The Driving Role of User Data in Publisher Transformation (Postprint)

Authors: Dou Linqing, paired thoroughbreds, Liang Jie

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

Abstract

In the era of big data, the value of data resources in the publishing industry is becoming increasingly prominent, among which user data possesses even greater value due to four characteristics: extensive population coverage despite being a small industry, high inherent data dimensions of products, low data acquisition costs, and sustainable annual traffic. User data construction can not only assist publishing houses in expanding their digital business but also help paper book businesses break through sales ceilings, thereby facilitating the transformation of publishing houses by altering the mindset and business models of publishers. Therefore, with the in-depth development of media convergence, user data construction is destined to become a foundational business for publisher transformation, warranting continuous and sustained investment by publishing houses in the future.

Full Text

Abstract

In the era of big data, the value of data resources in the publishing industry has become increasingly prominent. User data, in particular, holds even greater value due to four distinctive characteristics: extensive population coverage despite being a niche industry, inherently high-dimensional product data, low data acquisition costs, and sustainable annual traffic flow. The development of user data infrastructure not only helps publishers expand their digital businesses but also enables paper book operations to break through sales ceilings. By transforming the mindset and business models of publishing professionals, it facilitates industry-wide transformation. Therefore, as media convergence deepens, user data infrastructure will inevitably become a foundational business for publisher transformation, warranting continuous long-term investment.

Keywords: user data; media convergence; big data; publishing industry; internet thinking

Introduction

As technology evolves, the ways in which information and content are disseminated have changed dramatically. However, for various reasons, the publishing industry has yet to reach a consensus on the importance and necessity of user data infrastructure. On one hand, publishers continue to discuss the concept of "user-centric thinking" that other industries embraced over a decade ago, yet hesitate to implement actual user data development initiatives. On the other hand, the industry has successively championed "content is king," "technology is king," and "channel is king," with the notion that "the future of publishing is 'user is king' " only emerging recently [1]. As media convergence develops in depth, user data infrastructure will become an essential foundational business for any publisher seeking transformation and the source of long-term value creation. Consequently, user data development can serve as a long-term strategy for publishers, deserving continuous investment over the next decade or even longer.

1. User Data as the Real Data Goldmine in Publishing

In the big data era, data-driven approaches have reconstructed every aspect of publishing, making the asset value of data increasingly apparent [2-4]. Publishing data can be categorized in numerous ways when examined by attribute, with vertical domains differing significantly and horizontal dimensions like users and content showing vast disparities. Broadly speaking, however, all publishing data falls into two categories: professional data and public data. The distinction lies in whether data demonstrates value only within its vertical domain (professional data) or across all vertical domains (public data). Content data belongs to the former category, as its value is realized only within content-related industries. User data, by contrast, constitutes public data—professional domain attributes represent merely one of many user data attributes. Since diverse content competes for users' limited attention, only user data can enable deep cross-industry integration between publishing and the broader national economy.

Why, then, is user data in publishing more valuable than in other industries? This stems from four key characteristics of publishing user data: extensive population coverage despite being a niche industry, inherently high-dimensional product data, low acquisition costs, and sustainable annual traffic. Any one or two of these characteristics would guarantee access to high-quality user data resources in other industries. Although publishing' s output value is relatively small, books as a special product category achieve coverage of 60-70% of the national population [5].

Book products differ from most commodities by naturally carrying multiple high-dimensional data attributes. Even before production, during the topic selection

stage, the target user group is clearly defined—whether children, women, or seniors. Is the product purchased by mothers for their children, or by adult children for their parents? Each book can thus be assigned a clearly defined demographic profile before production. Since book products inherently possess rich data dimensions, the sales process essentially transfers these predefined tags to the user group through transactions. The essence of this process is not people finding books, but books finding people—it is a user tagging process.

How costly is user data acquisition currently? The collection process itself reveals the answer: each publisher employs hundreds of ground promotion teams that collect user data by distributing flyers—millions or even tens of millions annually. These flyers not only remain with users rather than being discarded but also accompany them throughout their daily lives. Remarkably, this requires no additional payment to the promotion teams, enabling near-zero-cost acquisition of millions of high-quality user traffic. Multiplying this by a conversion rate yields valuable user data. The actual cost? Perhaps a value-added service concept bundled with books, a digital topic selection idea accompanying the publication, or simply adding a row to the topic selection form for user data collection methods. The only unavoidable cost is time, which may become the greatest expense in user data infrastructure development.

Currently, internet demographic dividends have disappeared, with leading apps from BAT and ByteDance capturing most top-tier online traffic. Only limited opportunities remain in vertical demographics and scenarios. Online promotion must therefore follow rules set by these traffic giants, whose monopoly has driven traffic costs so high that customer acquisition expenses now exceed traditional media advertising costs. Consequently, even internet companies are seeking offline traffic, giving rise to the new retail phenomenon—essentially transforming existing offline traffic into visualizable, operable user data assets through enriched online services. Publishing similarly possesses substantial offline traffic that, like an engine, provides annual continuous coverage. Yet without operationalizing this traffic, it goes to waste each year. This book-borne offline traffic, along with publisher brands and user data, represents valuable intangible assets that publishers must rigorously protect and properly utilize.

2. User Data Helps Paper Book Business Break Through Sales Ceilings

Generally, user data is perceived as relevant only to digital business, not to paper book sales breakthroughs. However, by observing online data changes and offline book sales on the Yuemenhu platform, we can identify replicable patterns from typical cases. User data not only expands digital business but also helps sell more paper books, enabling paper book operations to break through existing sales ceilings—a valuable capability amid reduced book numbers and rising paper costs.

Increasing paper book sales involves two primary approaches: finding increment

from stock and finding increment from increment. Both require first establishing data connections with existing users to collect their information. The keyword for "finding increment from stock" is conversion—transforming readers into users. These are distinct concepts: a reader' s value ends after purchase, while a user' s value begins at purchase.

How can publishers find increment from stock? The mature approach involves deep bundling with books—not merely adding QR codes for value-added services, which yields low conversion rates. The optimal method is intervening during the golden early stage of topic selection, even adding corresponding content to the topic selection form to design digital services as an inseparable part of the book usage experience. This way, paper books generate revenue while digital products or services collect user data. Additionally, established offline book promotion activities can be digitized or combined with online elements to collect previously untouchable user data online, forming visualizable and operable user data assets —another method for finding increment from stock.

The keyword for "finding increment from increment" is marketing—an external expansion method for publishing user data. Three relatively mature approaches currently exist: digital product "offline stores," online brand amplification of offline brands, and matrix dissemination; and user 裂变 (fission) dissemination. Portal platforms can enable digital product "offline stores" by generating unique QR code posters for different bookstores (or even for authors and WeChat influencers), creating offline sales channels for digital products and services. Bookstores can recommend these to their offline traffic without inventory, returns, or stocking, earning channel revenue shares from sales. Online brand amplification and matrix dissemination extend and magnify existing brands on the internet—brands being the most stable foundation for publishers to find incremental users. This is carried out through mobile products, so a scaled, matrixed mobile product system can rapidly create online brand effects.

User 裂变 (fission) dissemination is currently one of the most suitable and cost-effective models for publishers to find increment from stock. By benefiting both inviters and invitees, a single existing user can attract multiple new users. However, WeChat has imposed many restrictions on user acquisition, such as prohibiting inducement to share, making direct user acquisition more straightforward within publishers' own home-field apps.

Convergent publishing should integrate not only owned content resources but also user needs. By remaining user-centric, user data infrastructure becomes an offensive and defensive strategy. Even without independently operated digital businesses, it plays an invaluable role in breaking paper book sales ceilings, maintaining market share, and enhancing differentiated competition.

3. User Data Drives Upgrading of Publishers' Mindset and Business Models

What distinguishes editors at the "Dedao" APP from traditional publishing house editors? From an individual perspective, there is essentially no difference—reportedly, several editors from a publishing house became "Dedao" APP editors immediately after Luo Zhenyu' s talk there that morning. The difference lies in how these editors operate: "Dedao"APP editors establish direct connections with users through the app, accessing first-hand user data. One could say they are editors armed with user data, with user data empowering them as an "individual combat system." This user data is not WeChat public account data or third-party platform Excel sheets. Rather, user data collection and operation occur within a single scenario with substantial operational and service autonomy. I term this "home-field user data," characterized by its ability to self-cycle and improve products and services. Conversely, when collection and operation occur in separate scenarios or operational autonomy is lacking, this constitutes "away-field user data."

How do some publishers on the Yuemenhu platform leverage user data? They treat user data infrastructure as a long-term strategy, using user data as the basis for business decisions to support rapid paper book development. To prioritize this strategy, they eliminate many distractions, even sacrificing short-term online revenue, which they view as an interference factor. Consequently, their online business KPIs focus not on revenue but on home-field user data.

If publishers value user data, they should look beyond registration and app download numbers. To enhance user data value, they must increase user activity and encourage more time spent on their home-field apps. Some publishers have purchased large amounts of premium foreign content to provide free access. This approach boosts user stickiness: user data supports precise topic selection and marketing for paper books, while free digital products and services collect user data and activate user value, ultimately achieving integrated development of traditional publishing.

4. User Data Redefines the Publishing Industry

With user data, a publisher' s business essence shifts from selling books, products, or knowledge to managing users. Once a user data strategy begins, books, products, and knowledge become less important. Publishers must instead focus on users who want to learn, read, improve, and experience—their communities, relationships, and data. Books and knowledge products thus become tools for establishing, maintaining, and strengthening connections with users at certain levels. In other words, selling books versus using book sales to find long-term user communities are fundamentally different concepts. Only through community operation can publishers obtain long-term value.

After identifying user groups through online, offline, product, service, and activ-

ity channels, publishers must understand users better than competitors to continuously provide needed products and services—this constitutes their long-term value and profit moat. Therefore, at the outset of user data strategy development, an editor must clarify four questions: which reader data is currently being collected by your books or products, how this data collection is achieved, what profit and revenue models are derived from reader data, and what competitive moats are formed through superior reader understanding. With these considerations clear, you realize you are not merely selling books but using them to find people. By aligning yourself with these users, you become their advocate for interests and needs, shifting from a seller's to a buyer's market. This better leverages editors' natural advantages in content curation and resource integration, transforming user needs behind you into competitive business advantages. This continuous influx of user data, brought by books or other channels, becomes the source of new business growth.

Thus, in the book-selling era, publishers aimed to sell books, with users merely tools for fulfilling sales targets. In the user economy era, user relationship management becomes central, with books serving only as a medium for user operation. After transformation and upgrading, publishers become not just content production companies but "content + user" data service companies.

The publishing industry's greatest advantage lies in its content capabilities, with content being one of the most cost-effective ways to acquire user data. Using content data as a stepping stone to obtain user data, then user needs as the core to drive publishing through user data, the industry can enter a broader spiritual and cultural domain. By leveraging its user data characteristics and advantages, publishing can achieve dual empowerment of user data and content data for other industries, even positioning itself at the entry point and center of this domain. Only then can equal and valuable cross-industry integration between publishing and the broader national economy occur.

References

[1] Zhou Zhonghua, She Guang, Xing' anfu. The Future of Publishing and New Retail [J]. Publishers, 2019(2): 42-43.
[2] Zhang Bo, Lei Jin, Ren Dianshun. Value Discovery and Activation of Traditional Publishing' s Legacy Content from a Data Asset Perspective [J]. Science-Technology & Publication, 2018(10): 140.
[3] Yang Yang, Zhang Xueqian. Innovative Practices in the International Publishing Industry in the Big Data Era [J]. Editorial Friends, 2018(12): 5-9, 19.
[4] Hao Yang. Implementing the National Big Data Development Action Plan: Serving Digital Publishing Transformation and Upgrading with Big Data Thinking [J]. Science-Technology & Publication, 2016(1): 9-12.
[5] Yan Xiang, Peng Tianshe. Five Implementation Scenarios for Convergent Publishing [J]. Science-Technology & Publication, 2018(6): 46-50.

(Authors' affiliations: Beijing Chuyue Information Technology Co., Ltd.;

Petroleum Industry Press; Editorial Department of Beijing University of Technology Journal)

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