Economics & Business Management
ISSN Print: 3079-5214
ISSN Online: 3079-5222
https://doi.org/10.63313/EBM.9160

The Current Status, Problems, and Countermeasures of the Publishing Industry's Transition Toward Data-Driven Development

Chen Liu
University of Shanghai for Science and Technology, Shanghai 200093, China
Email: 242602896@st.usst.edu.cn

How to cite this paper: Liu, C. (2026). The Current Status, Problems, and Countermeasures of the Publishing Industry's Transition Toward Data-Driven Development. *Economics & Business Management*, 5(1), 57-66. ISSN Print: 3079-5214; ISSN Online: 3079-5222.
https://doi.org/10.63313/EBM.9160
Published: 2026-03-10

Copyright © 2026 by author(s) and Erytis Publishing Limited.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/



Abstract

In the context of the emerging era of digital intelligence, the high-quality development and transformation of the publishing industry depend not only on digital technological innovation but also on the central role of data. As a critical factor of production, data plays an indispensable role in promoting the transformation of the publishing industry toward a data-driven paradigm. Within this context, publishing forms are gradually evolving from traditional print-based publishing to data-centered data publishing. This transformation enables the publishing industry to achieve multidimensional improvements in the process of digital development. Based on the contemporary technological and social context, this paper examines the current status of the publishing industry's transition toward data-driven development, analyzes the problems encountered in the transformation of publishing business models, and proposes policy recommendations and strategic pathways for the future development of data publishing. Based on literature analysis and an examination of industry practices, this paper analyzes the current status, identifies key obstacles including technological bottlenecks, copyright dilemmas, and international expansion challenges, and proposes development pathways such as building a new data circulation ecosystem, establishing a compliance governance system, and reshaping narrative systems to promote international markets.

Keywords

Publishing; Data Elements; Data Publishing; Data-Driven; Generative Artificial Intelligence

1. Current Status of the Publishing Industry's Transition Toward Data-Driven Development

1.1. Policy Support and Infrastructure Development

With the continuous iteration of digital technologies, data elements have increasingly become a key factor of production, and their driving role in the

transformation of the publishing industry has become more prominent. China was among the first countries in the world to formally establish data as a factor of production. On October 25, 2023, the National Data Administration was officially established in Beijing, marking a new stage in the application of data elements. On December 31 of the same year, in order to implement the decisions and deployments of the CPC Central Committee and the State Council, the National Data Administration released the Three-Year Action Plan for "Data Elements ×" (2024–2026), which clearly proposes promoting high-quality economic development through the integrated application of data. At the same time, according to the arrangements of the Central Economic Work Conference, the deep integration of digital technology and the real economy has continued to advance. The introduction of the Implementation Plan for Promoting Common Prosperity through the Digital Economy further strengthens, at the strategic level, the core role of data elements in promoting balanced and adequate development.

As the industry-level implementation of these national strategies, China has issued the Development Plan for the Publishing Industry during the 14th Five-Year Plan Period in order to promote the high-quality development of the publishing sector. The plan focuses on improving the supply of high-quality content and stimulating innovation vitality in publishing, thereby further building China into a leading publishing power. Transforming raw data into data elements and subsequently into data productivity has become a key pathway for promoting the development of new quality productive forces in the publishing industry. Through collaborative optimization, data generates a "multiplication and division effect," demonstrating the powerful enabling role of data elements in new quality productive forces and becoming an important driving force for the intelligent digital transformation of society [1].

The full realization of the value of data elements relies on strong infrastructure support. China has established twenty National Scientific Data Centers in different fields to carry out work related to scientific data. With the continuous improvement of policies and infrastructure related to data, the hub nodes of the "Eastern Data, Western Computing" project have gradually advanced, and the overall national computing capacity has been continuously enhanced. The construction of data centers not only drives industrial investment but also promotes data connectivity between eastern and western regions, thereby maximizing the value of data resources. Driven by both top-level policy design and infrastructure development, the digital transformation of the publishing industry has gained strong policy support as well as a solid computing foundation.

1.2. Evolution of Publishing Forms: From Digital Publishing to Data Publishing

Data publishing is a new form of integrated publishing centered on data. It refers to

taking data as the key element and treating text, images, audio-visual materials, games, and animation as different forms of data representation. Publishing activities are organized around data mining, collection, indexing, storage, and computation. Through the construction of data models, publishing eventually develops toward the application of publishing data and the provision of publishing data services [2]. Some scholars also argue that data publishing refers specifically to the publication of scientific data, which can enhance multiple aspects of scientific research and further improve the quality of research data [3]. Data publishing not only emphasizes the importance of data itself but also focuses on data insights and user needs, thereby providing users with more tailored solutions and better services [4]. The development of publishing has generally gone through three stages: the digitization of carriers, the service-orientation of data, and the construction of an intelligent ecosystem. Technological iteration and policy support jointly drive the publishing industry to shift from "information transmission" to "knowledge empowerment."

Data elements possess unique attributes that distinguish them from traditional factors of production, including non-rivalry, self-reinforcement, and scenario dependence.

Non-rivalry. Data can be repeatedly used by multiple entities, and the cost of replication is extremely low. This not only results in extremely high utilization efficiency but also allows unlimited frequency of use. Compared with traditional production factors, it significantly reduces cost consumption and can generate benefits across multiple fields. For example, Beijing provides citizens with a public data platform covering areas such as the economy, tourism, healthcare, meteorology, and food, enabling the public to consult and use data without limitation.

Self-reinforcement. In enhanced publishing, whether it is the association of knowledge, the expansion of content presentation, or the selection of dissemination channels, all rely on data as the fundamental support [5]. According to Metcalfe's Law, network value is proportional to the number of connections, and this relationship achieves its greatest effect in peer-to-peer connected networks. Taking the Zhihu platform as an example, users' original behavioral data such as questions, answers, and likes are processed through algorithms to achieve precise content recommendations. At the same time, newly generated user behavior data feed back into the algorithmic model, forming a virtuous cycle in which the more data are used, the more valuable they become.

Scenario dependence. In the digital economy, whether data can function as a factor of production is measured by whether it possesses value. Therefore, the value of data elements is particularly prominent and important [6]. The release of the value of data elements depends on specific scenarios. The higher the compatibility between data elements and their application scenarios, the greater the value that can be generated. For example, the Chinese Classics Ancient Books Database of Zhonghua Book Company digitizes the semantics of ancient texts in depth,

improving the efficiency of researchers conducting academic studies and promoting scientific progress. If the same data are applied in public education to create more interactive experiences, their value can be further extended in the cultural field. This demonstrates that the same data asset may generate very different values in different scenarios.

1.3. Exploration of Data Publishing Applications at Home and Abroad

Based on the unique attributes of data elements, diversified application models of data publishing have emerged worldwide. Through the analysis of typical domestic and international cases, these practices can generally be summarized into three categories.

Academic platforms: promoting the open sharing of research data. Such platforms focus on facilitating the circulation and reuse of scientific data. In China, a relatively mature data repository is ScienceDB, a data knowledge base built by the Computer Network Information Center of the Chinese Academy of Sciences. It provides data services for researchers, academic journals, and research institutions, and the datasets are openly shared [7]. Internationally, platforms such as Zenodo, an open-access data repository operated by research institutions—provide global researchers with services for data storage, sharing, and access, greatly promoting interdisciplinary research and rediscovery.

Commercial services: providing high-value data decision support. These services take data as the core and provide users with in-depth intelligence services. For example, the Wind Financial Terminal integrates global capital market data and provides professional analytical models and tools, offering indispensable decision-making support for most financial institutions in China. Internationally, the Bloomberg Terminal, as a benchmark in global financial data services, provides real-time financial data analysis and subscription services.

Public cultural platforms: constructing new forms of cultural data services. This type is oriented toward the public sphere and aims to transform cultural and public knowledge into widely usable data assets. In China, a representative example is the Cihai Knowledge Service Platform developed by Shanghai Century Publishing Group, which digitizes and semantically links knowledge across disciplines to form a convenient public knowledge inquiry platform. Internationally, government open data platforms such as data.gov stimulate social innovation by releasing public data.

2. Problems Faced: Major Obstacles to Data-Driven Transformation

2.1. Technological Bottlenecks: The Proliferation of Data Silos and Barriers to End-to-End Integration

Data, as the core element of data publishing and the foundation for artificial intelligence applications, plays a crucial role in determining the quality and

effectiveness of AI development. At present, China's publishing industry has actively begun to employ artificial intelligence technologies to support the upgrading and development of the entire publishing workflow. However, several challenges persist, including inconsistent data standardization, the persistence of "data silos" caused by legacy systems, and the shortage of data-related professionals.The persistence of "data silos" formed by legacy systems, and insufficient application of data technologies due to a shortage of data-related professionals.

Although technologies such as AIGC have been introduced in the publishing industry to assist content creation, the depth of their integration remains limited. In the context of the deep integration and development of the publishing industry, publishing institutions have continued to accelerate the digital conversion of existing print products and the development of emerging digital products. However, most publishing institutions still tend to treat data elements merely as a form of publication [8]. The exploitation of data value largely remains confined to improving efficiency in isolated stages rather than extending across the entire publishing industry chain. Behind this situation lie not only insufficient technological investment but also structural issues, particularly the lack of interdisciplinary professionals in the field of data publishing.

2.2. Copyright Dilemma: Uncertainties Surrounding Data Ownership and the Compliance of AIGC-Generated Content

In the process of generating content for data publishing, generative artificial intelligence plays an indispensable role. The training of large-scale models can assist with basic textual tasks, idea generation, and other preliminary work. However, when AIGC systems collect and process web-based content data, the ownership of copyright becomes difficult to trace, and the authenticity and reliability of the generated content are often difficult to verify. Copyright law emphasizes the protection of expression, whereas model training essentially involves the extraction of non-expressive elements such as ideas and stylistic patterns. This form of non-expressive use conflicts with the "idea–expression dichotomy" in copyright law, resulting in a lack of clear and operational standards for determining fair use [9].

The development of artificial intelligence has brought significant transformations to social resources and has profoundly reshaped the role of the content industry. AI has fundamentally transformed data into a form of resource [10]. As an essential source for generative artificial intelligence, content has become an indispensable component in this process. Nevertheless, issues such as the tracking of data usage and the distribution of benefits remain unresolved. It is therefore necessary to establish new governance mechanisms to better regulate data-related intellectual property rights and to provide institutional safeguards for data ownership, collection, utilization, and transactions. Such measures will further stimulate the market vitality and innovative potential of the content industry within the

framework of data publishing.

Beyond copyright issues related to content production, attention should also be paid to the underlying logic of content generation. Although AIGC, as an artificial intelligence technology, may provide relatively neutral data drawn from online sources, its operational logic ultimately depends on human design and configuration. The potential bias and orientation embedded in model-generated language therefore warrant careful consideration. The values conveyed through content data are often transmitted in subtle ways. Only through more standardized and responsible governance of language models can the long-term and sustainable development of AIGC be ensured.

2.3. Challenges of international expansion: cultural differences and technical standards.

Data publishing faces multiple barriers in its international expansion, including those related to technology, intellectual property, and culture. First, as data has become a core element of publishing, many countries regard it as a key driving force in the development of the publishing industry. However, data standards have not yet been unified across countries, making it difficult to establish a shared platform for data publishing. In addition, digital technologies also affect the construction and operation of such platforms. Overseas markets place high demands on user experience and platform stability, and some domestic data publishing models are not yet fully capable of adapting to these requirements.

Second, significant differences exist among countries in terms of legal regulations and intellectual property protection concerning the copyright of data publications. When data publications are disseminated in overseas markets, disputes related to piracy and copyright infringement may arise. Due to the cross-border nature of these disputes, rights protection becomes more complicated and the costs of safeguarding intellectual property rights are considerably higher.

Furthermore, data publishing involves multiple forms and modes of information dissemination. It is not only a means of transmitting information but also an important medium for cultural expression. If the cultural meanings embedded in data cannot be effectively communicated through appropriate narrative strategies and aligned with the cultural context of the target market, the phenomenon of “cultural discount” may occur, thereby hindering the effective international dissemination and localization of data publications.

In order to promote the sustainable development and global dissemination of data publishing in the current era of digital intelligence, this study analyzes the issue from three dimensions—technology, institutions, and communication—and proposes development pathways that may facilitate the publishing industry's transition toward a data-driven paradigm.

3. Discussion and Policy Implications: Development Pathways for Promoting Data-Driven Publishing

Data has become a key driver of development in the contemporary publishing industry, accelerating its digital transformation and enabling traditional print-based publishing to evolve toward data publishing. This transformation has also given rise to new cultural forms within the publishing sector. However, data publishing still faces several challenges, including incomplete institutional frameworks, insufficient market adoption, and limited adaptability in practical application scenarios. These issues constrain the further commercial development of data publishing. Against this background, how to effectively utilize data elements and fully realize their value has become a critical issue for the future development of the publishing industry.

3.1. Overcoming Technological Barriers: Building a New Ecosystem for Data Circulation

Supported by digital technologies, data publishing enables more efficient processing and utilization of publishing data. Publishing institutions can draw upon the data governance practices of leading international publishing groups and adapt these models according to their own organizational characteristics. For instance, Elsevier has made substantial investments in foundational data infrastructure, building standardized datasets that support AI model training and enable users to conduct data analysis, inform decision-making, and address complex research and business problems [11].

In the context of rapid digital technological development, it is essential to further strengthen the application of artificial intelligence in the publishing industry. Publishing enterprises should actively integrate data elements into their operations, combine the traditional strengths of publishing with emerging forms of data publishing, and establish unified industry data standards. Promoting data circulation among publishing institutions and cultivating high-level data professionals are also critical steps toward improving the efficiency and security of data utilization. By leveraging large-scale datasets and connecting multiple sectors of digital publishing—including books, digital games, and video media—the publishing industry can extend its value chain and promote the large-scale development of data publishing while gradually eliminating the long-standing problem of data silos.

3.2. Improving Institutional Frameworks: Establishing a Compliance Governance System for Data and AIGC

The rapid development of data publishing and the emergence of new publishing business models have elevated data to a central production factor in the content industry. Through effective data utilization, publishing enterprises can improve managerial decision-making, reduce resource consumption, and enhance

operational efficiency. Data-driven management also strengthens traceability in business operations and facilitates the optimized allocation of internal resources [12]. However, the sustainable development of data publishing requires robust monitoring and governance mechanisms to ensure that emerging issues are addressed in a timely manner.

Encouraging innovation and broader participation from multiple stakeholders is therefore essential for the continued development of data publishing. In terms of regulatory governance, market supervision should not rely solely on a single authority but should gradually move toward a more open and adaptive monitoring framework. Such a framework would allow market actors greater flexibility while enabling regulatory institutions to adjust policies according to evolving market conditions [13].

At present, the reform of market-oriented allocation of data elements has entered a critical stage, and the institutional mechanisms related to data publishing still require further improvement. A key challenge lies in balancing data security with data-driven innovation and development [14]. To address this issue, it is necessary to gradually establish a data property-rights registration system to facilitate the traceability of data ownership. Central mechanisms in this process include data classification and graded ownership confirmation, as well as a mechanism for the separate operation of data-related property rights[15]. In addition, clarifying the roles and responsibilities of different actors within the data publishing ecosystem and establishing a hierarchical governance framework will help support the future testing and evaluation of regulatory policies.

3.3. Reshaping Narrative Systems: Expanding International Markets through Platform-Based and Localization Strategies

In today's world, where globalization and digitalization are increasingly intertwined, cultural outreach has become an important indicator of a country's soft power and industrial competitiveness. At the Belt and Road Forum, China proposed the construction of the Digital Silk Road, aiming to promote international cooperation in building data infrastructure and to strengthen cultural exchange among participating countries. In this context, promoting Chinese culture through data publishing and effectively "telling China's stories" requires close alignment with the cultural contexts and practical realities of target markets in order to achieve successful international dissemination.

To address the challenges of entering overseas markets, publishing enterprises should ensure that their data formats, interfaces, and technical standards are compatible with widely used international standards, thereby enabling integration with multiple global platforms. In addition, collaboration with leading international academic publishing groups—such as Elsevier—can facilitate the integration of high-quality data content into global distribution networks and knowledge

ecosystems. Equally important is the localization of cultural narratives. By employing technologies such as data visualization and Artificial Intelligence Generated Content (AIGC), the cultural connotations embedded in Chinese data resources can be transformed into market-oriented "data stories," helping to bridge cultural gaps in international communication.

4. Conclusion

Furthermore, strengthening international cooperation is essential for the development of data publishing. Countries should promote the establishment of unified cross-border data flow agreements and create specialized data governance organizations to address cross-border disputes and regulatory challenges related to data publishing. Through enhanced international coordination, a more stable and transparent institutional environment can be created for the cross-border circulation and utilization of data resources, thereby supporting the sustainable development of the global data publishing ecosystem.

In summary, as a critical factor of production in the era of digital intelligence, data elements should be fully integrated into the transformation of the publishing industry toward a data-driven paradigm. This transformation requires expanding the market for data elements, enriching digital application scenarios, strengthening the cultivation of digital professionals, and promoting deeper integration between digital technologies and digital industries. These measures will contribute to improving the efficiency of data utilization and enhancing the overall innovation capacity of the publishing industry in the digital age.

At the same time, it is essential to improve market regulation and copyright protection mechanisms for data publishing, while recognizing the important role of Artificial Intelligence Generated Content (AIGC) as a key driver of digital publishing. By leveraging technological capabilities and market advantages, data publishing can foster new industries, new models, and new business formats. Such developments will further reshape the traditional value chain of the publishing sector and accelerate the transition toward a more intelligent and service-oriented publishing ecosystem. As a result, the empowering value of data within the data publishing market will become increasingly evident and will serve as a significant driving force for high-quality development.

References

[1] Wang, X.Q. (2025). Legal application dilemmas and solutions of the "three new elements" of new quality productive forces empowered by data elements. Science & Technology Progress and Policy, 42(7), 21–32.

[2] Zhang, X.X. (2025). "Data Elements × Publishing": From digital publishing to data publishing. Document, Information & Knowledge, 1–10.

[3] Huang, G.B., Qu, Y.J., & Wen, D.Q. (2024). Conceptual differentiation between data publishing and digital publishing. Publishing Journal, 20, 24–31.

[4] Zhou, R.T. (2024). Data publishing in the Web 3.0 era: Constructing a new ecosystem of

co-creation and collaboration. Editors' Friend, (7), 38–44.

[5] Hua, N., & Dong, W.J. (2025). Data-driven enhanced publishing strategies and implications: A case study of 12 high-impact chemistry journals. Chinese Journal of Scientific and Technical Periodicals, 36(2), 192–198.

[6] Yang, J., & Tao, A.T. (2024). Exploring the value of publishing data elements in the era of digital intelligence. Chinese Editors Journal, (9), 28–33.

[7] Xu, J., Tang, W.H., & Xia, X.Y. (2020). Analysis of the current status of data publishing and countermeasure discussion from a practical perspective. Chinese Journal of Scientific and Technical Periodicals, 31(11), 1331–1337.

[8] Cong, T., & Xi, X.Y. (2024). Research on data element-driven innovation in publishing services: Based on the "Data Elements × Three-Year Action Plan (2024–2026)". Chinese Editors Journal, (8), 82–88.

[9] Xing, Y.H. (2025). Challenges and optimization paths of publishing data governance in model training scenarios. Publishing Research, (8), 13–20.

[10] Deng, X.L., & Bu, L.X. (2024). Humanistic constraints in the technological wave: The transformation of the publishing industry's role under generative AI. China Publishing Journal, (16), 26–33.

[11] Zhang, X.F. (2025). Research on data technology empowering AI product development: A case study of the Elsevier academic publishing group. Publishing Journal, 33(4), 83–93. https://doi.org/10.13363/j.publishingjournal.2025.04.010

[12] Yang, J. (2025). Data elements empowering new quality productive forces in the publishing industry. Publishing Journal, 33(4), 74–82.

[13] Chen, M. (2024). Data publishing in the context of open science: Functions, dilemmas, and governance paths. Science-Technology & Publication, (10), 80–88.

[14] National Data Administration. (2025). Digital China Development Report (2024). Beijing: National Data Administration.

[15] Cai, F., & Lei, Y. (2024). Publishing data as a factor of production: Origin, value and breakthrough paths. China Publishing Journal, (22), 41–46.