

Functional Evolution of Domestic Traditional Publishing Resource Management Systems (Post-print)

Authors: Bai Huarui

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

Abstract

The publishing resource management system is the most important digital platform for publishing houses, responsible for the management and application of core digital assets such as digital books, journals, images, audio, and video. This paper summarizes and organizes the evolution of domestic publishing resource management systems and the underlying logic by analyzing the changes in publishing resource management requirements.

Full Text

Preamble

For a long time, printed books served as the primary carrier of knowledge, with publishers responsible for its creation, editing, and distribution. With the development of computers and the internet, publishers have been swept into the wave of digitalization, whether actively or passively, constructing in-house information platforms to enable digital-based creation, editing, and dissemination. The publishing resource management system, abbreviated as “resource repository,” represents one of the most critical platforms in a publisher’s digital infrastructure, primarily addressing the collection, storage, retrieval, download, and application of digital resources. The Fangzhi Zhihui Publishing Resource Management System, in which the author participated in research and development, currently holds one of the highest market shares among resource repository products in China’s publishing industry. Drawing on years of industry experience, this article summarizes and analyzes the evolutionary path of publishing resource management systems.

Functional Description of Publishing Resource Repositories

While publishing resource repositories share similar functions with those in other industries—namely, unified collection, storage, management, and application of resources—they differ in their accommodation of the unique characteristics of digital resources in the publishing sector. At the business module level, publishing resource repositories primarily comprise six modules: resource management, copyright management, knowledge system management, data exchange, system management, and publishing application.

Resource management constitutes the core function of publishing resource repositories, enabling scientific and orderly management of omnimedia resources, including acquisition, processing, storage, conversion, publishing, and retrieval. It also provides resource access, management, and content-oriented analysis and conversion services for other application systems within the publishing house. The system offers fast and efficient resource access, a unified storage and management platform for publishing resources, and fragment-level content access, management, and permission control capabilities, fully supporting both unstructured and structured resource management.

Copyright management handles copyright information and contractual documentation for publishing resources. It tags relevant copyright information from resource libraries, uploads copyright contract documents, stores this information in a dedicated copyright database, and automatically establishes associations with corresponding resources. This module supports not only author copyright information management but also third-party authorized copyright registration and management, including expiration reminder functionality.

Knowledge system management builds upon high-quality publishing resources, utilizing constructed lexicons to perform large-scale structured, fragmented, and knowledge-based processing of content resources according to disciplinary characteristics. This enables deep knowledge revelation and forms new digital content products. Through association and recombination at the knowledge application unit level, it can rapidly generate and customize personalized product content, thereby providing readers with systematic, personalized, and efficient knowledge services. Knowledge system management should encompass keyword management, subject term management, and domain ontology management, with each submodule operating independently while establishing content associations.

Data exchange represents a platform-oriented feature of resource repositories. As the “resource pool” within a publishing house, the repository must support not only manual or offline data import and export but also integration with existing internal systems. It needs to connect with ERP systems, editorial systems, rich media production systems, and publishing systems to enable bidirectional data flow. This requires a dedicated data exchange module to support common integration methods, transmission protocols, and standard data formats, achieving configurable system integration.

System management comprises foundational functions for any digital system, and resource repositories are no exception. These include standard user, role, organizational, and permission management features. However, because resources represent a publishing house's core assets, publishing resource repositories demand higher standards for permission control and security protection than typical digital systems. Permission settings must be configured across multiple dimensions, such as branch departments, resource categories, and copyright tiers, with precision extending to specific fields and resource types. Additionally, custom entry database configuration represents a distinctive feature of publishing resource repositories and another core function of system management. Given the wide variety of books and journals, the structural fragmentation of publications in the knowledge service era has created diverse knowledge entries requiring different knowledge bases for storage—such as papers, encyclopedias, entries, regulations, cases, and Q&A formats. Consequently, the ability to freely define entry databases, configure their attributes, attachments, tags, and relationships with other resources constitutes the primary function of this module.

Publishing application serves as an extension module for resource utilization, primarily providing resource acquisition capabilities to users or systems. Its forms mainly include websites and open API platforms. Websites enable end-users to access resources through search, navigation, recommendation, browsing, and reading functionalities, while open API platforms facilitate third-party platforms to call and retrieve data. Access methods include authorized API calls and authorized page component calls, with the open API platform controlling third-party data access through temporal, frequency, and content-level restrictions.

The Resource Management Era

Publishing represents a traditional knowledge production and dissemination industry. Books and journals have long served as knowledge carriers. With the advent of the information age, their production processes shifted to professional typesetting software, ultimately generating electronic files for print and online distribution, primarily in PDF, Word, and EPUB formats. At the digital level, comprehensively describing and managing a book requires basic metadata, refined promotional information, cover files, main text files, illustration files, and associated multimedia attachments. These digital resource management needs shaped the initial functional form of publishing resource repositories.

Demand for domestic publishing resource repositories began to emerge broadly in the first decade of the 21st century, entering a rapid growth phase around 2010. National technical renovation projects in digital publishing launched around 2012 further propelled demand into an explosive period.

Repositories during this era primarily managed metadata and various electronic files for books and journals, with different publishing sectors developing specialized electronic file management capabilities—for instance, educational publish-

ers requiring lesson plan and courseware libraries, and electronic audio-visual publishers needing media asset libraries. Beyond file management, copyright management also emerged as a partial requirement, focusing on copyright information management and authorization information management. This period lacked an independent knowledge system module, with classification and keywords existing only as resource attribute fields.

Regarding data exchange, beyond manual batch upload and export applications, integration primarily involved ERP systems and digital reading websites. The main workflow consisted of ERP systems synchronizing book project information into the repository, collection personnel uploading attachment information, and website operators publishing e-book files to reading platforms.

In terms of publishing applications, some publishers required repositories to include an intranet publishing platform for internal user resource retrieval and acquisition, without external access.

The Knowledge Service Era

The arrival of the mobile internet era triggered substantial industry transformation. In knowledge services, this shift manifested as a move toward fragmented reading. Structuring books—refined carriers of authoritative knowledge—to provide professional knowledge services through fragmented knowledge became a new trend in publishing resource applications after 2014, with the state investing heavily in knowledge service project support beginning in 2015.

Fragmentation requirements for knowledge services created new functional demands for resource repositories in this era. Fragmentation processing, 入库解析 (repository parsing), entry storage, knowledge system construction, knowledge system annotation, micro-copyright management, and database production and publishing emerged as mainstream requirements.

Fragmentation processing generally extends international universal standards, such as the DocBook standard for books and JATS standard for journals, or industry-customized XML standards like the entry XML standard developed by Shanghai Lexicographical Publishing House. Large-scale adoption of these XML standards made structured processing and XML-based typesetting new innovation points in publishing technology. Structured processing addressed legacy book and journal resources, while XML-based typesetting served collaborative compilation platforms, promoting composite publishing—simultaneously enabling print publishing while generating structured digital resources for multi-channel distribution.

Fragmentation processing, repository intake, and storage created demand for processing workflow and upload client functionalities. Publishers assigned processing tasks to vendors, who accessed tasks through the platform, completed processing, and uploaded content via clients. The platform then performed content validation through schema and business rule checking before parsing and

importing into the repository.

Fragmented knowledge in the publishing industry is also called “entries.” Entry structures vary significantly across different resources, unlike previous publishing resources where books, journals, images, and audio-video files had relatively fixed structural fields and attachment types. Entry structures require separate definition based on publication type, necessitating that resource repositories provide custom entry database functionality. This requirement also drove evolution in the repository’s underlying technology architecture.

The massive volume of fragmented entries demands organized structure, as traditional classification systems cannot adequately support enhanced knowledge service experiences. Consequently, independent keywords, subject terms, and even domain ontologies emerged as new organizational methods for entries. How to rapidly, effectively, and cost-efficiently construct lexicons and knowledge graphs, and how to quickly and accurately annotate entries based on these lexicons, became technical challenges of this era. The industry generally adopted a combination of manual and intelligent-assisted methods for construction and annotation, bringing machine learning and artificial intelligence technologies into the application scenarios of publishing resource repository platforms.

Micro-copyright, relative to whole-book or whole-issue copyright, addresses how to determine and utilize copyrights after fragmentation. Entry applications must first undergo micro-copyright module verification before external service provision.

Building professional databases based on entries, images, audio-video, and other resources became the mainstream of knowledge services in this era. Standard functionalities for professional databases primarily integrated search capabilities combining faceted search, natural language analysis, and intelligent Q&A; resource navigation according to knowledge systems; reading and intelligent recommendation capabilities for knowledge entries; and interactive capabilities including personal and platform collections, notes, sharing, and annotations. In essence, these features enable users to acquire knowledge more precisely and conveniently.

The Data Middle Platform Era

Strictly speaking, the data middle platform era does not signify the end of the knowledge service era. During this stage, constructing knowledge services based on advantageous resources remains a key direction for publishers’ digital transformation. However, data middle platform-ization has indeed become the primary direction for continued advancement in resource repositories for traditional publishers since 2017.

After more than a decade of digital transformation, many publishers have built numerous systems and platforms. However, these systems remain fragmented, connected only through data exchange in serial fashion, resulting in functional

redundancy, cost waste, and operational difficulties. Furthermore, publishers have achieved few successful cases in operating websites or databases independently, partly due to insufficient operational capabilities. Consequently, as microservices, SaaS services, and middle platform architecture became mainstream technologies for platform construction, they also drove the evolution of resource repository middle platform-ization.

The data middle platform practices a philosophy of complete data openness, striving to eliminate data barriers between systems and build an ecosystem with unobstructed data flow [1]. Its core components are the data center and open API platform. The data center encompasses not only the publisher's own book and journal resources, structured entry resources, and knowledge systems but also industry resources, book reviews, and sales data collected from the internet. Intelligent analysis of this data provides support for various departments in topic selection analysis, sales analysis, and product development.

Simultaneously, because publishers' original resources often suffer from poor self-operation, they are typically entrusted to CNKI, Tongfang, and internet companies for operation, preventing publishers from controlling or obtaining user feedback. Therefore, the open API platform has become a primary function of the data middle platform. Through technical means, it authorizes data access to third-party platforms via refined permission settings and traffic control to achieve data delivery. Beyond API methods, embedded SDKs and H5 components can also be implemented to protect copyrights. For example, Shanghai Lexicographical Publishing House's Judian platform uses SDK embedding in apps to enable entry search and display.

Through the open API platform, publishers can control their resources while obtaining user usage feedback without directly operating knowledge service platforms themselves—an ideal approach for resource-rich publishers to achieve self-operation. Additionally, for large publishing houses and groups, the data middle platform provides data support for multiple internal systems, avoiding redundant storage and saving development and maintenance costs.

Due to large data volumes and high access frequency, the data middle platform has also driven another technological evolution in resource repositories, with cloud-based infrastructure, containerized cluster operations, and distributed storage becoming mainstream technologies.

From this 梳理 (analysis), we can observe that each evolution of resource repositories is essentially driven by advancements in computer technology, platform technology, and internet technology. Currently, as 5G technology, artificial intelligence, blockchain, and VR/AR technologies mature, they will drive another evolution in traditional publishing, and resource repositories, as foundational platforms, will undoubtedly be reshaped by these trends.

[1] Liu Changming, Lu Lan, Xu Jian. “New Infrastructure” Middle Platform Capabilities Support Smart Media Development [J]. *China Media Technology*, 2021(2): 10-13.

Author Profile: Bai Huarui (1985-), male, from Luoyang, Henan, holds a master's degree and serves as Development Director at Beijing Founder Electronics Co., Ltd. Research interests: knowledge services.

(责任编辑: Chen Xuguan)

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

Source: ChinaXiv – Machine translation. Verify with original.