Reflections and Practices on Building a Content Middle Platform for Publishers (Postprint)

Authors: Liu Kunxiang, Miao Lijin

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

Abstract

As the digital transformation, upgrading, and integrated development of publishing houses continue to deepen, numerous backend systems and frontend applications have emerged, while issues such as the disconnect between frontend and backend inherent in the traditional "frontend-backend" architectural model have gradually become apparent. Drawing upon the middle platform concept from the internet industry, China Machine Press has constructed a content middle platform specific to the publishing sector, thereby establishing a novel "frontend-middle platform-backend" architectural model. This has accelerated the circulation of content resources, reduced system development costs, and provided robust support for the scientific, standardized, and efficient operation and management of digital content resources.

Full Text

Preamble

Title: Thoughts and Practices on Building a Content Middle Platform in Publishing Houses

Authors: Liu Kunxiang, Miao Lijin (China Machine Press, Beijing 100037)

Abstract: As digital transformation and integrated development deepen in publishing houses, numerous back-end systems and front-end applications have emerged, while problems inherent in the traditional "front-end–back-end" architecture—such as disconnect between front and back ends—have become increasingly apparent. Drawing inspiration from the middle platform concept of the internet industry, China Machine Press has built a content middle platform unique to the publishing sector, establishing a new "front-end–middle platform–back-end" architecture. This has accelerated content resource circulation, reduced system development costs, and provided robust support for scientific, standardized, and efficient operational management of digital content resources.

Keywords: content middle platform; digital resource management; digital content distribution; copyright protection

CLC number: G641 **Document code:** A

Citation format: Liu Kunxiang, Miao Lijin. Thoughts and Practices on Building a Content Middle Platform in Publishing Houses[J]. China Media Technology, 2021(02): 7-9.

The term "middle platform" originates from the internet industry, corresponding to front-end and back-end. It refers to a collection of shared middleware in system architecture that helps enterprises rapidly integrate back-end resources to build front-end applications for end consumers and customers [1]. Middle platforms are generally divided into data middle platforms and business middle platforms. In the process of digital transformation, publishing houses have also attempted to build their own middle platform systems. For example, the "Renrong Rongzhi Knowledge Service Platform"developed by Posts and Telecommunications Press integrates and analyzes user, resource, and operational data from various systems for decision-making and display purposes, representing a typical data middle platform. China Machine Press (hereinafter referred to as CMP) is currently developing an integrated publishing operation platform that centrally provides common business modules such as channel management, order processing, and after-sales service across traditional and digital, B2B and B2C, and online and offline business lines, representing a typical business middle platform. This paper primarily introduces how CMP, drawing on middle platform concepts, has established a content middle platform unique to the publishing industry to enhance the management and operational efficiency of digital content resources.

[Figure 1: see original paper] Front-end-Back-end Architecture Diagram

1. Challenges in Digital Content Management and Operations for Publishers

As digital transformation and integrated development deepen in publishing houses, numerous back-end systems and front-end applications have emerged. The so-called back-end refers to business systems such as digital content resource management platforms (hereinafter referred to as resource management platforms) and ERP systems, primarily used for resource storage management and workflow operations, typically deployed on local intranets. The so-called front-end refers to content publishing platforms such as e-book platforms, databases, and online course platforms that face end users, including Web applications, mobile apps, H5 pages, and mini-programs, commonly deployed on the cloud. Content resources and data managed by the back-end are continuously transmitted to the front-end for publication and consumption. The commonly used "front-end-back-end" architecture is shown in Figure 1.

The back-end focuses on control and compliance, requiring security and sta-

bility, thus changing slowly with limited external interfaces. In contrast, the front-end focuses on business innovation, requiring rapid response and frequent product iteration. As the number of platforms increases and complexity grows, the traditional "front-end–back-end" architecture poses increasing challenges for publishers' digital content management and operations, mainly manifested in the following five aspects:

1.1 Disconnect Between Front-end and Back-end, Poor Content Circulation, and Version Inconsistency

Currently, most resource management platforms established by publishers only serve as storage warehouses without integration with front-end applications. Content resources for front-end applications often require manual download from the resource management platform, followed by separate upload to each front-end application for publication. This process is inefficient and difficult to manage. Moreover, it easily leads to inconsistencies between content published on the front-end and the final approved versions stored in the back-end, resulting in quality issues.

1.2 Single Resource Format, Incomplete Metadata, and Difficulty Meeting Diverse Front-end Needs

For management convenience and limited storage space, back-end resource management platforms typically store only one original format file. Metadata annotations contain only information required for back-end resource management, particularly with resource classification systems often having only a single dimension (e.g., books classified only by Chinese Library Classification). Front-end applications require secondary conversion and annotation, making it difficult to meet the diversified needs of multi-terminal, multi-scenario, and multi-channel front-end environments.

1.3 Incomplete Copyright Protection Mechanisms, Risking Digital Asset Loss

Since front-end applications are developed and operated by different teams without unified technical standards and security requirements, resource protection mechanisms vary significantly, with some lacking protection entirely. Additionally, many stages in the publication process still require offline manual operations on resource files, creating vulnerabilities for digital content asset loss during front-end publication and usage.

1.4 Lack of Unified Resource Authorization and Authentication Mechanism, Hindering Standardized Operations

With front-end applications operating independently, it is difficult to uniformly manage and monitor which front-end applications (especially external third-party platforms and channels) are authorized to access back-end resources, for

how long, whether IP and concurrent access restrictions apply, and how user access behavior data is handled.

1.5 Repeated Development and Resource Waste

Front-end applications are iterating faster with increasing variety, raising the risk of failure. In actual development, nearly every front-end application must redevelop common functional modules such as resource format conversion, e-book readers, video players, and resource classification management, while also considering adaptation for different terminals like desktops and mobile devices. This leads to longer development cycles, higher costs, non-unified technical standards, wasted hardware resources, and severe redundant development.

To address these issues, CMP learned from the middle platform concept of the internet industry and decided to establish a "content middle platform" unique to the publishing sector between the back-end responsible for content resource management and the front-end responsible for content distribution.

2. Content Middle Platform: Concept and Practice

2.1 Concept of Content Middle Platform

The content middle platform is a middle platform architecture unique to content resource-intensive industries such as publishing. Strictly speaking, it can also be considered a special type of data middle platform. While data middle platforms in the internet industry primarily provide shared services such as data exchange, distribution, analysis, and mining around business data like production and marketing [2], the "content middle platform" in the publishing industry focuses on industry-specific digital content resources. It provides front-end applications with shared services for unified format conversion, copyright protection, product packaging, and distribution management of content resources, forming a closed-loop "front-end-middle platform-back-end" architecture. This bridges the functional vacuum between current back-end content management and front-end content distribution, making the form of digital content resources in the middle platform better suited to the diversified needs of front-end consumption.

The "front-end-middle platform-back-end" architecture is shown in Figure 2 [Figure 2: see original paper].

2.2 Functional Design of Content Middle Platform

The functional design of the content middle platform primarily focuses on providing more powerful and shareable business support services for front-end applications. After development, it is deployed on the cloud like front-end applications. It mainly includes the following seven components:

2.2.1 Resource Extraction Service This service promptly responds to front-end application resource publication requests, extracting required original resources from the back-end to the cloud-based middle platform on demand and caching them for further processing. Resources not scheduled for publication remain in the back-end.

2.2.2 Format Conversion Service This service converts original content resource files extracted from the back-end into various formats required by front-end applications and stores them in the cloud to meet the invocation needs of different front-end applications. For example, original book files are converted into various mainstream e-book formats in the middle platform, while original video files are automatically converted into various common streaming media format files at different resolutions to cover the requirements of most front-end applications.

2.2.3 Copyright Protection Service This service uniformly applies copyright protection processing to files scheduled for publication, including watermarking, file encryption, and HTTPS encrypted access.

2.2.4 Product Packaging Service This service uniformly packages digital content resources to be published into product objects, enabling packaging of both individual resources and sets of related resources. (1) Packaging individual resources: For example, encapsulating encrypted e-books in various formats with specific readers; encapsulating sliced streaming video files with specific players; encapsulating knowledge entity entries with RDF frameworks. (2) Packaging resource sets: For example, reorganizing resources in a specific domain (which may include e-books, audio/video, chapters, test questions, etc.) into a resource collection according to certain dimensional relationships, and packaging them as resource bundles according to agreed standards for use as underlying knowledge resources for front-end database applications.

Packaging refers not only to encapsulating resource files themselves but also to encapsulating resource metadata. After packaging, front-end applications do not directly access resources but uniformly access packaged resource objects in the middle platform. Providing services through packaged objects not only enables resource reuse across different front-end applications and facilitates unified management and maintenance of resource presentation methods and technical specifications in the middle platform, but also enhances resource access security since resources never leave the publisher.

2.2.5 Metadata Management Service Resource metadata obtained from the back-end typically contains only basic information required for management, with resource classification dimensions being relatively singular. To meet the diversified content publication and consumption needs of front-end applications, the middle platform can optimize and expand resource metadata. For exam-

ple, adding third-party e-commerce classifications for e-books; tagging video resources with keywords and adding professional classifications for their domains; establishing relationships between resources for front-end recommendation applications.

2.2.6 Interface Management Service Front-end applications access middle platform resources through a unified interface gateway, eliminating the need for manual resource uploads to different front-end applications. Through interface management, unified authorization management for front-end access to content resources can be achieved. Personalized settings can be configured for front-end application access scope, format, time limits, IP addresses, authentication methods, access frequency, and concurrent users, enabling refined resource authorization management.

2.2.7 Data Analysis Service The middle platform can centrally collect front-end user behavior data on content resource access (such as access frequency, duration, frequency, and evaluations) through data collection tools embedded in resource objects. After statistical analysis, this data is fed back to both front-end and back-end systems to drive front-end user experience improvements and back-end content product optimization. Real-time data analysis can also monitor and verify abnormal access behaviors.

2.3 Application Practice of Content Middle Platform

The content middle platform has been widely applied in CMP' s digital business operations. Several typical application cases are described below.

2.3.1 Flexible Distribution of E-book Resources Through Content Middle Platform CMP' s e-books need to be published not only on proprietary platforms but also on numerous third-party cooperative channels. Previously, this relied on manual file uploads, requiring e-book files to be processed and formatted according to third-party platform requirements before being copied to partners, with access and sales data not immediately available to the publisher. This approach was inefficient, had non-unified technical standards, and lacked security guarantees. After building the content middle platform, both internal and external front-end applications can access any required e-book resource format through unified middle platform interfaces, ensuring resources never leave the publisher, authorization is uniformly managed, and data is fully controlled. Currently, CMP has completed distribution integration with its self-built Engineering Technology Digital Library, as well as with Renrentian Shudian' s Changxiangzhixing platform and Zhongxin Jinqiao' s Kezhi platform.

2.3.2 Automated Video Transcoding and Multi-channel Publishing Through Content Middle Platform Similar to e-book resources, CMP' s vast video resources also need to be published on numerous internal and external

front-end applications. Different platforms and terminals have varying requirements for video format and resolution. Using the middle platform' s automatic conversion service, original video files are uniformly converted into three formats —MP4, HLS, and DASH (the latter two being fragmented streaming media formats)—with each format available in high, medium, and low resolutions. These are encapsulated with a custom-developed video player and uniformly provided to front-end applications through interfaces, achieving effects similar to e-book resource distribution. Currently, CMP has implemented video resource publishing integration with its self-built Equipment Manufacturing Video Library platform, print book QR code resource publishing platform, and several external platforms.

2.3.3 Content Middle Platform Supporting Marketing Promotion Activities During the COVID-19 pandemic, numerous offline marketing activities shifted online. CMP organized an online textbook exhibition requiring thousands of textbook PDF digital sample chapters to be prepared within a week. Using the content middle platform' s batch PDF processing capabilities and simple configuration of reader access permissions, the required digital sample chapters and information were quickly pushed to the online exhibition front-end application through interfaces, ensuring the event' s successful execution.

2.3.4 Rapid Development of Thematic Database Products Based on Content Middle Platform Leveraging the middle platform' s resource set packaging functionality, CMP has gained the capability to rapidly develop thematic database products based on current industry hotspots and market demands. For example, thematic databases developed using the middle platform model—such as robotics, UAV, and science and technology awards databases—have all received positive user feedback.

2.3.5 Content Middle Platform Driving Product Content Optimization Traditional print books operate in a one-way communication model, making it difficult to collect reader feedback. In the previous "front-end–back-end" architecture, user resource access behavior data was scattered across various front-end applications, limiting its holistic utility. After building the content middle platform, CMP has utilized its unified data collection and analysis capabilities to drive back-end product content optimization. For instance, analysis of user behavior data collected from front-end applications often reveals that many users show particular interest in specific domain content during certain periods, manifested by relatively longer dwell times on content related to that domain while browsing e-books and video products. CMP' s subsequent launch of detailed extended interpretation books or course products in these domains often achieves favorable market performance.

3. Conclusion and Outlook

The establishment of the content middle platform has helped CMP improve the efficiency of digital content resource distribution and the speed of front-end application iteration, reduced trial-and-error costs, strengthened protection of core content assets, optimized product content and user experience, and provided robust support for scientific, standardized, and efficient operational management of digital content resources. Simultaneously, the content middle platform requires continuous upgrading and improvement: on one hand, research is needed on how to coordinate with existing data middle platforms and business middle platforms; on the other hand, its core functions must be continuously enriched, particularly through strengthened application of new technologies such as blockchain, 5G, artificial intelligence, and VR/AR, forming a "content middle platform + technology middle platform" structure. Only such a content middle platform can empower traditional publishers to achieve truly deep integrated development.

References: [1] Ma Rundong. Zeyuan Software: Empowering the Middle Platform Capability of Publishing Houses [OL]. Shubohui, 2019-08-16. https://mp.weixin.qq.com/s/FV18FbJJiD4dQMCl_{AcBGg}. [2] Chen Xinyu, Luo Jiaying, Deng Tong, Jiang Wei, et al. Middle Platform Strategy: Middle Platform Construction and Digital Business [M]. Beijing: China Machine Press, 2019. 159-175.

Authors: Liu Kunxiang (born 1981), male, from Zhengzhou, Henan, Deputy Director (in charge) of Digital Technology R&D Center at China Machine Press, Associate Editor. Research interests: digital publishing. Miao Lijin (born 1976), male, from Wenzhou, Zhejiang, Deputy President of China Machine Press. Research interests: digital publishing and enterprise management. (Responsible editor: Chen Xuguan)

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