Constructing Digital Special Collections in the Age of Big Data: A Case Study of the Chinese Manuscript Literature Digital Resource Database (Postprint)

Authors: Han Songtao, Huang Chen

Date: 2023-11-07T10:52:02+00:00

Abstract

[Purpose/Significance] This paper aims to explore the various issues in digital special collections construction in the big data era through a case study of the Chinese manuscript documents digital resource database. [Method/Process] This paper investigates the specific processing methods for constructing multi-source, multi-format manuscript document resources, clarifies the construction objectives with academic innovation as the core requirement, and explores sustainable development paths, thereby summarizing the model for digital special collections construction. [Results/Conclusion] This paper argues that relying on core resource holding institutions, conducting collaborative resource construction, completing global integration of certain types of documents, and providing user-friendly academic tools constitute the main path and model for constructing academic digital special collections.

Full Text

Preamble

Journal of Literature and Data Science, Vol. 5, No. 3, September 2023

Exploration of Digital Special Collection Construction in the Big Data Era: A Case Study of the Chinese Manuscript Documents Digital Resource Library

Han Songtao[1], Huang Chen[1],[2],[3]
(1. Zhejiang University Library, Hangzhou 310027, China;
2. Center for Information Resources Analysis & Application of Zhejiang University, Hangzhou 310027, China;

3. Administration Center for China Academic Digital Associative Library, Hangzhou 310027, China)

Abstract:
[Purpose/Significance] This paper examines the construction of the Chinese Manuscript Documents Digital Resource Library as a case study to explore the challenges facing digital special collection development in the big data era. [Method/Process] The study investigates specific approaches for processing multi-source, multi-format manuscript resources, clarifies development goals centered on academic innovation needs, and explores pathways for sustainable development, thereby summarizing a model for digital special collection construction. [Result/Conclusion] The paper argues that the primary pathway and model for building academic digital special collections involves relying on core collection institutions, achieving resource co-construction through collaboration, completing global integration of specific document types, and providing convenient academic tools.

Keywords: Digital special collection; Manuscript documents; Crowdsourced co-construction
Classification Numbers: G256; K877; G250.74
DOI: 10.31193/SSAP.J.ISSN.2096-6695.2023.03.04

---

1. Background Review

Traditional special collection digitization has typically relied on institutional holdings, aligned with disciplinary development, and constructed databases through project-based initiatives. This approach often suffers from incomplete resources, rudimentary technical support, inadequate long-term maintenance, and low utilization rates, easily creating numerous resource silos. For years, the library community has sought viable solutions to these dilemmas, yet effective breakthroughs have remained elusive.

In early 2022, with funding from the National Ancient Books Digitization Project Special Fund of the National Ancient Books Consolidation and Publication Planning Office, Zhejiang University launched the "Chinese Manuscript Documents Digital Resource Library" construction project. As a major initiative included in the national ancient books digitization "14th Five-Year Plan," the project team faced urgent questions regarding top-level design: how to maximize the integration of Chinese manuscript resources worldwide, accommodate scholars' and the public's usage habits and needs, and create an academic research space based on manuscript resources. To address these challenges, the project engaged Professor Zhang Yongquan, a senior professor of humanities at Zhejiang University and renowned Dunhuang scholar, as chief academic expert to guide planning for resource collection, organization, service utilization, and research support. The project portal (https://xieben.cadal.edu.cn/) went live in June 2022, receiving dedicated

coverage from *Guangming Daily* [1]. Phase I of the project passed acceptance review organized by the National Ancient Books Consolidation and Publication Planning Office in November 2022.

The construction of the Chinese Manuscript Documents Digital Resource Library differs significantly from traditional special collection databases. Although only Phase I objectives have been completed, the concepts, planning, organization, and technical architecture developed during the process offer valuable points for discussion and reference. This paper uses the Chinese Manuscript Documents Digital Resource Library project as a case study to explore digital special collection resource construction in the big data era, aiming to propose a sustainable development model for resource-based academic research platforms in the digital humanities context, providing reference for academic and professional peers.

2. Multi-source, Multi-format Resources through Crowdsourced Co-construction

Manuscript documents refer to ancient books or textual materials written with soft or hard brushes on paper, primarily prevalent from the Eastern Han to Five Dynasties period, serving as the main carrier of Chinese civilization during this era. After the Northern Song Dynasty, as woodblock printing technology matured, printed editions gradually replaced manuscripts as the dominant format. The popularity of printed editions both accelerated the disappearance of early ancient manuscripts and shaped a manuscript system centered on contract documents, legal archives, folk religious texts, account notes, and popular literary works. Since the late Qing Dynasty, scientists and explorers both domestic and foreign have discovered some early manuscript documents in Gansu, Xinjiang, and Shaanxi, including ancient maps from the Western Han period and Jin Dynasty manuscripts of *Strategies of the Warring States* and *Records of the Three Kingdoms*, though these remain limited in number.

On June 22, 1900, the Dunhuang Mogao Grottoes Library Cave was opened, revealing a large number of manuscripts from the Tang Dynasty and earlier, shocking the world. Subsequently, during the Republican era, numerous manuscript collections including Turpan documents, Khara-Khoto manuscripts, local documents from the Song-Yuan period, and Ming-Qing archives were gradually made public, dramatically increasing the quantity of manuscript documents and returning them to scholarly attention [2].

2.1 Collection of Multi-source, Multi-format Special Collection Resources

Manuscript documents encompass numerous types and are widely distributed, creating significant uncertainties for resource collection. Taking Dunhuang manuscripts, Turpan documents, and local documents as examples, each cat-

egory exhibits distribution fragmentation, with images existing in multiple formats.

2.1.1 Dunhuang Manuscripts Dunhuang manuscripts are primarily held by four institutions: the National Library of China, the British Library, the Bibliothèque nationale de France, and the Institute of Oriental Manuscripts of the Russian Academy of Sciences. Among them, Chinese collections contain approximately 16,000 items, Russian collections about 19,000 items (including fragments), British collections about 14,000 items, and French collections about 8,100 items. The French and British portions are the most content-rich and research-valuable. Additionally, numerous other public and private institutions hold collections, with smaller quantities in Japan, the United States, Germany, Denmark, Sweden, and other countries [3]. Their images exist in formats corresponding to technological eras: early black-and-white film, later grayscale and color publications, and recent color digital images.

2.1.2 Turpan Documents Approximately 40,000 Turpan documents exist worldwide, scattered across various locations. Chinese institutions hold about 12,000 items, distributed among the Xinjiang Uyghur Autonomous Region Museum, the National Museum of China, Peking University Library, Shanghai Library, and Liaoning Provincial Archives, among others. Foreign collections reside at the German National Library, Ryukoku University Library in Japan, the St. Petersburg Branch of the Institute of Oriental Studies of the Russian Academy of Sciences, Helsinki University Library in Finland, the British Library, Istanbul University Library, and Princeton University Library in the United States [4]. Beyond these major collections, scattered holdings are numerous; for instance, Zhejiang University Library holds a small number of Turpan documents from the Gaochang Kingdom period. Research on Turpan documents started relatively late, with publications released sporadically and primarily in book form.

2.1.3 Local Documents Local documents are even more fragmented. The famous Huizhou documents, as the largest known collection of folk literature, were discovered in the 1950s during the collection and sale of ancient books at the Tunxi Ancient Books Bookstore. Subsequently, institutions such as the Chinese Academy of Sciences began large-scale acquisition, initiating the mass collection of folk documents [5]. According to incomplete statistics, over 800,000 Huizhou documents are held by institutions and individuals worldwide, with Sun Yat-sen University Library holding approximately 338,000 items—making it an exemplary local document collection in terms of quantity, temporal span, content richness, and systematic record-keeping [6]. However, published books have only released tens of thousands of items, with most Huizhou documents still stored in major public collections.

Similarly, Zhejiang local document collections are scattered. Zhejiang Normal University holds about 100,000 items across its library and Center for Exca-

vated Documents. Zhejiang University Library holds nearly 10,000 items, with its collection continuing to grow. Zhejiang University City College has also begun collecting local documents, with holdings exceeding 10,000 items. These public collections were primarily gathered from folk sources. Public institutions also hold local documents preserved as archives; for example, Longquan in Zhejiang houses a batch of judicial archives stored in the city's archives [7], and the famous Lanxi Fish Scale Registers are held in the Lanxi City Archives [8]. Personal collections also exist, such as the Shicang Contract Museum established by an individual, which holds large quantities of contract documents. This demonstrates that local document collections involve numerous institutions, making integration challenging. Except for publications like the 8,000-item Shicang contract collection, Longquan judicial archives, and Lanxi Fish Scale Registers, most other Zhejiang local documents remain in early stages of organization, requiring multi-party collaboration for database collection.

2.2 Organization of Special Collection Resources

Such vast, fragmented, and diverse manuscript documents urgently require standardized organization during collection. Classification and indexing are crucial methods for resource organization. The Chinese Manuscript Documents Digital Resource Library established classification and metadata standards for collection data from the outset.

2.2.1 Classification of the Chinese Manuscript Documents Digital Resource Library Classification serves browsing convenience and enables de facto inter-document association. The library employs parallel geographical and content-based classification systems.

Geographical classification recognizes that newly discovered or published manuscripts from specific regions have demonstrably driven distinctive academic research, emphasizing macro-level comprehensiveness. For original resources, manuscripts are classified by region, considering global influence, initially divided into three categories: Turpan documents, Dunhuang manuscripts, and Khara-Khoto documents. Turpan documents refer to manuscripts discovered in Jin-Tang ancient tomb complexes in the Turpan region of Xinjiang since the late 19th century, representing the primary physical remains of paper documents from the Wei-Jin and Six Dynasties period. Dunhuang manuscripts primarily refer to manuscripts from around the Tang Dynasty and earlier discovered in the Dunhuang Mogao Grottoes Library Cave, plus a small number of printed documents. The copying period spans over 600 years from the Wei-Jin Six Dynasties to the early Song Dynasty, focusing on the Tang and Five Dynasties as the main physical remains of paper documents from that era, succeeding Turpan documents and preceding Song-Yuan printed editions and later manuscripts. Khara-Khoto documents refer to paper manuscripts and a few printed documents discovered at the Khara-Khoto site in Ejin Banner, Inner Mongolia. These documents succeed Dunhuang manuscripts,

with copying and printing dates spanning the Northern Song, Liao, Jin, Western Xia, Yuan, and Northern Yuan periods, primarily in Western Xia and Chinese scripts, covering traditional classics, history, philosophy, and literature, as well as Buddhist and Taoist scriptures, contract documents, and official archives—precious materials for studying this period, particularly the Western Xia Dynasty.

Content-based classification initially divides into four categories: local documents, Ming-Qing archives, folk religious manuscripts, and folk opera-novel manuscripts. Local documents, relative to official archival documents, refer to Song to Republican period folk documents primarily handwritten and discovered over the past century. Ming-Qing archives refer to documents from imperial courts and government agencies at all levels, with approximately 20 million items extant. Folk religious manuscripts refer to copied scriptures of dozens of folk religious sects popular among ancient commoners, possessing multiple values in folklore and religious studies and representing an important component of traditional Chinese culture. Folk opera-novel manuscripts mainly refer to opera scripts, musical scores, opera theory, storytelling texts, and miscellaneous drama stories preserved through handwritten copies, forming a category corresponding to printed editions.

These manuscript documents, viewed chronologically from the Wei-Jin Six Dynasties through the Qing Dynasty, provide continuous resources reflecting literature, history, thought, folklore, and religion across these historical periods. The above represents first-level classification, with second and third levels beneath; for instance, local documents contain 13 second-level and 47 third-level categories.

2.2.2 Indexing of the Chinese Manuscript Documents Digital Resource Library Resource indexing primarily employs metadata, with the initial principle of enabling mapping to Dublin Core standards. Using manuscript metadata as an example, the 15 Dublin Core elements are divided into three major categories—content, form, and collection—corresponding to relevant manuscript content. Some elements have been expanded; for example, "title" is expanded into self-assigned title and original title, since most manuscripts lack original titles, necessitating self-assigned titles with established rules. "Date" is expanded into three elements: dynasty, Chinese chronological era (reign title), and Gregorian calendar year. This serves three purposes: objective description, precise search points, and navigation functionality.

Resource indexing can also employ open tagging. However, single-dimensional tags cannot effectively serve academic resource classification. Tags from different dimensions used on the same "plane" cannot replace systematic classification. We note that Ranganathan's *Colon Classification* (7th edition) is essentially a "faceted classification system." "Tags and faceted classification share fundamental similarities; faceted classification is essentially multi-dimensional

(or multi-faceted) tags. Using dimensional concepts to standardize tags while maintaining their freedom within dimensions is the primary method for enabling tags to function as academic classification tools" [9]. Multi-dimensional tagging will also be an open resource indexing method for the Chinese Manuscript Documents Digital Resource Library.

2.3 Sustainable Development through Crowdsourced Co-construction

No single collecting institution can exhaustively cover literature resources in a given field. Given the wide-ranging sources, diverse formats, and complex indexing of manuscript documents, constructing such a resource library through single-institution efforts would result in only a selective, illustrative special collection with limited practical value. Building on collection strengths, co-construction and sharing represent important pathways. The key challenge for crowdsourced special collection construction involves integrating resources from the physical to pictorial, published, and even video formats, achieving seamless connection based on cyberspace integration.

Therefore, we incorporated open crowdsourcing functionality requirements into the special collection platform construction. On one hand, the system allows resource integration from different public and private collections through backend crawling and frontend editing, with source attribution and clear rights control. We designed an extensible metadata framework enabling online addition of descriptions for different sources of the same resource, as well as addition of texts, images, URLs, and even research papers. On the other hand, the system permits users to upload images, add annotations, and correct metadata, which becomes new content for the resource library after administrator review. Additionally, we are developing APIs and interface standards to enable interested partner institutions to implement linked data through interfaces, revealing maximum resources to users.

Research is fundamentally a cumulative process building on predecessors' achievements—the original form of crowdsourcing. The internet enables this diachronic process to unfold both synchronously and diachronously. Introducing crowdsourcing mechanisms into special collection construction allows different institutions and individuals to collaborate online, continuously improving resource completeness, perfecting resource description and organization frameworks, and tightly connecting resource services with research, effectively addressing sustainable development challenges for special collection resource libraries.

3. From Resource Services to Research Support

Traditional document services focus on discoverability and accessibility, remaining external to users' learning and research processes whether on-site or online. How to support and enhance resource services through platform technical functional modules, truly embedding them into users' academic life cycles, represents

an important hallmark of next-generation special collection resource libraries.

3.1 Virtual Integration and Associated Reconstruction

Due to multi-source characteristics, functions such as multi-source display and comparison, and multi-dimensional association revelation beyond single-document cataloging, have become important requirements for the Chinese Manuscript Documents Digital Resource Library. The resource library aims beyond simple document presentation, providing tools for users to perform operational tasks that reveal resource associations while showcasing users' discoveries and academic achievements, thereby increasing user engagement.

3.1.1 Multi-source Data Display and Multi-resource Association Revelation Documents are cataloged by item, yet single items often have multiple copies from different sources. Taking Dunhuang manuscripts as an example, a single Dunhuang scroll has produced various preservation outcomes due to historical changes and technological development, including early microfilm (mostly black-and-white images), book publications, and later color digital images from books and scanning. While later technology is generally considered more advanced with better image clarity, preserving only the most recent scans might seem sufficient. However, different preservation outcomes span long time periods. Dunhuang manuscripts primarily consist of Tang Dynasty copies, with the latest dating to the Song Dynasty—over a thousand years ago. Each photography or scanning session potentially damages the documents. Although early reproduction technology was relatively primitive, the completeness of early manuscripts was best. Moreover, each reproduction could involve "accidents," such as omitting corner positions or neglecting single symbols on verso sides. Therefore, for complete manuscript collection, gathering all numbered documents does not equal comprehensive collection; only by including all historical reproduction outcomes can data integrity be ensured. The Chinese Manuscript Documents Digital Resource Library has collected and indexed these materials, supporting multi-source display and providing comparison functions for images from different sources (see Figure 1 [Figure 1: see original paper]). Considering copyright and accessibility factors, some data are displayed only through URL access.

Furthermore, multiple manuscript documents have interconnections. Presenting them as single items in a database cannot fully reveal their 内涵 (connotations); their associations must be simultaneously revealed for optimal user utilization. We catalog by single item while providing combination functions, using virtual integration to allow individually cataloged documents to join groups, satisfying both requirements. To this end, we established two virtual special topics: Association and Group.

Association topics place different period manuscripts of the same document or document type together in a certain form, establishing interconnections to facilitate comparative research. For example, *The Analects* was a Tang Dynasty

"textbook." Different copies copied by students have been unearthed in both Dunhuang and Turpan, including at least several commentaries such as *Zheng's Commentary on the Analects*, *Collected Explanations of the Analects*, *Meaning and Order of the Analects*, and *Pronunciation and Meaning of the Analects*. Based on scanned original manuscripts, an "*Analects* Association Topic" could be established, presenting all discovered manuscripts from Dunhuang and Turpan together and establishing mutual connections. This would maximize the academic value of these manuscripts, compensate for gaps in existing historical records, and present users with specialized, diverse association effects.

Group topics primarily address special circumstances of contract documents and similar manuscripts. Although contract documents are cataloged by individual item, they often appear in batches from single families, termed "household attribution"—only household-attributed contracts possess high research value. From a cataloging standards perspective, the "title" field alone would generate numerous similar titles for multiple documents, making cataloging impossible for hundreds of items together. Assigning a general title would violate objective cataloging principles. Therefore, drawing on the academic "household attribution organization method," such manuscripts can be presented through group topics. In fact, these preserved documents are mostly original vouchers and records that were once intimately connected to the owners' production, life, social interactions, and emotional worlds, belonging to the same entity and mutually associated, thus forming a continuous whole demonstrating internal belonging. Using the household attribution concept to establish group topics for batches of documents from the same family allows linking all associated manuscripts. This both presents objective conditions of individual items for retrieval convenience and reveals combinatorial relationships and overall characteristics through group topics.

In summary, item-level cataloging combined with randomly combinable topic-based virtual integration represents an optimal solution for Chinese Manuscript Documents Digital Resource Library construction. Meanwhile, the aforementioned multi-dimensional tagging—dynamic clustering through dimensional tags—also serves as a method for addressing associations.

3.1.2 A Special Form of Virtual Integration—Reconstruction Reconstruction topics primarily apply to Dunhuang manuscripts. Due to various historical reasons, entire Dunhuang scrolls were torn into multiple sections or small fragments. Manuscripts originally belonging to the same item but now artificially separated are cataloged by holding institution numbers, dispersing these fragments across separate records. Therefore, the platform supports database builders and users in forming reconstruction topics by grouping related fragments and providing explanatory notes.

Project chief academic expert Professor Zhang Yongquan has long engaged in Dunhuang manuscript reconstruction. Reconstruction topic construction currently utilizes his research results to associate known scattered fragments under

single themes (see Figure 2 [Figure 2: see original paper]).

With the development of artificial intelligence technology, exploration and construction of smart ancient books platforms has begun implementation, primarily involving scholars establishing overall frameworks and related content, then using technical means for presentation. We are collaborating with computer science experts to explore machine learning methods for manuscript fragment reconstruction through shape analysis and handwriting comparison, using human-computer interaction to rapidly achieve resource reorganization and improvement. In the future, we hope to use AI technology to quickly realize systematic resource reconstruction, facilitating user learning and research.

3.2 Academic Community Based on Resource Space

Traditional special collection libraries provide one-way document services where users search for and obtain needed resources before returning to their work platforms to continue research. If resource space and research space could be connected or even integrated, allowing users to complete resource searching, indexing, annotation, and reorganization within a single space to form personal knowledge spaces oriented toward research directions, learning and research efficiency would be greatly improved. Simultaneously, such knowledge spaces could continuously enhance the resource space of the special collection library, creating a self-growing resource library—an important development direction for next-generation special collection resource platforms.

Our approach involves introducing a series of tools into the special collection resource platform, such as a GIS system based on historical geography, allowing users to freely annotate and organize platform and personal resources from spatiotemporal dimensions through personal spaces; providing general knowledge bases including era conversion, traditional-simplified character conversion, queries for official positions, place names, personal names, and objects, and multilingual translation; and offering visualization toolkits for convenient chart-based content display. These tools seamlessly integrate resources into learning and research environments, supporting multi-dimensional resource display to promote learning and research discoveries.

To this end, we are actively examining domestic and international peer solutions. Globally, multiple projects use URIs and W3C open standards to link other data for virtual integration and research space formation, including DDB, BNB, LC, DPLA, the Getty Museum, ResearchSpace, and Shanghai Library [10]. ResearchSpace, developed jointly by the British Museum and Oxford University, 致力于通过关联数据汇聚全球在线资源形成特藏知识空间，是一个透明穿梭于特定主题的全球资源平台。网站对自己的介绍是“连接、关联、语境知识表达”(Connect, communicate and represent knowledge with context) [11]。大学数字图书馆国际合作计划 (CADAL) 项目从 2017 年起与牛津大学开展合作，成功地将中国传统绘画、音乐文物两类资源与 ResearchSpace 系统进行了链接。我们希望借鉴 ResearchSpace 的框架，逐步将特藏资源平台构建成基于资源空间的学术社区。

4. Sustainable Development and Utilization

For a resource library to become a "growing organism," in addition to builders continuously updating content, user participation represents a primary method. When platforms evolve from document services to knowledge services, then to publishing services, and even become a cyber research room within users' learning and research spaces, the substantial content created by users can both serve as academic archives supporting learning and research and, beyond displaying academic papers and monographs on the platform, further expand the academic space of the special collection resources.

We believe several approaches are worth considering: first, the website must possess unique resources that compel users to visit; second, it must have integrative capacity to achieve completeness or relatively maximum aggregation of certain resource types; third, it must promote co-construction and sharing to advance multi-party collaboration; fourth, the website must provide a series of academic tools to form a resource-based research support platform.

The Chinese Manuscript Documents Digital Resource Library has planned and preliminarily implemented construction across these four dimensions.

4.1 Building on Unique Manuscript Resources at Zhejiang University

Zhejiang University's manuscript collections primarily include Dunhuang-Turpan documents, local documents, and religious manuscripts. The Turpan documents represent a particularly unique collection of 20 items, most retaining their original form as funerary objects without disassembly—a relatively precious survival globally. The project team plans to conduct 3D digital scanning of Zhejiang University's Turpan document collection.

Zhejiang University Library also holds a batch of local documents focusing on Zhejiang and adjacent provinces, primarily covering Zhejiang and Fujian regions, including Yunhe County, Songyang County, Jingning County in Lishui, and Taishun Township in Wenzhou, totaling approximately 15,000 items/books/pieces. The "Zhejiang Lishui Qingyuan Jushui Jiankeng Wu Family Folk Documents" represents the most valuable set, comprising over 1,200 items documenting the Wu family's economic transactions and folk activities from the mid-Ming to Republican period, possessing high research and utilization value. Since the 20th century, as historical research perspectives have shifted toward grassroots society, using folk historical documents has become a trend. Local documents are highly valuable for studying rural social organization, land rights structures, taxation systems, people's livelihoods, religious beliefs, and language, making them highly favored by scholars. Culturally, local documents serve as important carriers of historical culture, possessing artifact value like excavated cultural relics, with each document being unique and irreproducible. Unlike printed books, they are mostly handwritten manuscripts with limited circulation, essentially existing as "unique copies" passed down through generations. Therefore, the collection

and organization of local document literature constitutes both an academic foundation and frontier, possessing dual value for cultural preservation and academic research.

Zhejiang University Library's religious manuscripts include several Buddhist sutra scrolls 8-10 meters long, plus copies of Buddhist and Taoist classics and ritual texts, all quite distinctive.

Through purchase, donation, and organization of existing holdings, the quantity of these characteristic documents continues to grow.

4.2 Building a Turpan Documents Publication Site

In terms of digitization history, Dunhuang documents already have the relatively mature International Dunhuang Project (IDP). Launched in 1994, IDP's core work involves restoration, cataloging, and preservation of ancient documents and artifacts from Dunhuang and Xinjiang. With network technology development, IDP aims to collaborate with holding institutions to reunite these artworks through high-quality digital images, facilitating online access to increasing information for scholars and the public.

In contrast, no website currently systematically collects Turpan documents globally. The Chinese Manuscript Documents Digital Resource Library will prioritize Turpan documents as a special topic, establishing itself as a global publication center for integrating Turpan documents, promoting their preservation, research, and sharing, and establishing relevant regional centers.

4.3 Co-construction and Sharing through CADAL

The China Academic Digital Associative Library (CADAL), as a public welfare digital library project constructed by the Ministry of Education, has established extensive cooperation with domestic universities and dozens of renowned university libraries and national libraries abroad. The Chinese Manuscript Documents Digital Resource Library will rely on CADAL to promote cooperation with domestic universities and international manuscript holding institutions. Taking the millions of extant local documents as an example, the project portal plans to include published documents and collaborate with major holding institutions for joint publication.

4.4 Strengthening Construction of Academic Monographs and Reference Works

Focusing primarily on Dunhuang documents, two approaches are emphasized: first, listing reconstruction information and academic research results for each Dunhuang scroll after the document entry to provide academic reference; second, prioritizing the construction of Dunhuang studies reference works to strengthen the portal's academic research and reference functions.

5. Conclusions and Recommendations

Using the "Chinese Manuscript Documents Digital Resource Library" construction project as a case study, we have systematically examined special collection construction, organization, service, and utilization, making some beneficial attempts while focusing more on planning and reflection. During this process, we discovered that special collection resource construction requires conceptual and technological breakthroughs in sources, formats, organization, and presentation. Limiting scope to a single institution or location, targeting only one or two media types, relying entirely on standardized classification principles and controlled vocabulary indexing, and providing access through keyword search will inevitably render special collection resource libraries as isolated islands gradually submerged in internet 泡沫 (froth).

In the big data era, resource revelation is more comprehensive, network conditions support diversified division of labor and cooperation, and the exclusivity of digital resources is less prominent than print resources. Through crowdsourcing methods combining metadata and content, balancing discoverability and accessibility, we can construct digital special collections far more complete than traditional databases, with better user experience and positive cycles of utilization and sustainability.

In light of this, we position the Chinese Manuscript Documents Digital Resource Library as a resource-based research and learning platform oriented toward high-quality users including scholars and the public. First, it should 立足本地放眼全球 (立足本地放眼全球), digitally integrating as complete thematic resources as possible. Second, it should provide crowdsourced indexing tools, encouraging users to conduct multi-dimensional indexing to generate multi-directional associations. Third, it should open academic space to meet users' needs for reading, learning, resource aggregation, and even knowledge creation. Some of these concepts have been implemented in Phase I and received positive feedback, while others await continuous improvement and deepening in subsequent project phases. We hope to create such a new platform to attract users from different learning and research fields, continuously improving platform usability through their experiences while enabling their participation in special collection content generation, thereby developing a self-growing, self-organizing special collection construction model that becomes a paradigm for digital humanities research and new liberal arts laboratory construction.

References

[1] Du Yu, Lu Jian. "Chinese Manuscript Documents Digital Resource Library" Launched [N]. *Guangming Daily*, 2022-6-26(4).
[2] Zhang Yongquan. Manuscript Documents: Important Carriers of Chinese Civilization Transmission, In: *Viewing China from Archaeology* [M]. Beijing: Zhonghua Book Company, 2022.
[3] Zhao Yanchang, Li Xiaoguang. On the Reasons, Process, and Specific Distri-

bution of Dunhuang Documents Lost Overseas [J]. *Liaoning Provincial Museum Journal*, 2012(00).
[4] Wang Xiaoping. Empirical Research on Turpan Documents Lost Overseas [J]. *Popular Literature and Art*, 2016(2): 273-274.
[5] Yang Peina, Shen Bin. Toward a Folklore Document Studies: The Evolution of Collection and Organization Methods for Folk Documents in the 20th Century [J]. *Journal of Sun Yat-sen University (Social Science Edition)*, 2014, 54(5): 71-80.
[6] Wang Lei. Huizhou Documents, Huizhou Studies, and Digital Humanities [J]. *Library Tribune*, 2016, 36(9): 1-4.
[7] Yin Weiqin. On the Diversity of Judicial Basis in Republican Period Grassroots Courts: A Case Study of Longquan Sacrificial Land Dispute Judicial Archives [J]. *Zhejiang Social Sciences*, 2010(5): 92-96.
[8] Gao Chaoqun. Introduction to New Book *Collection of Lanxi Fish Scale Registers* [J]. *Researches in Chinese Economic History*, 2023(1): 103.
[9] Han Songtao. New Discussion on Digital Library Classification [J]. *Library Journal*, 2011, 30(10): 36-39.
[10] Zhang Zheyu, Zhang Lei. Open Data Construction and Digital Service Transformation in Memory Institutions [J]. *Library Tribune*, 2020, 40(5): 21-26.
[11] ResearchSpace [EB/OL]. [2023-6-25]. https://researchspace.org/.

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