product data, including books, newspapers, audio-visual materials, online publications, as well as author metadata, printing data, and copyright data. Fourth, government affairs data, including comprehensive office data, market supervision data, public service process data, personnel information, and financial data. Fifth, business data, including market transaction data, contract data, and copyright data from news publishing enterprises. Sixth, content data, including information and knowledge data, which comprises four types: factual knowledge, skill knowledge, principle knowledge, and human knowledge, forming a rich big database platform considered a unique data treasure and important foundation for knowledge services.

3.2.1 Data Collection

Data collection is the most critical factor in production transformation for the news publishing industry, aiming to construct new data development formats by integrating content data, market data, and user data while emphasizing the interactive nature of big data technology applications. Throughout this process, it focuses on building valuable and timely information development systems around core publishing business content, integrating data collection and feedback data from publishing institutions' content creation into actual production processes. Currently, China has numerous authoritative professional news publishers with rich specialized content resources that increasingly pursue resource-oriented and fragmented data content development during business transformation and upgrading. They emphasize using big data to provide abundant resources, ensuring effective industry data connection and adjustment, solving certain "data island" effects within enterprises, seeking rapid industry development, strengthening industry data development standards, and establishing targeted big data service systems. Additionally, meeting unified standardized data collection requirements for the news publishing industry and optimizing big data collection platforms are essential. Constructing a comprehensive data resource system based on big data analysis and application is necessary to enhance the timeliness and accuracy of big data applications.

3.2.2 Data Indexing

News publishers can collect massive amounts of data with big data technology support and then implement indexing work on the obtained data. Throughout this process, it is necessary to analyze and apply the predictive and early warn-

ing functions within the big data system, enhancing the innovative key role of data. Considering the current development status of China' s entire publishing industry, enterprises should more extensively utilize data indexing technology to effectively index knowledge and industry application content, establishing a knowledge indexing discipline research system that can lay a solid foundation for subsequent development of news publishers' knowledge services.

3.2.3 Data Computation

News publishers' data computation work includes cloud computing and statistical analysis, which can be implemented compatibly and simultaneously. First, analyzing cloud computing reveals that it contains massive amounts of big data that can combine with indexed data to optimize cloud computing processing and establish a secondary data research system based on processing results. Here, both cloud computing technology and data computation processes provide rich and diversified dynamic computing content, enabling technology expansion and virtualization, ensuring comprehensive adjustment of virtualized information resources that are truly incorporated into network links to satisfy unified resource management and intelligent scheduling. Objectively speaking, this constructs a brand-new news publisher data resource pool that continuously provides different resource services to certain users in need.

Regarding statistical analysis, it primarily combines the rapid development of computer technology to collect and process statistical data while meeting data analysis and storage development needs. Considering that general computer systems cannot effectively determine the objective complex patterns reflected by data, it is necessary to propose decision-making solutions when judging future development trends, conduct secondary indexing and computation of data content, and meet statistical analysis conditions. During this process, it is also necessary to quantitatively analyze data uncertainty, extract data content based on information patterns, and find optimal solutions for data indexing and statistical computation.

3.2.4 Data Modeling

In data modeling, it is necessary to analyze disciplinary system modeling and industry application modeling according to different application scopes. In this process, disciplinary system modeling primarily relies on existing disciplinary systems with mature theoretical knowledge systems that can build internal models based on relevant relationships and select rational data modeling solutions. Currently, big data modeling structures in the news publishing industry are quite mature and can be used for industry development application modeling. Their modeling scope covers various industry development fields and can expand data modeling scope based on current industry occupational foundation status, which also means expanding the business scope of news publishing enterprises. Moreover, data modeling is personalized and can construct corresponding big

data models for users' specific needs, primarily targeting data modeling development fields.

3.2.5 Data Service

In data service, it primarily combines data collection, indexing, statistics, and modeling to design data service systems, striving to provide rich and colorful big data services for all target customers while fully considering effective transformation and optimization of audit development content in the news publishing industry. Specifically, it means reasonably utilizing big data technology to comprehensively promote transformation from publishing to knowledge service dissemination for publishing industry institutions, such as transformation from artificial intelligence to development mechanisms, and from traditional business categories to big data-driven directions, comprehensively enhancing the overall business operation and development management efficiency of news publishing institutions.

Currently, news publishers should pursue using big data technology to optimize personalized knowledge service content, integrating data resources from different channels to ensure effective optimization of knowledge service platforms under digital sorting background, establishing user-centered retrieval and reading platforms, hotspot analysis platforms, data mining analysis platforms, and knowledge association service platforms. Meanwhile, they should construct knowledge graphs to provide users with digital, knowledge-based, and intelligent service projects. Today' s news publishers pursue digital publishing business development, so implementing personalized user retrieval and browsing terminal behavior tracking based on big data is understandable, as these approaches can precisely grasp current user demand patterns and customize personalized solutions for users' future knowledge needs.

Future big data model standardization construction and application should become the foundational key to current news publishing industry development. For example, China' s National Standardization Committee has established the "National Information Technology Standardization Committee Big Data Standards Working Group" to formulate a complete national big data field standard system, enabling news publishing to combine big data technology processes to collect multi-faceted resource data content and achieve effective information data docking. Therefore, in the future, the news publishing industry should also establish information-based business platforms based on big data sharing and exchange mechanisms, accelerate construction of major big data projects, and lay foundations for news publishers' big data system construction and business innovation optimization.

References

[1] Liu Chengyong. Cultivating Data Thinking and Constructing a Big Data System for News Publishing [J]. *Publishing Reference*, 2016(7): 4-8.

[2] Zeng Wen, Xu Hongjiao, Che Yao, et al. Research on Topic Selection Decision Analysis Models Based on Big Data in the Book Publishing Industry [J]. *Journal of Intelligence*, 2018, 37(8):
[3] Tian Jun. Research on New Business Forms of Big Data Application in the News Publishing Industry [J]. *Communication and Copyright*, 2017(8): 47-49.
[4] Wang Xinxin. Analysis of Dilemmas in Big Data Application in China' s News Publishing Industry [J]. *Network Security Technology and Application*, 2016(10): 120-121.
[5] Shi Jialiang, Wu Ding. Analysis of Obstacles to Big Data Application in China' s News Publishing Industry [J]. *China Publishing*, 2015(23): 29-31.
[6] Zhang Xinxin. Reflections and Prospects on Big Data Application in the News Publishing Industry [J]. *Technology and Publishing*, 2016(1): 3-8.
[7] Anonymous. Establishment of News Publishing Big Data User Behavior Tracking and Analysis Laboratory [J]. *News World*, 2017(4).
[8] Wang Tingting. Reflections on Transformation of Traditional Publishing Industry in the Big Data Era [J]. *Business Economy*, 2017(4).

(Author Affiliation: Peking Union Medical College Press)

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