

# A Preliminary Exploration of the Application Value of Big Data for Digital Publishing (Post-print)

**Authors:** Zhao Hongyuan

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

## Abstract

The emergence of big data technology has produced profound impacts on the development of the digital publishing industry. Across various dimensions—including content production, process management, dissemination modes, and delivery forms—the industry has demonstrated unprecedented characteristics: diversification of production entities, collaboration in process management, contextualization of dissemination approaches, and diversification of delivery forms. Digital publishing has gradually transcended the limitations of traditional publishing, developing toward an integrated convergence with multiple business forms such as books, journals, newspapers, radio and television, as well as internet publishing. Against this contextual backdrop, this paper elaborates on the application value of big data for digital publishing from perspectives including user profiling, product discovery and filtering, publishing process collaboration, supply chain structural transformation, and knowledge service deepening. The study posits that via big data-driven precise positioning, user experience optimization and iteration can be achieved; through the construction of knowledge association systems, content discovery and filtering can be realized; through the integration of internal and external enterprise resources, collaborative product operations can be promoted; through the optimization of supply chain structures, business risks can be controlled; and through enhanced in-depth resource development, knowledge service renewal and iteration can be accomplished.

## Full Text

### Preamble

**Theoretical Research: Media Theory**

**A Preliminary Exploration of the Application Value of Big Data to**

**Digital Publishing***ChinaXiv Cooperative Journal*

**Abstract:** The emergence of big data technology has profoundly impacted the development of the digital publishing industry. In aspects ranging from content production and process management to dissemination methods and delivery formats, unprecedented characteristics have emerged: diversified production entities, collaborative process management, contextualized dissemination, and diversified delivery formats. Digital publishing has gradually broken free from the limitations of traditional publishing, evolving toward the integration of multiple formats including books, journals, newspapers, radio and television, and internet publishing. Against this backdrop, this paper elaborates on the application value of big data for digital publishing from the perspectives of user profiling, product discovery and filtering, publishing process collaboration, supply-demand chain restructuring, and the deepening of knowledge services. It argues that through big data-enabled precise positioning, user experience can be optimized and iterated; knowledge association systems can be constructed to enable content discovery and filtering; internal and external enterprise resources can be integrated to promote collaborative product operations; supply-demand chain structures can be optimized to control operational risks; and deep resource development can be strengthened to achieve iterative updates in knowledge services.

**Keywords:** big data; digital publishing; knowledge association; knowledge services

**CLC Number:** G230

**Document Code:** A

**Article ID:** 1671-0134(2019)09-016-06

**DOI:** 10.19483/j.cnki.11-4653/n.2019.09.001

**Author:** Zhao Hongyuan

## 1. The Significant Value of Big Data for Digital Publishing

As big data development enters a stage of substantive application, it has generated profound impacts across all levels of society. At the industry level, decision-making bases have become more comprehensive and accurate, facilitating the formation of new core competencies and engendering deep transformations in enterprise production models, business models, and management paradigms, while blurring industry boundaries. At the level of thinking patterns, there has been a shift from relying on sample data to utilizing entire datasets, from pursuing precision to tolerating ambiguity, and from focusing on causal relationships to emphasizing correlations. At the level of wisdom and strategy, while data proliferation creates chaos in terms of “quantity,” it may simultaneously begin to reveal order in terms of “quality,” posing new challenges for social development and management models [1].

Against this backdrop, digital publishing has gradually moved toward diversifi-

cation: network publishing, mobile publishing, database publishing, convergent publishing, and knowledge services. The characteristics of digital publishing itself have undergone significant changes, evolving from simple digitization in content production, management processes, product forms, and dissemination channels to diversified content production entities, diversified delivery methods, systematic service processes, and intelligent content recommendation. The diversification of content production entities is manifested in the emergence of various content production models such as UGC (User-Generated Content), PGC (Professionally-Generated Content), OGC (Occupationally-Generated Content), and DGC (Device-Generated Content). Diversified delivery methods include paper books, audio books, VR/AR books, knowledge payment, and knowledge services. Systematic service processes are reflected in multi-level association systems that provide readers with complete knowledge systems. Intelligent content recommendation is based on precise recommendations grounded in user profiling.

Throughout the formation and evolution of these characteristics in digital publishing, the application of big data technology has been the primary driving factor. As the digital publishing industry develops, the application value of big data continues to grow.

### 1.1 Construction and Iteration of User Personas

User persona, a concept proposed by American interaction designer Alan Cooper in his book *The Inmates Are Running the Asylum: Why High-Tech Products Drive Us Crazy and How to Restore the Sanity*, originally meant “mask” and by extension refers to the “role” represented by the mask, primarily serving to describe user characteristics. Based on users’ actual and potential characteristics, user personas can be divided into static and dynamic profiles. The former focuses on users’ statistical characteristics (such as birth date, gender, education level, and occupation) and behavioral characteristics (such as consumption habits in daily necessities, housing, transportation, entertainment, social interaction, and learning). The latter addresses users’ possible, potential, and trending characteristics [2]. The former is demand-oriented, focusing on user behavior patterns in application scenarios, while the latter is lifestyle vision-oriented, emphasizing user trend prediction based on data mining [3].

From a big data perspective, the concept of users extends to all users, including actual or potential users as well as cross-industry potential users. The criteria for defining users lie in the results of value provision and value co-creation, rather than whether they have purchased the enterprise’s products/services or participated in its activities [4]. User profiling based on big data involves user labeling and constructing data models to study “full-sample” user characteristics, which helps accurately identify and depict target users, effectively connecting user attributes, behaviors, and expectations while greatly avoiding interference from subjective factors. It facilitates timely detection of user changes, promotes product/service upgrades or adjustments, and enables precise recommendations

tailored to individual users [5].

Compared with other industries, users in digital publishing exhibit greater complexity. First is the multiplicity of user roles. The same user may be not only a single service user but also a hybrid role combining several or all of the following: content creator, user, interpreter, and disseminator. Second is the interconnect-edness among users. Based on attention to common thematic content and users' own knowledge levels, collision or assimilation effects occur among users, forming diverse associative relationships. Finally, there is the dynamic evolution of user characteristics. As users' knowledge increases, their demand for content, roles in the content industry chain, and social relationships are in a continuous state of dynamic evolution. These characteristics of digital publishing' s target users require not only accurate description of users' actual features for static profiling to achieve precise product/service push but also a future-oriented approach to accurately grasp the development trends of user needs for dynamic profiling, thereby capturing changes in user interests and demands and achieving complete construction and continuous iteration of user personas.

## 1.2 Discovery and Filtering of Long-Tail Products

In October 2004, Chris Anderson, editor-in-chief of *Wired* magazine, proposed the Long Tail theory, arguing that “as long as storage and distribution channels are large enough, the combined market share of niche or slow-selling products can rival or even exceed that of a few hit products—in other words, numerous small markets can aggregate into a market energy that rivals the mainstream market” [6]. The significance of the Long Tail effect lies in “the accumulation of all non-popular markets forming a market larger than the popular market.”

In the era of big data, people have access to increasingly large amounts of information, thereby improving the long-standing information asymmetry phenomenon. However, big data has also triggered a selection dilemma. Faced with massive information, people feel overwhelmed, particularly struggling to identify large amounts of limited-value or even junk information, thus creating a paradox in big data practical applications. This is reflected in the market economy as the coexistence of power-law distribution [7] and long-tail distribution. On one hand, due to the concentration effect of attention, the proportion of head products becomes smaller while individual product sales increase. On the other hand, long-tail products proliferate with uneven quality, making selection and identification increasingly difficult. Solving this problem requires returning to the fundamental value of big data itself: achieving precise product recommendations through complete user profiling and product feature description.

The book publishing industry is a niche industry. On one hand, most books have limited readership and cannot achieve huge sales volumes. On the other hand, with so many books flooding the market, most struggle to find their target readers, with only a tiny fraction becoming bestsellers [8]. In the digital publishing stage, this characteristic becomes even more pronounced. As storage and

replication marginal costs approach zero and creation barriers are significantly lowered, the scale of published works expands dramatically. The polarization between power-law distribution and long-tail distribution becomes more intense in the digital publishing industry, making the discovery and filtering of long-tail products even more challenging.

Through big data technology, establishing complete user profiles helps identify user needs, while constructing complete knowledge association systems helps navigate and locate high-quality long-tail products. User evaluations and various citation relationships related to works can be used to filter out inferior long-tail products.

### 1.3 Collaboration in Publishing Processes

Collaboration refers to the process or capability of coordinating two or more different resources or entities to achieve a common goal. From the perspective of enterprises with independent legal status, collaboration includes both inter-enterprise and intra-enterprise collaboration. Inter-enterprise collaboration comprises vertical and horizontal collaboration. The most typical example of the former is outsourcing, while the latter involves joint development where two or more parties form project teams for collaborative development, with intellectual property rights and profits shared proportionally. Inter-enterprise collaboration can be divided into three levels based on tightness: low coupling, medium coupling, and high coupling. Intra-enterprise collaboration also has three levels: product structure collaboration, professional technology collaboration, and collaboration among various internal business departments [9].

Enterprise collaboration in the big data context emphasizes resource integration. Internally, it is necessary to break system boundaries and departmental silos, integrate multi-source heterogeneous data from inside and outside the enterprise, and through big data analysis, understand the enterprise's own strengths and weaknesses, market environment, and user characteristics and needs, thereby reforming and optimizing internal management and service processes to provide precise products/services. Externally, enterprises at different development stages inevitably generate various virtual or actual teams, leading to continuously expanding collaboration boundaries and correspondingly expanding and changing resource demands. Through big data analysis, various external resources including human, organizational, technical, and financial resources can be effectively integrated and dynamically adjusted according to actual conditions. This enables information sharing and utilization both within and between enterprises, optimizes external cooperation processes, and thus significantly enhances efficiency [10].

Compared with other industries, digital publishing processes involve more complex factors. User profiling, topic selection planning, expression method determination, delivery method selection, integration of external review resources and authors, association system construction, interactive function design, content su-

pervision, and copyright tracking and monitoring all require integrating various internal and external enterprise resources. Different needs demand corresponding identification and selection, especially when facing rapidly changing content creation innovations, which requires more collaborative publishing processes to achieve efficient product/service market operations.

The application of big data technology helps enhance the marketization degree of every link in the publishing process, providing decision support for resource integration at different stages of the product lifecycle through data analysis and mining. This enables the identification, selection, acquisition, allocation, activation, and organic integration of internal and external resources, constructs enterprise value chains [28], promotes effective connection and maximized resource utilization across development stages, timely identifies and handles operational risks, accelerates product updates and iterations, and achieves collaborative product operations.

#### 1.4 Optimization of Supply-Demand Chain Structure

Supply-demand chain research originated in the 1960s and generally refers to the network formed by enterprises or enterprise departments engaged in material acquisition, processing materials into intermediate or finished products, and delivering finished products to users. Participants in the supply-demand chain include suppliers, manufacturers, distribution centers, retailers, and end users. The supply-demand chain involves logistics, information flow, and capital flow among related entities [11]. Based on the type of enterprise playing a core role in the supply-demand chain, supply-demand chain models can be divided into distributor-centered, manufacturer-centered, and auctioneer-centered models [12]. Effective supply-demand chain management helps enterprises reduce costs, enhance competitiveness, accelerate logistics and information flow throughout the supply-demand chain, shorten product circulation cycles, and strengthen core competencies [13].

The emergence of big data technology has changed traditional supply-demand chain models by quantifying and analyzing supply-demand information and reducing uncertain risks in supply-demand chains. This is manifested in two aspects: first, the establishment of a point-to-point model (i.e., 定点 production and sales based on consumer demand) replacing the face-to-face model (i.e., traditional batch model); second, improved breadth, depth, and accuracy of supply-demand analysis and forecasting. The first manifestation means shortened communication distance between production and consumption ends for producers, leading to the differentiation and disintegration of the three supply-demand chain models mentioned above, with personalized demand satisfaction gradually becoming the mainstream production model. For consumers, it means the prevalence of on-demand purchasing, where users select appropriate products/services based on recommendations from big data analysis and their actual needs. The second manifestation involves the evolution from traditional sample analysis and profile analysis to precise analysis, which forms the basis for

predicting future trends and determining enterprise strategic decisions [14].

Compared with supply-demand chains in other industries including traditional publishing, digital publishing differs significantly in both composition and operation mode. In terms of composition, there are content supply-demand chains, technology supply-demand chains, and carrier supply-demand chains. The content supply-demand chain runs from authors to publishing units to platform operators (including self-built platforms by publishing units) to end users. The technology supply-demand chain runs from data processors (such as digital or audio-video processing of content) to publishing units to technology developers (platform or database development). The carrier supply-demand chain runs from authors to publishing units to channel providers to retailers and finally to readers. In each supply-demand chain, except for the intermediary publishing units, the relationships are not one-to-one or one-to-many as in general industries but many-to-many, with overlapping or collaborative relationships among roles on the chain. The former is exemplified by the overlap between author and reader roles, while the latter is exemplified by collaboration between data processors and platform developers.

The application of big data transforms both upstream and downstream of the entire supply-demand chain from relatively simple linear structures to multi-dimensional network structures [15]. Through profiling of end users and analysis of various roles on the supply-demand chain, precise matching of supply and demand in publishing can be achieved, optimizing the supply-demand chain structure.

### 1.5 Deep Development of Knowledge Services

Knowledge services refer to personalized intelligent services that take knowledge resources as objects and employ organization, association, mining, and reconstruction of information and knowledge as means. This is the core guiding ideology for designing today's knowledge and information resource sharing models, which provide users with solutions to specific problems—that is, “knowledge” [16]. Knowledge includes explicit knowledge and tacit knowledge. Explicit knowledge refers to knowledge that can be easily converted into symbols and expressed in language for dissemination. Tacit knowledge is hidden in the hands and minds of knowledge subjects, manifested as skills, know-how, insight, experience, and other knowledge that cannot be encoded in language or specific symbols [17]. The most typical manifestation of the former is publications, especially books, where systematic argumentation and exposition, convenient dissemination and delivery make them the optimal carriers for explicit knowledge. The most common expressions of the latter include master-apprentice inheritance, practical exercises, and brainstorming, characterized primarily by being “tacitly understood but not explicitly stated.” The two can also be mutually transformed: tacit knowledge can be converted through socialization, externalization can transform tacit knowledge into explicit knowledge, integration can transform explicit knowledge, and internalization can transform explicit knowl-

edge back into tacit knowledge [18]. Through continuous mutual transformation between explicit and tacit knowledge, knowledge innovation and dissemination are achieved.

The advent of big data has brought unprecedented changes to knowledge services in terms of knowledge production and updating, typical characteristics, service scope, and models. First, knowledge production entities have diversified, including both domain experts and ordinary users, professional institutions and individual producers, resulting in both systematic professional knowledge and fragmented application/popularization knowledge, which enhances the necessity of constructing knowledge association systems. Second, in terms of knowledge updating methods, there are both qualitative systematic updates and quantitative fragmented updates generated through knowledge interaction, which simultaneously serves the functions of user experience and knowledge output. Third, the typical characteristics of knowledge services are manifested in: facing autonomous needs and intelligent services, uncertainty, emphasizing user participation, on-demand use and payment, and group-based innovation [19].

From a social function perspective, digital publishing is no different from book publishing in that both serve the functions of entertainment and education. The former includes online literature, online games, digital music, and online animation; the latter includes e-books, digital newspapers, digital journals, online educational publications, database publications, and mobile publications. From the above analysis of digital publishing and knowledge services in the big data context, it is evident that knowledge services emphasize a user-centered approach more strongly. Based on users' information preferences and specific needs, personalized service methods are adopted to provide users with solutions to practical problems. Therefore, knowledge services can be regarded as the deepening of digital publishing in the education domain [20].

Leveraging big data technology, knowledge services can integrate not only existing systematic knowledge resources but also relevant fragmented knowledge resources from society, while generating new knowledge through knowledge interaction. On this basis, complete knowledge association systems can be constructed to achieve associations between systematic and fragmented knowledge, automatically establish connections between new and existing knowledge, and eliminate problems such as "information overload," "knowledge fragmentation," and "information silos," thereby promoting the deep development of knowledge services [21].

## 2. Leveraging Big Data to Achieve Transformation and Development in Publishing

### 2.1 Optimizing and Iterating User Experience Based on Functional Positioning

User experience, simply put, refers to users' subjective feelings, primarily concerning their cognitive impressions and responses before, during, and after using a product [22]. The main factors affecting user experience are fourfold: product strategy, user interface, technology, and operations [23]. Big data quantifies indicators such as user pleasure, engagement, acceptance, retention, and task completion to establish complete static and dynamic user profiles, thereby improving factors affecting user experience and achieving optimization and iteration of user experience.

The functional demands of big data for user experience can be divided into three categories: intelligent recommendations based on sales, demand analysis based on topic selection, and capability analysis based on creation. Therefore, optimization and iteration of user experience should proceed from these three aspects.

First, maximize the collection and organization of massive user data, analyze their static characteristics including statistical and behavioral features, identify potential users for products, and provide precise recommendation services with contextualized introductions to simplify user selection processes and enhance purchase experience.

Second, strengthen the collection and organization of unstructured data, particularly topics of common interest, expressed opinions, and ideas in potential niche fields. Through mining and analysis, discover the deep thoughts and trends reflected behind user data, plan topic selections for niche markets, and continuously track these changes during the topic selection process with constant iteration to ultimately form topics and products that meet user needs.

Finally, through tacit knowledge annotation and tacit knowledge map construction, identify potential authors who meet requirements and analyze their professional and creative capabilities. Tacit knowledge annotation is conducted through the externalization of tacit knowledge: first, individuals proactively provide information including personal details, experiences, created works, and content output during interactions, from which personal identification labels are selected; second, evaluations from others, mainly various comments, interpretations, and attitudes from surrounding people during interactions, provide more objective assessments from an external perspective; finally, the system automatically discovers and timely captures individuals' tacit knowledge, such as skills unconsciously displayed during person-to-person interactions [24]. Tacit knowledge map construction primarily involves analyzing associative relationships among data related to authors, including author-work relationships, relationships among authors in the field, and relationships among works in the

field. Based on authors' professional capabilities, creative abilities, and communication skills, the creative methods, expression methods, and dissemination collaboration for works are determined.

## 2.2 Building Association Systems for Content Discovery and Filtering

The value of big data is expressed through the correlation of big data—that is, the structure of big data [25]. The coexisting phenomena of information overload and information silos accompanying big data, together with the parallel existence of power-law distribution and long-tail distribution in digital publishing, result in most content (except for a few head contents) not receiving sufficient exposure opportunities to be known by users, while some inferior content cannot be processed in a timely manner, reducing the signal-to-noise ratio. To solve this problem, we must return to the fundamental value of big data: building association systems to fully utilize the multi-type, multi-dimensional, and multi-level associative relationships among knowledge, thereby achieving knowledge mining, discovery, verification, identification, and filtering [26].

Using big data technology to construct knowledge associations and achieve discovery of high-quality content and filtering of inferior content can be approached from several aspects:

First, build existing knowledge associations of explicit knowledge, knowledge content, and disciplinary knowledge through big data-based automatic annotation, social annotation, and knowledge element annotation. On this basis, combine outputs from knowledge interaction processes to continuously incorporate newly generated knowledge into existing knowledge systems while improving knowledge evaluation indicator systems (including output knowledge during interaction). Then establish tacit knowledge associations and personalized thematic knowledge associations to achieve knowledge discovery and filtering through mutual verification among knowledge and user evaluations.

Second, improve associative relationships from three dimensions: knowledge-to-knowledge, knowledge-to-person, and person-to-person. Knowledge-to-knowledge association refers to the association among explicit knowledge mentioned above. Knowledge-to-person association requires annotating people—that is, readers and authors—including both statistical and behavioral feature annotation. On this basis, through diverse associations between knowledge and authors, the authenticity of knowledge content can be determined. Content authenticity is identified through two approaches: first, similar to co-citation and coupling analysis in literature analysis, discovering associative relationships of specified knowledge content to make correct judgments; second, through user evaluations, drawing correct conclusions based on majority opinions.

Finally, comprehensively apply simple association, temporal association, and causal association rules to timely detect abnormal data, eliminate spam accounts and spam information, and reduce noise. Through simple association rules measuring the correlation between two types of variables (such as output

content and original content during interaction), the quality and motivation of output content can be judged. Temporal association calculations are performed on behaviors (such as content output behavior and timing) to determine the diligence of outputters or the nature of output content. Based on criteria including temporal sequence, strength, specificity, reproducibility, consistency, rationality, and similarity of associations, the likelihood of causal associations is determined to discover new knowledge and judge the correctness of existing knowledge [27].

### **2.3 Integrating Internal and External Resources to Promote Collaborative Product Operations**

The role of big data in optimizing digital publishing processes is primarily realized through data analysis and mining to provide decision support. This enables the identification, selection, acquisition, allocation, activation, and organic integration of internal and external resources at different stages of the product lifecycle, constructs enterprise value chains [28], promotes effective connection and maximized resource utilization across development stages, timely identifies and handles operational risks, accelerates product updates and iterations, and achieves collaborative product operations.

Leveraging big data technology to integrate resources and promote collaborative product operations should proceed from the following aspects:

First, during the topic selection planning stage, considerations should be made from both content and primary creator perspectives. Regarding content, through big data analysis and mining of existing massive literature data, based on multiple indicators including attention levels, word frequency statistics, co-citation degrees, and coupling degrees, research hotspots in disciplines, blind spots in discipline niche markets (such as research or popularization), and focuses of concern for readers at different levels can be timely identified to determine topic selection directions (i.e., ideas). Then, through user profiling analysis of user application scenarios, expression methods suitable for different scenarios are determined. Regarding primary creators, while analyzing and mining existing massive literature data, an author database should be established to track and analyze authors' professional research directions, academic attention levels, media attention, academic dissemination, user attention, creative styles, fan groups, and dynamic behaviors, enabling precise author selection based on work application scenarios. For topic selection analysis based on copyright introduction, market analysis of the copyright work's output region and analysis of the author and work's influence and prospects in the input region should be added.

Second, during the content production stage, two main aspects should be considered: first, contextualized expression and iteration; second, timely establishment and updating of content associations. The former involves segmenting user application scenarios, integrating relevant data resources for different scenarios,

and achieving “create once, publish many times.” On this basis, expression and delivery methods matching application scenarios are determined, such as immersive deep or shallow reading, companion reading (parent-child reading falls into this category), and accompanying reading (audio books), which can respectively adopt professional exposition or popular narration, audio or text formats. During the creation process, authors’ creative thinking is continuously assisted based on user feedback to improve delivery design and achieve product updates and iterations. The latter involves continuously establishing and improving content associations while content is published and updated, achieving value-added and iterative delivery as described above.

Finally, during the product dissemination stage, big data technology should focus on solving the connection problem between products and application scenarios to improve product delivery efficiency. This connection is achieved through two pathways: first, finding target users and their usage scenarios through big data analysis and achieving intelligent precise product recommendations through user profiling; second, designing targeted incentives to stimulate users’ active dissemination behavior and achieve viral product spread. Intelligent recommendations are realized in two ways: one is active recommendation based on integrating user and application scenario connection channels—for example, smart speakers are most suitable for accompanying reading scenarios, so audio books suitable for accompanying scenarios are recommended to users through smart speaker physical distribution channels; the other is association rule mining-based recommendation (here “association” refers to association in the general sense, not knowledge association), which finds strong association rules by analyzing metrics such as lift, confidence, and support to provide users with extended recommendations.

#### **2.4 Optimizing Supply-Demand Chain Structure to Control Operational Risks and Reduce Costs**

The primary role of supply-demand chain optimization lies in timely identification of operational risks and cost reduction. The two functions of big data in supply-demand chain optimization are mainly achieved through historical data analysis of both upstream supply and downstream demand ends, enabling enterprises to judge various capabilities and risks on the supply end and improve operational levels while discovering and creating user needs.

Optimizing the supply-demand chain through big data technology to enhance operational efficiency should proceed from two aspects:

From the supply end perspective, the focus should be on strengthening capability analysis, including production capacity, responsiveness, management capability, and risk prevention capability of the supply end. The supply end of digital publishing includes content suppliers, technology suppliers, and carrier suppliers, with content suppliers being the key focus, encompassing various UGC, PGC, and OGC. First, through big data collection of digital publishing units’ own busi-

ness data, user experiences, and interaction data, the enterprise' s positioning and core functions should be analyzed and determined for dynamic correction and adjustment. On this basis, through analysis and investigation of content suppliers such as PGC and OGC, a knowledge production sharing and collaboration network should be established to determine systematic content sources and supply rhythms. Screening and review methods for UGC content producers should be established, along with strict review methods for content to ensure output content legality. Methods for discovering and cultivating UGC content producers should be developed to gradually develop them toward PGC and improve their content production quality. Combined with manual review and user feedback, processes for screening, evaluating, and deeply processing high-quality content should be established.

From the demand end perspective, the focus should be on discovering unmet content needs in the market through analysis of various semi-structured and unstructured data, primarily based on knowledge element analysis of content, output content during content interaction (such as book reviews), and various self-media content. Through relevant and causal association analysis, the basic expression ideas needed are determined, and further, content expression methods are determined through application scenarios presented by user profiling. Second, user basic characteristic data, consumption records, geographic location data, online interaction data, offline behavior data, web public data, and social-generated data [29] are used to determine the relationship between content production elements and user data. Finally, through matching with content producers on the supply end, the division of labor and cooperation among different content producers is determined to meet market demands.

## **2.5 Strengthening Deep Resource Development to Achieve Iterative Updates in Knowledge Services**

The deep development of publishing resources through big data involves, on the basis of integrating owned resources, completing the organization, association, mining, and reconstruction of knowledge to build the basic architecture of knowledge services. It further expands and integrates external resources with timely updates to form iterative updates of knowledge service products, helping users continuously achieve knowledge discovery.

Promoting continuous upgrading of knowledge services through big data-integrated resources should primarily proceed from the following aspects:

First, under the premise of determining knowledge service product themes based on the aforementioned target user profiling, organize required content resources and conduct deep processing. Construct a knowledge annotation system including knowledge element annotation, tacit knowledge annotation, dynamic annotation, and social annotation to establish a complete knowledge association system. Provide users with personalized knowledge systems and achieve knowledge discovery at the individual level.

Second, continuously annotate output content generated through knowledge interaction during knowledge service product operation processes, establishing associations with existing knowledge content. Timely acquire newly generated knowledge content and information from society, and incorporate it into the knowledge system based on knowledge annotation. While forming dynamic knowledge associations, continuously expand the boundaries of the knowledge system to achieve iterative updates in knowledge services.

Finally, establish personalized associative relationships based on users' behavioral characteristics and connections between knowledge to provide users with personalized knowledge systems and achieve knowledge discovery at the individual level. For example, a user's motivation for learning the input hypothesis may be researching second language teaching theory, knowledge service operations, or writing theory research. Therefore, it is necessary to comprehensively examine users' browsing content, created content (including citations), interactive content, and other behavioral characteristics to provide them with associated content related to second language teaching, knowledge service operations, or writing theory research, enabling precise content recommendations that effectively promote user knowledge absorption, drive knowledge output, and achieve knowledge sharing and collaboration.

Compared with traditional publishing, digital publishing integrates multiple formats including books, journals, newspapers, radio and television, and internet publishing. Its boundaries with knowledge services are increasingly blurred, emphasizing not only systematic, associative, and interactive content but also contextualized expression, personalized content, and precise recommendation to improve the effectiveness of idea delivery. These characteristics are both the inevitable result of big data development in the digital publishing field and are strengthened through big data technology.

Big data has profoundly impacted user profiling, long-tail operations, process reengineering, supply-demand chain optimization, and knowledge service deepening through the integration, analysis, association, mining, and reconstruction of various information and data resources. This has formed a truly user-centered pattern of content production, interaction, and delivery, fundamentally overturning traditional publishing operation models. Against this backdrop, digital publishing enterprises must drive comprehensive transformation and development in the publishing industry by optimizing user experience, discovering and filtering content, promoting collaborative product operations, optimizing supply-demand chain structures, and updating knowledge services.

## References

- [1] Jiang Haoduan. The Essence of Big Data and Its Possible Impacts[J]. China Economic Report, 2013(6): 16-22.
- [2] Huang Chengqun, Yu Yanping. Exploration of Modern Publishing Talent Skills Based on Digital Publishing Characteristics[J]. Publishing Reference,

2017(8): 40-42.

[3] Dai Fuping, Xin Xiangyang, Zhang Huimin. Dynamic User Persona: Describing Users is Creating Users[J]. *Art & Design*, 2018(3): 94-96.

[4] Zhang Huimin, Xin Xiangyang. Four Dimensions for Building Dynamic User Personas[J]. *Industrial Design*, 2018(4): 59-61.

[5] Xie Kang, Wu Ji, Xiao Jinghua. User Profiling and User Behavior Analysis Based on Big Data Platforms[J]. *China Informationization*, 2018(3): 100-104.

[6] Hao Shengyu, Chen Jingren. User Personas in the Big Data Era Help Enterprises Achieve Precision Marketing[J]. *China Collective Economy*, 2016(4): 61-62.

[7] Deng Fengyi, Deng Hairong. The “Long Tail Effect” of Digital Publishing in the Big Data Era[J]. *Publishing Research*, 2014(10): 27-29.

[8] Hu Haibo, Wang Lin. A Brief History of Power-Law Distribution Research[J]. *Physics*, 2005(12): 889-896.

[9] Lian Xiaochuan. Power Law, Long Tail Theory and Book Publishing[J]. *Journal of Shaanxi Normal University (Philosophy and Social Sciences Edition)*, 2007(3): 110-116.

[10] Wang Wei. Talking About Collaboration Again: Types of Collaboration (Part 1)[J]. *CAD/CAM and Manufacturing Informatization*, 2013(2): 13.

[11] Liu Guquan. Interpreting Big Data and Big Collaboration[J]. *Zhixie Intelligent Manufacturing WeChat Official Account*, 2014(5): 15.

[12] Chai Yueting, Han Jian, Wu Cheng. Agile Supply Chain and Its Management[J]. *China Mechanical Engineering*, 2000, 11(3): 352-355.

[13] Li Gang, Liu Lu, Chen An. Research on the Operation Mode of Supply Chain Management[J]. *High Technology Letters*, 2001, 11(2): 71-73.

[14] Liu Liwen. Discussion on Implementation Strategies of Enterprise Supply Chain Management[J]. *Computer Integrated Manufacturing Systems-CIMS*, 2000(2): 15-18.

[15] Zhang Yankui. Discussion on the Relationship Between Big Data and Supply Chain Logistics Service Development[J]. *Manager*, 2016(28): 325.

[16] Zhang Chi, Zhong Ying. Research on the Construction of Publishing Supply Chain 2.0 Based on Big Data[J]. *Science-Technology & Publication*, 2015(5): 105-107.

[17] Fu Lei. Design and Implementation of Knowledge Element Annotation System[D]. Hubei: Central China Normal University, Educational Technology, 2009: 1-61.

- [18] Zhang Hongtai. Discussion on the Mutual Transformation Between Tacit Knowledge and Explicit Knowledge[J]. Journal of Chongqing University of Science and Technology (Social Sciences Edition), 2011(7): 162-163.
- [19] Zhao Shiyong, Hong Xiaonan. The Dialectical Relationship Between Explicit Knowledge and Tacit Knowledge[J]. Studies in Dialectics of Nature, 2001, 17(10): 20-24.
- [20] Qin Xiaozhu, Li Chenhui, Mai Fanjin. The Connotation, Typical Characteristics and Conceptual Model of Big Data Knowledge Services[J]. Information and Documentation Services, 2013(2): 18-22.
- [21] Liu Zhi. Research on Knowledge Services of Digital Publishing Enterprises[D]. Hunan: Xiangtan University, School of Public Administration, 2011: 1-57.
- [22] Zhao Hongyuan. Construction of Knowledge Association System from the Perspective of Publishing[J]. China Media Technology, 2019(1): 52-55.
- [23] Li Bin. Redefining User Experience: Clearing Seller Position and Returning to User Perspective[J]. China Businessman, 2018(7): 51-53.
- [24] Lao Cao. Four Factors Affecting User Experience[J]. Everyone is a Product Manager, <http://www.woshipm.com/pd/1664.html>.
- [25] Qin Tieshui, Wang Qiong. Discussion on the Construction of Expert Tacit Knowledge Maps[J]. Journal of the National Library of China, 2007(2): 58-62.
- [26] Liu Wenyuan, Li Shaoxiong, Wang Xiaomin, et al. Big Data Knowledge Discovery[J]. Journal of Yanshan University, 2014(5): 377-380.
- [27] Zhao Hongyuan. Construction of Knowledge Association System from the Perspective of Publishing[J]. China Media Technology, 2019(1): 52-55.
- [28] Zhao Yanfeng. Specious Causal Associations[J]. Scientific World, 2013(8): 114-117.
- [29] Liu Fei, Yan Jianyuan. Case Study on Enterprise Process Reengineering Based on Value Chain Analysis in the Big Data Era[J]. Logistics Engineering and Management, 2014(7): 226-228.
- [30] Liu Dequn. Transformation of Film and Television Text Creation and Production Models Driven by Big Data[J]. Journal of Nanjing University of Posts and Telecommunications, 2019(4): 38-45.

(Author' s Affiliation: Shanghai Century Publishing Group)

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

*Source: ChinaXiv –Machine translation. Verify with original.*