IIAI Open Conference Publication Series
IIAI Letters on Informatics and Interdisciplinary Research
Vol.006, LIIR400
DOI: https://doi.org/10.52731/liir.v006.400

Digital Technology Applications for Discovery of Contextual Links among Records - Developments, Challenges, and Prospects Regarding Utilisation of Historical Information Resources From Medieval China

Yichen Wu *

Abstract

Compared with those developed for born-digital records, ICT applications for digitised historical documents remain relatively immature, awaiting further development. In response to growing interest in local historical research, the utilisation of digital technologies to discover contextual links among scattered records has yielded noteworthy outcomes. This paper examines the current state of digital technology utilised for Chinese historical research and education through case studies. Three representative platforms, THDL, CGKSP, and CHLAD, which rediscover and present contextual links among fragmented Chinese medieval records, were selected for in-depth analysis. This paper examines the potential of introducing archival discourse on context to enhance the representation of historical record aggregates with greater consistency and integrity on these platforms.

Keywords: Archival science, Chinese historical records, Context, Digital technology application, Knowledge discovery

1 Introduction

With its indispensable role in handling historical records, the concept of context, even with its varied definitions and scopes, is essential in both archival science and historical research [1][2][3]. In this era, the digital transformation of records has further amplified the significance of context for archivists, but questions regarding the concept's applicability and function persist. Although this concept has received limited attention from information professionals, the issue requires further exploration.

Besides these considerations of the concept of context, it is reasonable and natural for informatics specialists to develop applications for automated processing and broader access to records as information resources [4][5]. Although not the first type of information resource to be processed, historical records – formed in the past rather than the present – have been integrated into platforms empowered by digital technologies for research and educational purposes for over a decade, especially in the context of emerging digital humanities. The employment of digital tools to quantitively analyse large volumes of historical data, such as in cliometrics, suggests that the application of digital technology to historical records can be traced back to earlier advances in information and communication technologies (ICT) [6]. Whereas historians have increasingly acknowledged the supportive effect and profound influence of digital technologies on the re-

* University of Tsukuba, Tsukuba, Japan

search and education of Chinese history, the discovery of contextual links among dispersed Chinese medieval historical records, rather than statistical analysis of the figures of and within them, remains a relatively innovative subfield. Furthermore, the design and development of such applications are short of a sophisticated consideration of the concept of context.

By defining context as the information about the historical environment that characterises record creation, transmission, and utilisation, this paper examines recent advancements in digital technology applications that discover and present links among records through the lens of context. This specified mode of knowledge discovery both supports and is supported by the latest trends in Chinese medieval local historical research.

2 Reliance on Records in Recent Chinese Medieval Historical Research

As in other national historiographies, research on Chinese medieval history heavily relies on records as historical materials and academic information resources [7]. However, such research places even greater emphasis on records and documentation, particularly due to their booming number and scope during the Ming and Qing Dynasties [8]. On the other hand, local research focusing on these periods requires more access to dispersed and scattered records due to the shortage of systematic and centralised official preservation of those records for local purposes.

As a traditional and widely relied-upon academic information source, records from authorities during the Ming and Qing Dynasties have long been widely accessed and cited by historians. Moreover, there has been a trend in which researchers increasingly utilise private documents and other records outside official custodianship, as historians' attention and focus are drawn to local and village history in medieval China more [9]. Against this background, documentation publications, information platforms, and historians themselves are contributing an increasing number of records to historical research - many of which have not undergone processing or quality assurance by archivists. For research on local history beyond the centralised administration system, particularly in Southeastern and Southern China, concerns extend not only to the aggregation of records but also to how these records are organised and represented through schemas, frameworks, and standards that structure otherwise separately maintained but collectively required information resources. When addressing this challenge, it is not sufficient for historical researchers to represent historical records in a logical and detailed order. Besides them, it is more crucial to ensure that records are presented in ways that enable historical documents to be translated in either a narrow or broad context in favour of the needs of historians or archival traditions. For scattered and fragmented private records, it is critical to organise and represent them with their shared contextual substance, contextual information, and inter-record linkages. This context-based approach enables the understanding of these records not just as isolated documents, but as integral units—reflected in archival notions such as fonds, record groups, or series—and as evidence of creators and creations that form a continuum within historical events and existence.

However, the lack of context in record representation worries users, providers, and organisers of private records and official documents, especially in local historical research. Similar to the current principles of respect for the original order, documentation from medieval China was primarily preserved in the order and organisation established by authoritative institutions where it was last circulated, utilised, and kept for transactions and administration. However, within the vein of the empire's developed administrative system, correspondence between the central court and local authorities, or among different functional sections across various civic levels, often

resulted in a complex organisation. Records from a single section were typically sorted into different inventories based on their authorship and style. As a result, sent documents and their replies were stored separately. Consequently, the original document sent and the response received are often found in distinct inventories, and the latter usually cited only a simplified version of the former in medieval documentation systems. Furthermore, when received records responded to or addressed an issue from a third-party department or local authority, they rarely included the complete originating cause of the response. These conditions often compel historical researchers to consult multiple inventories or even different fonds to compile a comprehensive list of records within a single context. This dispersion of records from a single transaction or affair, combined with preserved context, implies compromises among various contexts of records across different points in their lifecycle: the terminal agency where records are kept in their final preservation was commonly taken as the primary context, while previous agencies, where records were sent or received earlier, conceded. Despite efforts to represent and preserve the most recent state of records before archiving, limitations in traditional archival infrastructure result in compromises that limit more comprehensive descriptive practices.

In addition to the challenges posed by the nature of paper-based record preservation and the complexities of bureaucratic structures, traditional publishing approaches impose further constraints. Due to the geographical distance between different archival repositories and the difficulty of directly accessing the original records held in repositories, edited, compiled, and published collections of historical records have become one of the primary means for historical researchers to initially review and investigate records related to their research topics, long before the emergence of digital repositories and online databases [10]. However, for those results of the collaboration between the editors and records custodians, some obstacles remain to the more efficient and convenient utilisation of records by historians.

On the one hand, published historical records are often sourced from a single or limited number of fonds or inventories. Constraints such as time, cost, or authorised publication scopes through appraisal may prevent the inclusion of all original records referenced by others in a single publication. On the other hand, historians' efforts to aggregate records based on their logic and content around a given theme may also unintentionally overlook the citation relationships and the more profound contextual relevance among records. Although this analysis is based on the examination of conditions in Chinese medieval local research and studied records in this field, the discussion and exploration of digital technology solutions for current and potential difficulties can theoretically be expected to be applicable and extendable to all sources of dispersed and scattered records required by historical research.

3 Digital Attempts to Discover and Present the Context of Historical Records

In recent decades, there has been a growing trend toward digitising historical records and providing access to their digital surrogates. For digital collections and platforms focused on Chinese historical research, the first stage primarily involved the digitisation of records and their supplementary materials. The next stage was focused on extracting numerical data from records, describing them, and conducting automated statistical analysis. A case in point of this second step is the automated processing of historical records in the field of cliometrics and economic history, where an enormous volume of numbers and figures written down in records are required to be analysed statistically for macro-level description of quantified historical facts and some-

times even historical trends as part of conclusion in historical studies.

However, recent developments differ significantly from these earlier stages, as they aim to process and present records holistically to enable the automated discovery of knowledge or insights [6]. As the novel stage progresses, the processing of digitised historical records reveals their context, which is vital for comprehending records in historical research. To examine how digital platforms represent context, particularly in terms of links and relationships, this paper reviews qualified cases that illustrate these emerging trends.

3.1 The Selection of Cases

Since this paper aims to review not only practical situations of but also the development of suitable platforms for discovering and presenting contextual links among Chinese medieval records in historical research, the selection of appropriate cases was conducted through a qualitative review of the literature describing the development of platforms meeting these conditions.

To minimise the potential negative impact on the understanding of Chinese historical research due to linguistic barriers and a lack of basic knowledge of Chinese history, the writing language of the reviewed papers was limited to Chinese. Given the restricted writing language, the China National Knowledge Infrastructure (CNKI) was chosen as the primary source of papers describing the targeted digital technology applications, as it is the largest single source of academic papers in Chinese. According to the search syntax by CNKI, the search expression could be written as SU = ("Information technology" OR "digital technology") AND ("historical databases" OR "digital history" OR "digital humanities") AND ("archives" OR "records") AND ("context" OR "link"). Here, the SU refers subject. Additionally, the literature search was limited to the following classifications: "Theory of Historiography", "History of China", "China Ancient History", "Library Science and Digital Library", and "Archives Science and Archival Undertaking".

After searching on CNKI, a screening of the papers describing or mentioning the targeted platforms was conducted. During this screening, a set of simplified rules was developed and applied to candidate papers.

Firstly, the criteria of selected platforms were established, which means a publicly accessible website employing digital technology application(s) on the databases aggregated and/or provided by itself for utilisation. Regarding the function of the selected platform, it shall be demonstrated through the discovery and presentation of knowledge among the aggregation of historical records. Secondly, the processed raw data in their databases shall be historical records collected and kept as archival objects, allowing historians to use these databases as an alternative to traditional archives for part of their functions. Thus, since the databases are not exact digital replicas of authentic archives, they should be considered as a kind of online digital repository or collection.

Then, given the growing focus on the trend of historical research, the records mainly offered by these databases are limited to those created during the Chinese Ming and Qing Dynasties. For the influence and representativeness of the platforms, the compositions of the offered historical records should be as comprehensive and extensive as possible. On the other hand, the developers shall consciously target historical research and learning as their platform's utilising scenario, with historical researchers and learners as intended users. Furthermore, among the discovered or rediscovered knowledge in the records, there must be context and links that are based on the context of the records. Lastly, the design, selection, and mechanism of the applied digital tech-

nologies must be depicted and explained in the published literature.

After reviewing the literature from CNKI, platforms with the largest volumes of historical records in a specific region or genre were selected for their representativeness and excellence as historical record repositories, which are crucial for knowledge discovery and historical research. To facilitate the anticipated comparison among different types of digital technologies and their application patterns, only one platform was retained when similar digital technologies were applied in a similar manner. Finally, a further extension to related papers written in English was conducted based on the authors of the Chinese writings.

Following the steps mentioned above, three cases – the Taiwan History Digital Library (THDL) [11], the Chinese Genealogy Knowledge Service Platform (CGKSP) [12][13], and the Chinese Historical Local Archives Database (CHLAD) [14][15] – were selected for subsequent comparison analysis.

3.2 The Cases' Profiles

These selected cases were chosen for their representativeness as online platforms that apply digital technologies to furnish historical records in their respective contexts. Notably, these platforms differ in the types and approaches of digital technologies employed. With the shared goal of providing historical records for research and education, they apply digital technologies to discover and present contextual links among records from different repositories and provenances, as shown in Table 1. All of them are designed and developed as experimental projects to explore automated or semi-automated approaches to processing historical records as research materials through the support of digital technologies.

Table 1: Basic information of the selected platforms

Platform	THDL	CGKSP	CHLAD
Published time	2009	2015	2012
Development Organisation	National Taiwan University	Shanghai Library	Shanghai Jiao Tong University
Types of records	Private deeds and official documents	Genealogies	Diverse private records
Sources of records data	Publications, archival repositories, and private collections	Library collection and indexed genealogies from outside collections	Original private collections
Provenance of records	Transaction parties and authorities at various levels	Families and their houses	Ordinary citizens of towns and villages

THDL (URL: http://thdl.ntu.edu.tw/) was launched in 2009 by digital humanities scholars at National Taiwan University. As a platform that integrates official and private documentation core databases for issues regarding Taiwan during the Ming and Qing Dynasties, THDL provides aggregated and arranged full-text transcriptions of records concerning the same or relevant transactions or communications. Referring to the original or photographic versions of the records, THDL employs text-mining algorithms to automate the discovery and representation of potential contextual relevance among historical records [11].

CGKSP (URL: https://jiapu.library.sh.cn) is a historical knowledge platform launched by the Shanghai Library in 2015, catering to both professional historians and amateur researchers. With the participation of librarians, this platform introduces broader perspectives on knowledge discovery and management. It is built upon a core database of global Chinese genealogies and empowered by linked data of people, places, time, genealogies themselves, and other entities involved in their creation [12][13].

CHLAD (URL: http://ndfwx.datahistory.cn/) was started in 2012 as a platform to promote the academic utilisation of private records collected by the Department of History at Shanghai Jiao Tong University. In collaboration with the Press and Library of the same university, the platform relies on a standardised and specified metadata schema and metadata field search function to link private records, primarily from the southeastern provinces of Ming and Qing China, according to historical entities mentioned within them. Designed and constructed with input from historians, the platform's development incorporates principles and practices around provenance from archival science [14][15].

3.3 Sources, Scales, and Presented Forms of Records

The genres of incorporated records, such as official documents, private and business land deeds, and genealogies, clearly distinguish these selected platforms. Additionally, their sources, including compilations of records, academic monographs extracting records, and central or local chronicles, are fundamental for their diverse characteristics. Furthermore, the presentation formats of records can influence how users utilise the platforms simultaneously.

While these platforms share a common aim of supporting historical research and education through digital technologies for knowledge discovery, they focus on contextual links among historical records, each with diverse yet representative sources. The THDL collects and provides Ming-and-Qing records related to Taiwan, sourced from extant publications and accessible repositories, offering what is likely the most comprehensive list of Taiwan's private and public historical records from that period. The CGKSP focuses on Chinese genealogies, offering the broadest scope for this special genre. Based on a published catalogue and continuous updates, the CGKSP database may be the most extensive digital surrogate and reference collection of Chinese genealogies from the Ming and Qing Dynasties. The CHLAD, of its term, is built on unique and exclusive data sources, mainly as private records collected by the history faculty at Shanghai Jiao Tong University through years of fieldwork. Most of these records were created and circulated for transactions in the southeastern provinces during the Ming and Qing Dynasties. Compared to other local record collections, CHLAD stands out due to its spatial and temporal range, as well as the platform designed for its utilisation. The representativeness of these platforms is evident in the vast number and broad range of records they include as historical information resources. The THLD comprises approximately 40,000 land deeds and over 37,000 official

documents. CGKSP and CHLAD provide more than 56,000 genealogy entries and 35,000 local records, respectively. These substantial volumes make them among the most comprehensive platforms in their respective categories and scopes at the time of their founding, ensuring their representativeness and value for historical researchers and enthusiasts.

In addition to the types of sources and quantities of records, the platforms differ in their presentation formats. Both THDL and CHLAD provide the full text of every record included, while CGKSP adopts a different approach. The latter platform only provides partial access to certain genealogies, which can be searched and reviewed. This limitation is a result of the platform's goal, as expected by its developers, to include all existing Chinese genealogies on a global scale, not just those in their own collection. Due to the limited accessibility of original manuscripts, CGKSP provides as much descriptive information as possible for each genealogy and provides references for access through their custodians. Regarding the presentation of full-text records, CGKSP and CHLAD primarily provide photographs of original documents, whereas THDL offers only transcribed texts.

3.4 Primary Digital Technology Applications and Their Patterns

These cases, with similar development goals and target users, utilise distinct types of digital technology in various patterners to facilitate the discovery of contextual links among records, which are necessary intermediate outcomes for historical research. This diversity highlights the varied directions and potential solutions that coexist for advancing specialised knowledge discovery based on archival data. Faced with these differences, there is a clear need to examine the adaptability of various digital solutions and platform frameworks to enhance the efficiency and effectiveness of information retrieval in a specific context.

At THDL, the primary digital technology employed is text mining based on custom-developed algorithms. Both private deeds and official documents undergo text extraction and are tagged with the extracted words and phrases according to developer-specified rules. These rules help identify proper nouns - such as names of historical figures and places - by recognising exclusive terms that typically appear in specific positions relative to the proper nouns within the content of records. The THDL team developed a custom algorithm called *Word-Clip* to extract and tag these words, presenting them as properties of the records in both content and context. The tagged terms are then used as metadata and the basis for automatically comparing the given description elements, such as individuals or organisations involved in transactions, geographic locations of transactions, and dates of record creation. This facilitated the automated discovery of latent connections among records based on the shared contextual information.

On CGKSP, metadata forms the foundation for its digital technology applications. However, the primary phase of metadata preparation is carried out manually by experts. In this way, the platform primarily processes data derived from the records to describe them rather than the content of the records themselves. Based on expert-created metadata, CGKSP utilises linked data as the core component of its framework. On this platform, the metadata structure deviates from the traditional approach; instead, it redefines metadata items as entities and their associated properties. Guided by linked data principles and designed to support historical research and education in a digital humanities environment, the platform structures the description of records using the Resource Description Framework (RDF). This approach facilitates more precise and reliable discovery of correlations among entities described by and used to describe records, many

of which form part of the context of records. Through this approach, CGKSP applies digital technologies directly to the contextual information of the records, not to the contentive information, the plain text, of records. Each entity is assigned a unique Uniform Resource Identifier (URI), which together forms a lightweight semantic web of records on CGKSP [13]. While the platform supports traditional information retrieval through a metadata-based search interface, it offers a knowledge discovery service that identifies contextual links among records, such as shared creation dates, geographic locations, or families with the same surname of creators.

CHLAD employs relatively straightforward digital technology solutions. Like CGKSP, it focuses on the role of metadata in the discovery and presentation of contextual links among records for historical research. However, CHLAD does not adopt linked data structures. Instead, it devotes effort to designing a detailed and delicate metadata schema through close collaboration among digital technology specialists, librarians, and historians. This design process draws upon archival theory and practice, particularly the concept of provenance—the origin or creator of a record—which is treated as a key contextual factor linking records. All metadata on CHLAD is manually created and entered by experts. The platform then offers a search interface with fully searchable metadata fields. Notably, the "family" field, unique to CHLAD, represents both the administrative divisions and familial information of the record creators simultaneously. This design illustrates a historian's conviction that research and interpretation should be grounded in the aggregated records generated by social units, such as families or communities, rather than isolated documents. Similar to CGKSP, records in CHLAD are linked through shared historical entities, such as family, geographic locations, dates, and causes. However, this is achieved without the use of more advanced metadata frameworks or semantic technologies. In this way, CHLAD provides a knowledge discovery tool in the form of information retrieval to support historical research and education while requiring less labour for the organisation and search of contextually related records.

In summary, these three representative platforms demonstrate the diverse applications of digital technology for discovering contextual links among records in the research and education of Chinese medieval history. Beyond the distinct compositions of their development teams and the varied involvement of professionals such as librarians and historians, this diversity indicates the need to evaluate the applicability and adaptability of different digital technologies for the same purpose. In detail, the divergence raises questions about whether the semantic web or metadata can provide a direct mapping of contextual entities and their relationships. Moreover, the divergent technological solutions adopted by these platforms prompt consideration of whether digital technologies should be applied to deal directly with the contextual information of the records, as their attributes, or the content information, as their text. More importantly, it is a haunting problem of which side should be prioritised: the design and development of technologies, the framework and algorithm, or the manual generation of contextual information by experts. These considerations will provide guidance for future innovations in platform design and technological advancements in this field, offering a reference framework that is acceptable to both developers and users.

4 Role of Context Notion in Development of Digital Technology Applications

Beyond the digital technological solutions for discovering contextual links among historical records, other basic and supportive features are critical to these platforms; however, they are often overlooked because they do not directly involve the components and procedures for re-

constructing lost contextual links. Among them, the most influential ones are those presenting rediscovered links and defining the scope of context.

More importantly, these often-neglected aspects of the analysed digital technology applications may be significantly affected by the developers' understanding and perception of the concept of context. In this way, the potential improvement in performance from these perspectives implies a greater reliance on the existing notion and expertise of context from the field of archival science.

4.1 Presenting Form of Contextual Links among Records

After discovering contextual links using various digital technology solutions, another key task for these platforms is to present and effectively communicate the discovered knowledge to users. In addressing this challenge, the developers of the platforms mentioned above appear to have adopted more conservative approaches.

THDL provides a more comprehensive and integrated representation of its records, along with the contextual links among them. This applies to both incorporated private deeds and official documents. THDL represents these records through diagrams that display directly or indirectly linked items based on their contextual information. In the case of private deeds, links are established based on land parcels that were traded, divided, or transferred. For official documents, links are established through citations and references, which collectively build their functional and creation contexts. A strong contextual link can generate a diagram that spans dozens of records, presenting a chain of linked records at a glance. Regardless of the number of records included, the primary advantage of this presentation format is that it displays linked records in chronological order and integrates records that may not share direct citation or textual relevance, all in a cohesive relational web.

In contrast, CGKSP and CHLAD represent linked records primarily through lists. Even though they provide visualisations of the temporal and spatial distribution of records, which include the geographic location and date as shared contextual information, the records themselves—whether aggregated by contextual entities, attributes, or common geographic or temporal information—are presented in a list format. This mode makes it challenging to reveal indirect links or to bridge gaps where contextual information is not shared among records. For example, if Record A shares the same creator with Record B, and Record B shares the exact geographic location as Record C (based on the location of the transactions they document), the list format definitely cannot effectively reveal the indirect contextual link between Record A and Record C in a single output as a list of searched results.

This limitation shows a potential area for improvement in the presentation and discovery of contextual links among records. It also suggests that a more effective approach to platform development may involve a more inclusive design that integrates diverse experiences from various pioneers. The model and scope of different context elements should also be considered when determining the presentation format of contextual links. Here, the notion perceived by the developers is primarily reflected in the extent to which a historical record's context is preserved – whether records indirectly related to it still hold a position within its context.

4.2 Scope of Context Discovered

Furthermore, a review of the three platforms reveals a lack of consensus on the scope of context and contextual information that digital technologies should discover and present for historical research and education. Generally, the contextual links aimed at by these platforms can be summarised as relationships constructed based on the shared or similar entities or properties of their records. However, concerning specific and individual varieties of contextual links, there are still significant differences among these three platforms.

On THDL, the processing of land deeds prioritises the geographic location of the transacted land as the primary context. However, this is technically a property of the recorded item, such as the land itself, rather than the entities themselves existing in a transaction. In contrast, for official records on the same platform, developers place greater emphasis on citations to other records, which serve as the direct cause and legal foundation for those records.

For genealogical records on CGKSP, the scope of context is broader, encompassing the family's surname, lineage, or household, as well as the temporal and spatial distribution of record creation and even information about current custodians.

In the case of CHLAD, unique elements, such as family affiliation and the causes leading to the creation of private records, have been introduced as a pillar of knowledge discovery [14]. For the platform's developers and early users, these entities, properties, and events are considered crucial to historical research and education, as they facilitate the creation and inclusion of these items within the broader context.

Although it is expected for the scope of context in knowledge discovery to vary based on the source, provenance, record type, and user needs, such diversity also reflects a dynamic, competitive trend in the field's progress. For those representative platforms and projects, there are no commonly accepted and adopted routines in their design and development processes yet, only coexisting solutions and approaches. Conversely, the lack of a reference model that defines the scope of context poses a universal challenge, which limits the interoperability and convenience of research or educational projects conducted across multiple platforms. As a practical instance, a historical research project designed with a unified scope of context, intended to be explored and requested in consistent formats, will inevitably find that the needs of its researchers can only be partially met by some platforms, not all of which contain relevant records related to the given historical topics and themes. Therefore, a standard reference framework for contextual links discovery among historical records could support the quality, consistency, and evaluative potential of these services for both academic communities and the broader public.

Even though there has been clear and functional progress in the discovery of contextual links among records for historical research and education through digital technology applications, several challenges remain. As shown in Table 2, the diversity in the types and patterns of applied digital technologies, the forms of presentation for discovered contextual links, and the understanding of the scope of the context all require a thorough review with a critical reflection on the needs of both historical researchers and amateurs. On the one hand, historical research is usually personal and context-specific, shaped by individual research objectives and thematic focuses. On the other hand, it is feasible for digital technology applications to establish referential models that promote broader usability and interoperability. In particular, the definitions of context and its scope could benefit from simplification, especially in light of the absence of a universal metadata standard, as the THDL developers found in their work. CHLAD's incorporation of the archival science concept and principles of provenance into its digital framework has highlighted the potential value of drawing on broader theories and practices related to context - particularly those surrounding and extending from the notion of provenance - to enhance the scope of context and

to develop more robust technical frameworks for the discovery of contextual links among records [15].

Table 2: Comparison of the digital technologies applied on platforms

Platform	THDL	CGKSP	CHLAD
Applied digital technology	Text mining	Linked data	Search engine based on metadata
Processed objects	Original text of records	Metadata of records	Metadata of records
Represented form of records	Transcribed text	Digital image	Digital image
Presented form of contextual links	List and diagram of linked records	List of linked records	List of linked records

4.3 The Potential Enhancement from Archival Expertise on Context

Although widely used in everyday and academic circumstances, there is no universally accepted definition or scope of context across distinct academic fields and professions. Similarly, a noticeable gap in understanding the notion exists between archival science and history. However, by drawing on the enduring tradition of supporting access to and reviewing records in historical research, consistent engagement with records enables archivists to develop more systematically considered approaches to describe records with a carefully captured notion and quality of context.

Firstly, archival expertise can provide a steady and general basis for a more comprehensive description framework of context. Since the last century, archivists have developed standards concerning the context and description of records. For entities and properties in modern administrative and official documents, archival professionals have a relatively long history of developing demonstrations to enhance the presentation of their context. Although not specifically tailored for historical records, these approaches can still be adapted and implemented for digital technology applications that discover contextual links among them. Along with the recent focus on describing records in a broader and more complex landscape of context, especially as the Records in Context model has been developed and launched as the novel outcome of these efforts, it is reasonable to extend existing archival models to include more comprehensive contextual entities for historical records utilised in research and education [16].

Secondly, the archival discourse on the distinction and relevance between provenance and context could be borrowed to clarify the presenting form of contextual links, which could be enriched to be more reasonable and practical for historical researchers and learners. Besides taking the scope for context and contextual information as a reference, the evolving notion of context in archival science can also provide a structure for arranging records on these platforms, which are contextually relevant to each other. This could address the question of how closely related records should be represented together during the development of these platforms. Furthermore, a presenting form that distinguishes records aggregated by precise provenance from those by only relative context could be developed for more practical effect and support of this category of digital technology applications.

Lastly, not only the presenting form of contextual links among historical records but also the presenting form of historical records themselves could benefit from the understanding of context provided by archival professionals. Within the designed framework of these platforms, the archival notion of context can be used to justify why often-overlooked information in records, beyond the transcribed plain text, is essential for platforms that prioritise their context.

From a broader perspective, it is anticipated that the targeted development of digital technology applications will get a universally applicable reference from archival expertise on the presenting form of historical records, contextual links among them, and, more importantly, the scope and elements of their context, all of which are currently plagued by inconsistencies.

5 Conclusion

At present, the discovery and representation of contextual links among records constitute a specialised subfield within the broader discipline of knowledge discovery and digital humanities. In response to the growing focus among historians on local history and private records as historical information resources, both in China and globally, this subfield has made notable progress. According to the examination and investigation of the selected cases, it is evident that one of the reasons for this progress is that medieval Chinese history has greatly benefited from the abundance of surviving private and official documents from the Ming and Qing dynasties. Given clear and diverse advancements in this area, it is both timely and reasonable to anticipate further utilisation by targeted user groups. Moreover, continued development could be effectively supported by the integration of archival management theories—particularly those related to the description of records and their context—and by developing a standardised reference framework, which encourages a universal scope and format for discovered contextual information items and links based on them.

References

[1] K. Theimer, "Archives in Context and as Context," Journal of Digital Humanities, vol. 1, no. 2, 2012.

[2] P. Burke, "Context in Context," Common Knowledge, vol. 28, no. 1, 2022, pp. 11–40.

[3] R. Hirsch, "The permanence of Provenance: The 'Two Traditions' and the American archival profession," Journal of Archival Organization, vol. 8, issue 1, 2010, pp. 54–72.

[4] F. Guanyan, "Making better out of technologies: Responses of Interpares to Digital Records Management challenges," Proc. the 10th International Joint Conf. Knowledge Discovery, Knowledge Engineering and Knowledge Management, 2018, pp. 387–397.

[5] K. Theimer, "A distinction worth exploring: 'Archives' and 'Digital Historical Representations'," Journal of Digital Humanities, vol. 3, no. 2, 2014.

[6] B. Shen and P. Yang, "Digital technology and the notion of history: A study of the relationship between historical database and historical methodology in China," Historiography Quarterly, no. 2, 2017, pp. 87–95+159. [In Chinese]

[7] A. Walsham, "The social history of the Archive: Record-Keeping in early modern Europe," Past & Present, vol. 230, issue suppl_11, 2016, pp. 9–48.

[8] Y. Wang and J. Yang, "The Archives of the Ming and Qing Dynasties and the 'Study of Archives of the Ming and Qing Dynasties'," Academic Journal of Humanities & Social Sciences, vol. 7, issue 3, 2024, pp. 34–41.

[9] Philip C.C. Huang, "County Archives and the study of Local Social History," Modern China, vol. 8, no. 1, 1982, pp. 133–143.

[10] Y. Ding and X. Cao, "The significance and principles of archival compilation work - Basic knowledge lecture on archival compilation (Part One) [档案史料编纂工作的意义和原则——档案史料编纂的基本知识讲座(一)]," Archives Science Bulletin, issue 1, 1980, pp. 55–60. [In Chinese]

[11] S. Chen, "Information Technology for historical document analysis," PhD dissertation, College of Electrical Engineering and Computer Science, National Taiwan Univ., Taipei, 2011. [In Chinese]

[12] C, Xia and L. Zhang, "The application of Linked Data in Shanghai Library's service of Genealogy Digital Humanities," Library Journal, vol. 35, no. 10, 2016, pp. 26–34. [In Chinese]

[13] C. Xia, W. Liu, and L. Zhang, "Implementation of a Linked Data-Based Genealogy Knowledge Service Platform for Digital Humanities," Data and Information Management, vol. 2, issue 1, 2018, pp. 15–26.

[14] Y. Qian, Z. Xing, and X. Shi, "From collection resources to intelligent data: Construction of intelligent digital humanities platform for local historical documents of Shanghai Jiao Tong University," Digital Scholarship in the Humanities, vol. 36, issue 2, 2021, pp. 439–448.

[15] M. Tang and J. Zhang, "Research on the process of organising Local Historical Documents from the perspective of Provenance: A case study of Shanghai Jiao Tong University Library [来源视角下的地方历史文献整理流程研究—以上海交通大学图书馆为例]," Culture Journal, no. 12, 2016, pp. 219–222. [In Chinese]

[16] Experts Group on Archival Description, International Council on Archives, "Records in contexts: Conceptual model," Nov. 2023; https://www.ica.org/app/uploads/2024/01/ric-cm-1.0_0.pdf.