# Construction of a Unified Resource Center for Publishing Units Based on the Middle Platform Concept: A Case Study of Higher Education Press (Postprint)

**Authors:** Mao Fang

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

## Abstract

With the integrated development of the publishing industry entering the stage of deep integration, unified management and distribution of all-media content resources has become a pressing requirement. This paper presents the background and significance of constructing a unified resource center for publishing organizations based on the middle platform concept, the implementation pathway of "Three Unifications and Five Centers," and key considerations for implementation.

## Full Text

## Preamble

**Title:** Building a Unified Resource Center for Publishing Houses Based on the Middle Platform Concept: A Case Study of Higher Education Press
**Author:** Higher Education Press Co., Ltd., Beijing 100029

**Abstract:** As the integrated development of the publishing industry enters a stage of deep convergence, the unified management and distribution of all-media content resources has become an urgent necessity. This paper introduces the background and significance of constructing a unified resource center for publishing houses based on the middle platform concept, the implementation path of "Three Unifications and Five Centers," and key considerations for implementation.

**Keywords:** middle platform; unified resource center; educational publishing; deep convergence; data standards
**CLC Number:** TN976
**Document Code:** A

**Article ID:** 1671-0134(2021)06-129-04
**DOI:** 10.19483/j.cnki.11-4653/n.2021.06.041
**Citation Format:** Mao Fang. Building a Unified Resource Center for Publishing Houses Based on the Middle Platform Concept: A Case Study of Higher Education Press [J]. China Media Technology, 2021(06).

## 1. What is a Middle Platform?

A middle platform is an enterprise internet architecture that corresponds to front-end and back-end systems, referring to a collection of shared middleware within a system. The middle platform can be divided into two critical dimensions based on different functions and roles: (1) **Business Middle Platform**, which constructs componentized, service-oriented capabilities with high cohesion, low coupling, and high reusability, empowering agile development of front-end applications and enabling shared business capabilities and integrated business processes across the entire network. (2) **Data Middle Platform**, which focuses on data aggregation, data governance, data modeling, and data services.

## 2. Background and Significance of Middle Platform Application in Publishing Houses

Content resources are the core assets of publishing houses. The construction, management, and application of content resources constitute the core work of publishing houses and form the foundation for their integrated development. After years of digitalization efforts, publishing houses have accumulated substantial digital resources and developed numerous websites, platforms, and applications to meet various needs. However, alongside these achievements, publishing houses face several challenges: platforms have been built independently, making it difficult to effectively integrate and share resource data, operational data, and user behavior data; functional modules are redundantly developed; asset management remains extensive; data quality is uneven; file versions are chaotic; and existing technology development models lack competitiveness in rapidly responding to user needs and iterative system upgrades. For publishing houses or groups with multiple business systems, utilizing a middle platform to integrate and aggregate content resources, conduct standardized, asset-based, and intelligent management, build a comprehensive and rational digital asset control and application system, and establish a public capability reuse middle platform that supports rapid iteration of various business platforms is an inevitable requirement for promoting deep media convergence and achieving high-quality development.

The overall objective of building a unified resource center for publishing houses is to construct a data middle platform and business middle platform, integrate traditional and emerging media, and build a "new ecosystem" for digital publishing and digital education. The overall system design is shown in [Figure 1:

see original paper].

The unified resource center is built on a dual middle platform architecture of "data middle platform + business middle platform,"with a focus on implementing the "Three Unifications and Five Centers."

## 3. Implementation Path

### 3.1 What to Unify?

**3.1.1 Unified Data Standards** The "unification" of a unified resource center begins with unified data standards, which can be implemented at three practical levels: data models, metadata standards, and classification standards.

#### 3.1.1.1 Data Models

Data models should be established for various products and other data, including data structure (required components, optional components), data operations, personnel roles, and data constraints. Reference standards for data models in the educational publishing field include the "Central Cultural Enterprises Digital Transformation and Upgrading" project standard *GCZX 14—2014 Digital Resource Content Indexing Rules for Books*, national standards such as *GB/T 36642-2018 Information Technology—Learning, Education and Training—Online Courses*, *GB/T 29802-2013 Information Technology—Learning, Education and Training—Test Item Information Model*, *GB/T 36453-2018 Information Technology—Learning, Education and Training—E-textbook Information Model*, *GB/T 36098-2018 Information Technology—Learning, Education and Training—Virtual Experiment Component Packaging*, and the international EPUB standard. However, these standards are insufficient on their own; publishing houses need to modify, improve, and supplement missing standards based on these to build an enterprise data standard system.

#### 3.1.1.2 Multidimensional Metadata

Metadata standards should be established for various types of data (including 24 categories such as materials and products), encompassing dozens of metadata elements across multiple dimensions: general information, technical information, educational information, copyright information, relational information, lifecycle information, and intelligent proofreading information (see ). Reference standards for metadata include *GB/T 21365-2008 Information Technology—Learning, Education and Training—Learning Object Metadata*, *JY/T 0607-2017 Basic Education Teaching Resource Metadata*, *CY/T 97—2013 E-book Metadata*, *GB/T 36642-2018 Information Technology—Learning, Education and Training —Online Courses*, *GB/T 35427-2017 Core Metadata for Book Copyright Assets*, *GB/T 36097-2018 Information Technology—Learning, Education and Training—Virtual Experiment Component Metadata*, and *GB/T 35430-2017 Information and Documentation—Descriptive Metadata Elements for Periodicals*. Publishing houses similarly need to integrate and refine these standards to form applicable enterprise standards.

**Table 1: Composition of Educational Resource Metadata**

| Metadata Category | Elements |
|---|---|
| General Information | Title, creator, creation time, identifier, language, description, etc. |
| Technical Information | Format, size, duration, dimensions, resolution, location, etc. |
| Educational Information | Subject specialty, target audience, resource type, knowledge points, etc. |
| Copyright Information | Copyright holder, licensing method, rights type, rights restrictions, etc. |
| Relational Information | Belongs to, contains, different versions, sequence, matching, etc. |
| Lifecycle Information | Version, status, contributor, role (revision, review), etc. |
| Intelligent Proofreading | Intelligent review results |

#### 3.1.1.3 Various Classification Data Dictionaries

An educational publishing resource classification system should be established to form data dictionaries that automatically update and synchronize globally across all business platforms within the publishing house. Classification systems in higher education publishing include undergraduate major catalogs, vocational college major catalogs, course types, resource types, languages, image categories, and key publisher projects. Reference standards include *National Standards for Teaching Quality of Undergraduate Major Categories*, *Technical Specifications for Modern Distance Education Resource Construction*, *GB/T 28825-2012 Classification Code for Learning Objects*, *GCFH 4—2017 Resource Type Classification and Code*, and *TCADPA Digital Content Resource Classification and Code*.

### 3.1.2 Unified Identification and Entities 3.1.2.1 Unified Identification for Various Resources and Relationships

Reference standards for educational publishing content identification include *CY/T 81—2012 Unique Identifier for Press and Publication Digital Resources*, the National Digital Composite Publishing Engineering standard *GCFH 5 Resource Identification Application Specification*, *CY/T 126—2015 Digital Copyright Identifier (DCI)*, and the international standard *GB/T 32867-2016 ISLI (International Standard Link Identifier)*. Based on unified standards, a globally

unique identification system should be established to implement full lifecycle data management, ensuring data consistency and timely updates during cross-platform circulation through systematic unique identifiers, thereby forming an efficient dynamic asset management mechanism.

#### 3.1.2.2 Entity Deduplication

Based on unique identification, resource entity file deduplication must also be achieved. All new resources use a composite primary key of ID and MD5 value to ensure uniqueness. When a resource with the same MD5 value is detected in the system, it is automatically deleted or flagged as duplicate. When users modify resources online and version changes occur, the number of MD5 values under the same ID can be used to track resource update frequency and other data.

### 3.1.3 Unified Interfaces and Process Specifications (1) Unified Data Exchange and Interface Specifications

This includes data reception standards and data output standards, reflecting the positioning of the unified resource center as a data middle platform.

**(2) Unified Content Resource Production and Processing Workflow**

Under the all-media publishing concept, workflow reengineering should adapt to multiple product forms, publishing formats, and distribution terminals, supporting both online and offline production modes. Flexible workflow stages and personnel role permissions should be designed and 沉淀为业务中台 (crystallized as shared capabilities of the business middle platform) for output to various online operation platforms.

## 3.2 What Centers?

**3.2.1 Asset Center** In enterprise accounting systems, assets refer to resources formed by past transactions or events, owned or controlled by the enterprise, and expected to generate economic benefits. Press and publication units possess substantial digital content for publishing or information services, representing important assets. Establishing an effective content asset management model to improve asset efficiency is a key consideration for publishing houses. The crucial distinction between traditional content management systems and unified resource centers is the shift from resource management to asset management. This approach manages resources—including traditional media assets and new media resources scattered across various operation platforms—as assets, enabling content resources to demonstrate value independent of their carriers and facilitating the transformation of publishing from manufacturing to content and internet industries.

The primary requirement for content resource asset management is clear ownership. As media convergence deepens, copyright situations become increasingly complex, highlighting the importance of copyright management. Integrating copyright management into content resource management—rather than

establishing a separate copyright management system—is a more appropriate approach, linking later-stage content with earlier-stage topic selection and contract data to enable automatic annotation of digital content resource copyright information.

The core of content resource asset management is establishing a full lifecycle management mechanism for content assets. The entire chain—from topic planning, manuscript submission, contract signing, review and processing, to publication, application, and revision—is no longer fragmented but seamlessly integrated. Based on full lifecycle data, content assets undergo recognition, procurement, filing, processing, uploading, importation, review, acceptance, warehousing, usage, updating, allocation, auditing, evaluation, accounting, impairment, and inventory, enabling accurate grasp of digital asset quantity and value from a business perspective.

The goal of content resource asset management is maximizing asset value. Through the unified resource center' s identification and full lifecycle data tracking of content assets, content resources can be discovered and utilized. Multi-channel distribution fully taps potential, making resource usage, circulation, and revenue transparent and motivating departments to explore and utilize content assets. Meanwhile, the unified resource center supports requirements for opening, interfacing, and integrating content assets, supports introduction and output of content resources, integration and being integrated, supports transfer, sale, and write-off of asset usage and ownership rights, and supports full-process, cross-media, cross-platform tracking management, copyright protection, and circulation transactions to maximize asset value.

**3.2.2 Data Center** A key objective of the unified resource center is data aggregation to eliminate data silos, building a unified data middle platform for the entire publishing house and gradually evolving it into a big data center for the publishing unit, fulfilling both basic resource and innovation engine functions. The data center manifests in five main aspects:

**Extensive Data Scope.** The unified resource center will integrate data from all content systems and operation platforms, including portal websites, digital libraries, QR code service platforms, iExercise, digital course platforms, academic frontier online, new-form textbook networks, smart vocational education, college student online platforms, online malls, and data from relevant external third-party distribution channels.

**Comprehensive Data Types.** The unified resource center accesses both traditional print book content and digital media content; both structured and unstructured data; both static content data and dynamic user interaction data; and both internal data and externally crawled data. Overall, this includes over a dozen types of product and material data, author data, topic selection data, contract data, conference data, information data, internet data, system log data, user behavior data, customer service data, operational data, and tag data.

**Diverse Data Access Methods.** Data access to the unified resource center primarily occurs through three methods: API interfaces, log files, and Dataworks data tools. Various application systems can choose appropriate access methods based on their circumstances while maintaining dynamic data synchronization.

**Intelligent Data Retrieval.** The unified resource center provides global intelligent search functionality, enabling cross-platform precise retrieval and full-text search of text content, which represents a significant aspect of its role as a global data center.

**Multidimensional Data Dashboards.** Unified resource center dashboards serve both management decision-making and comprehensive analysis needs, enabling data asset inventory, asset value visualization, and data-driven, intelligent, real-time decision-making and supervision to identify and grasp development trends. They also meet data viewing needs of different personnel types, including content asset managers, frontline business staff, operation monitoring personnel, and external partners.

**3.2.3 Capability Sharing Center** The business middle platform of the unified resource center focuses on providing public foundational tools and services around the content production process, strengthening technology integration, avoiding redundant development, and agilely and rapidly adapting to the diversified and personalized online education market. Key capabilities include:

**Resource Upload and Processing.** Provides public resource upload interfaces and customizes resource processing templates with different parameters based on business system needs, including storage, transcoding, multi-bitrate compression, slicing, watermarking, and encryption.

**AI Capabilities.** Establishes unified AI capabilities providing natural language processing, image recognition, speech recognition and synthesis, and other services for human-machine collaborative assisted production to improve efficiency, optimize workflows, and ensure product quality. Specific applications include ideological screening, intelligent tagging, intelligent summarization, text intelligent proofreading, automatic subtitles, intelligent segmentation, and relationship building, with ongoing training of recognition models tailored to publishing house content characteristics and requirements.

**Workflow Customization.** The unified resource center provides content production, review, and publication workflow customization functions, including stage transitions and associated departments, roles, and personnel.

**3.2.4 Distribution Center** The unified resource center serves as the digital resource distribution hub for the entire publishing house, providing content distribution services to various operation platforms and enabling unified distribution and free organization from products to commodities. Content distribution through the unified resource center allows accurate understanding of access

data from various operation terminals, enabling unified copyright protection and traffic monitoring to provide credible data for resource co-construction and sharing.

**3.2.5 Resource Aggregation and Rapid Knowledge Base Production Center** Due to its aggregated massive resources and automated/manual annotation based on disciplinary knowledge systems, the unified resource center can achieve knowledge aggregation—such as aggregating and displaying knowledge by specialty, position, or domain to form thematic resource libraries with customizable configuration and dynamic updates. Thus, the unified resource center becomes a resource aggregation and rapid knowledge base production center, meeting personalized knowledge service needs of operation terminals.

## 4. Implementation Considerations

### 4.1 Overall Planning with Step-by-Step Implementation

The unified resource center is a system engineering project involving the entire organization, requiring overall planning and phased implementation. Taking the Higher Education Press unified resource center as an example: Phase I focused on initializing the global architecture, building the data public layer, initializing key data applications, completing unified user center integration and first-batch platform data access, and supporting the construction of a second-generation new-form textbook network based on the unified resource center, with preliminary completion of the data standard system. Phase II continues to expand the breadth and depth of data access, optimize data annotation, retrieval, tracking, and dashboards, refine data standards and models, iteratively optimize the middle platform architecture, and enrich middle platform functions to cover major businesses with both data and business middle platforms. Phase III strengthens applications, enhances AI capabilities, expands application scenarios to empower new products and services, with system standardization and intelligence reaching maturity and the basic resource and innovation engine roles of data being fully realized.

### 4.2 Complex Data Integration

Data integration is the cornerstone of the unified resource center, where completeness, quality, reliability, and implementability are all critical. Taking Higher Education Press as an example, the organization has numerous existing business systems with rich business forms, databases involving MySQL, MongoDB, and SQL Server, large data volumes (some single tables exceeding ten million records), some databases lacking timestamps reflecting content changes, complex deployment environments involving multiple public clouds, private clouds, and third-party hosting facilities, and business systems supported by diverse third-party technical teams. This complexity poses challenges for data integration, requiring flexible responses based on business realities, balancing

current needs with long-term goals, and selecting appropriate data access methods and synchronization tools.

### 4.3 Empowerment Rather Than Control

Empowerment rather than control is a principle that should guide middle platform construction. The middle platform exists for the front-end; its purpose is to better serve the front-end, enabling enterprises to continuously align their capabilities with user needs and smoothly channel back-end resources to users in response to demand. A common pitfall to avoid during middle platform construction is considering the middle platform superior to the front-end, where the middle platform establishes all rules, processes, and systems for the front-end to simply follow. The success of a middle platform strategy is not measured by technological advancement or converted capabilities, but primarily by whether the front-end wants to use it, likes to use it, and finds it easy to use. Therefore, middle platform construction should pay attention to coordinating relationships with the front-end, emphasizing empowerment over control.

### 4.4 In-House Technical Capability Support

The construction of a unified resource center involves numerous system integrations and is not accomplished overnight, requiring gradual iteration and continuous upgrading to reach maturity. Relying entirely on external technical capabilities can be passive during this process; having certain in-house technical support is highly beneficial for system design, customization, optimization, and progress control. In-house technical teams can integrate new systems using established data access methods, adjust and improve existing data models, and perform independent system iteration and maintenance, significantly improving project success rates.

## Conclusion

Traditional publishing houses face many difficulties and risks in digital publishing, but not developing represents the greatest difficulty and risk. The path of integrated development must be persisted despite its challenges. The middle platform is a new concept for publishing houses, and building a unified resource center based on the middle platform concept is a choice made when publishing houses' integrated development reaches a certain stage. It is also a comprehensive system engineering project and strategic support for achieving the integration and unification of traditional and emerging publishing. To realize the effectiveness of deep integration beyond the technical architecture requires a series of supporting enterprise standards, regulations, process norms, and institutional mechanisms. Both aspects are indispensable, mutually reinforcing and dependent, jointly promoting practical and profound integrated development and transforming integration advantages into development effectiveness.

**References:** [1] Shen Jie. Construction of a Unified Resource Management

Center for Educational Publishing Transformation and Upgrading [J]. China Editor. 2020(6): 64-68.
[2] General Office of the CPC Central Committee, General Office of the State Council. Opinions on Accelerating the In-depth Integrated Development of Media [EB/OL]. [2021-02-05]. http://www.gov.cn/zhengce/2020-09/26/content_{5547310}.htm.
[3] Publishing Resource Database. Overview of China' s Press and Publication Standardization Technical Organizations and Standards [EB/OL]. (2021-02-01). https://mp.weixin.qq.com/s/uJH7LzBtdMkVTIANySHm_g.

**Author Biography:** Mao Fang (1981-), female, from Henan, research interests: digital publishing, integrated development, standards, educational informatization.

**(Responsible Editor: Zhang Xiaojing)**

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