International Journal of Digital Humanities
https://doi.org/10.1007/s42803-022-00048-5

ORIGINAL PAPER

Treating a genre as a database: a digital research methodology for studying Chinese local gazetteers

Shih-Pei Chen[1] · **Calvin Yeh**[1] · **Sean Wang**[1] · **Qun Che**[2]

Received: 16 June 2021 / Accepted: 10 August 2022

Abstract
As digital humanities research relies on the digitization of sources, many of its applications are based on access to data on a huge scale that makes quantitative analyses and distant reading (or a birds-eye view) possible. Based on this assumption, we show how the genre of Chinese local gazetteers, with its volume, consistent structures, and broad geographic and temporal range, provides an ideal case to benefit from the digital approach. This paper introduces the Local Gazetteers Research Tools (LoGaRT), a suite of research tools designed for studying Chinese local gazetteers based on the philosophy that any comprehensive genre, such as Chinese local gazetteers, when accompanied with tools that aim to bring a collective lens to the genre, can greatly enrich the ways that scholars approach it and can transform the genre into a research infrastructure that enables new types of research. We report on how LoGaRT opens up new perspectives for researching Chinese history by showing case studies and the scholarly breakthroughs made by our research group. With this paper we hope to provide one possible answer to the question of how digital methods can contribute to humanities research. Even though LoGaRT is developed for a specific Chinese genre, we argue that the proposed research methodology and the corresponding user workflow and tools developed in our software can be applied to other genres or collections of sources when certain criteria are met.

Keywords Chinese history · Historical data · Digital humanities research · Text tagging · Collection analysis

1 Introduction

Digital humanities (DH) is an umbrella term that encapsulates the broad activities of applying digital technologies to the study of the humanities. The large amount of cultural heritage resources, digitized as images or texts, produced during the global wave

✉ Shih-Pei Chen
schen@mpiwg-berlin.mpg.de

[1] Max Planck Institute for the History of Science, Berlin, Germany

[2] Shanghai Jiao Tung University, Shanghai, China

Published online: 06 October 2022

of digitization in the first decade of the twenty-first century, inspired humanities scholars to seek for "killer applications" that used digital computational methods to enrich and advance current humanities scholarship. In the past decade, an increasing number of DH projects have been launched in all corners of the world, exploring different technologies for solving core research questions in many humanities disciplines. They explore technologies developed in computer, information, social, and natural sciences – relational and graph databases, knowledge representation methods, visual analytic techniques, social network analysis, textual analysis, machine learning, bioinformatics – and adapt them for the humanities. Among these endeavors, the ability to deal with a "large amount" of data is a core feature where digital methods shine – computers are able to process digital resources at a speed that a scholar alone would not manage in a lifetime. Now with the availability of digitized resources and fast computing power, we have the possibility to read and process unprecedentedly vast amounts of research materials and derive new findings that would not have been possible using traditional methods. In the past few decades, although many large textual collections of cultural importance have been digitized and compiled into databases, most of them still focus on delivering contents online rather than providing analytical tools beyond browsing, searching, and reading. Among them, Michel et al., (2011) proposed a digital methodology for cultural studies by building a keyword search tool for millions of digitized books to help scholars study the usages of terms in the huge collection. Michel's work is among many others that echo Moretti's "distant reading,"[1] which proposed an approach for understanding a huge collection of literature works by omitting the details of individual works and only looking at the quantitative, visual representation of the whole collection in order to obtain theoretical, structural knowledge of the collection (Moretti, 2000). Both cases use a collection that is comprehensive and statistically representative of a greater unit: in the former case, it was 4% of all books ever printed[2]; while in the latter case it was the entire genres — the epistolary, the gothic, the historical novel, as well as the literary novels published in certain countries (Moretti, 2005). Despite their triumphs, neither work leaves any practical tools or software implementations to enable other humanities scholars to apply the same methodology to other collections. On the other hand, several general-purposed DH infrastructural tools have been developed to support various computational, textual, and linguistic analyses for large textual collections.[3] However, without a clear research methodology behind them, it is up to the experience and creativity of humanities scholars to apply such tools to answer individual research questions.

This paper introduces LoGaRT, a suite of research tools designed for studying digitized Chinese local gazetteers. Beyond providing searching and reading tools, LoGaRT is built to embed the methodology of "treating a genre as a database": it orientates scholars to use a large-enough digital collection to inquire into the

[1] Such as Jockers (2013).

[2] Note that this corpus represents more English language publications for the period between 1800 and 2000.

[3] For example, CLARIN (Hinrichs & Krauwer, 2014) and HathiTrust Digital Library (https://www.hathitrust.org/) both provide large amounts of textual resources as well as modular text analysis tools. Voyant (Sinclair & Rockwell, 2016), on the other hand, allows users to upload a digital collection and provide a rich set of linguistic and data visualization tools to analyze a text collection.

whole genre. Within LoGaRT, scholars are equipped to collect locally situated data across historical China from existing digital local gazetteers. We regard such data as the echoes of a scholar's query in a digital collection, and argue that such echoes should statistically represent all what one could ever find in the whole genre, *when* the digital collection is large enough with regard to its underlying genre. Therefore, the overall collective patterns of such echoes, represented as statistical, visual, and geospatial distributions, should be sufficient approximations that can lead a scholar to understand the overall outlook of her question answered by the genre.

This concept of focusing on the collective value rather than the individual entries of a collection has been embedded in many digital humanities projects. For example, in the China Biographical Database (as well as in the other prosopography databases), biographies of hundreds of thousands of historical figures are used to derive patterns of how kinships and social networks might affect individuals' career paths.[4] In Taiwan History Digital Library, tens of thousands of correspondences between the imperial court and local officials of the Qing empire are used to draw temporal and political patterns of Qing emperors' attention toward this remote island (Chen, 2011; Tu, 2011). Mapping the Republic of Letters uses metadata of seventeenth- and eighteenth-century letters among intellectuals across Europe and Americas to builds an interactive visual tool to produce maps, charts, and other data visualizations that highlight the geographical, temporal, and social network structures of intellectual communities during the Enlightenment (Edelstein et al., 2017).

LoGaRT advances from these proposals in a way that it is built to be a research platform that can support diverse research interests. We abstracted the methodology of treating a genre as a database, and developed a practical workflow that is supported by our software. Within LoGaRT, scholars can issue queries toward the genre, collect data from the available digital collection, and obtain bird-eye views of the data; in the meantime, they are able to validate single data points and iteratively adjust one's queries. With this methodology implemented as a software, scholars are able to apply this methodology to their individual research interests. In this paper, we also report on the research projects derived from LoGaRT to demonstrate how the proposed methodology practically advances our understanding of Chinese history. We believe that this methodology and the practical workflow of software design proposed in LoGaRT are useful for studying any large digital textual collections beyond Chinese local gazetteers.

2 What are the Chinese local gazetteers

Local gazetteers, or *difangzhi* 地方志, is a long-standing genre in Chinese history that records local information on the topography, history, flora and fauna, produces and commodities, officials, biographies, and literature of a locality. A local gazetteer can cover an administrative region – province, prefecture, state, county, or a

[4] Harvard University, Academia Sinica, and Peking University (2021). *China Biographical Database.* https://projects.iq.harvard.edu/cbdb. For other examples of prosopography databases, see Repertorium Academicum Germanicum (RAG). https://rag-online.org/

geographical region, such as river basin or mountain. Local gazetteers have been the main genre of local geographical writing in historical China since the twelfth century, and this method of recording regional information has been practiced by local governments for eight hundred years. This practice was also transmitted to historical Korea, Vietnam, and Japan. They are usually compiled by local officials in collaboration with local literati, in the form of a book, mainly for the purpose of governing the local region (Lai, 1983; Huang et al., 1993; Cang, 2013; Dennis, 2015).

Inherent characteristics of the recorded data combined with its structural nature and the comprehensiveness of its coverage in terms of time, space, and themes makes the genre of local gazetteers an ideal test case for employing digital methods.

2.1 Data nature of local gazetteers

Every compilation of each local gazetteer can be considered as a data collection project carried out by the editors (usually local officials) to investigate and collect up-to-date data for purposes of practical administration, documentation, or for memorizing the past. Data are usually complied in categories and presented as chapters. Figure 1 shows a short but typical Table of Contents from a prefectural gazetteer. It divides the contents into six *juan*'s (volumes or chapters): Geography, Infrastructure, Food and Money, Political Matters, People, and Literature. Within each volume, the editors listed all the related information they could find: all the mountains, rivers, and other topographical features under Geography; all the governmental bureaus, buildings, schools, and temples under Infrastructure; number of households, collectable taxes, local products under Food and Money; all the people who ever served as local officials under Political Matters; all the literature works written by local literati or officials under Literature. A region, especially an administration one, often has multiple editions of gazetteers compiled at different historical times. This enforces

Juan 1: Geography
Map of Jinmen. Map of Gifu. Field division.
History of establishment. Territory. Geographical features.
Local customs. Mountains & rivers. Monuments.
Juan 2: Infrastructure
City's infrastructure. Bureaus. Schools. Altars.
Postal stations. Temples.
Juan 3: Food and Money
Households. Taxes. Public services.
Horses. Budgets. Local products.
Juan 4: Political Matters
Local officials. Responsibilities. Famous officials.
Rituals and ceremonies. Military affairs.
Juan 5: People
Selection. Merits & martyr. Virtue and filial.
Local gentry. Seclusion. Foreigners. Immortals.
Juan 6: Literature
Steles. Poems and Songs.

Fig. 1 A Table of Contents from Shuentien Prefectural Gazetter, 1583. The right side shows a rough translation. (Image source: Zhongguao Fangzhi Ku database by Beijing Erudition.)

the "data" nature of the genre: a newer edition of a gazetteer was often compiled in response to a need for up-to-date local data.

Thus, each gazetteer can be understood as a *data-base*, from which local officials, literati nation-wide, and scholars from later periods can find relevant data of the locality. It is worth noting that data kept in gazetteers, due to its regional characteristics, are often not found in archives kept by the central court, which makes local gazetteers valuable sources for understanding regional conditions.

2.2 Structural nature of local gazetteers

Although local gazetteers present data in a textual manner (Ba, 2004, pp. 26–27), the structure of their content (that is, table of contents) is guided by categories rather than chronology. What is particularly surprising is the consistency of categories in the gazetteers over the course of eight-hundred years. While the editors of gazetteers seem to have total freedom in deciding which categories to include in the work, there is a small set of common categories that re-appear in high proportions of gazetteers throughout the vast geographical coverage and long historical period. The strong structural nature of the genre has a great advantage: when treating all the local gazetteers as one unit, it is highly likely that one can find the same category of data in many gazetteers, thus making the genre a "conceptual database" for collecting thematic data across geographical regions and time periods.

In terms of how a particular category of data is organized down to the section level, they are often further organized into sub-categories and then eventually as lists of entities: lists of topographical features, of buildings, of flora and fauna, of people, etc. For example, Fig. 2 shows the first page of the chapter on local products. It starts with a list of sub-categories of products: grains, vegetables, fruits, woods, bamboos, flowers, herbs, drugs, aquatics, birds, animals, artefacts, and commodities. Then, under each category, it lists every type or species, such as grains, that one can find (or can expect to find) in this region. Sometimes only the names are listed. But in some cases, such lists also contain key features of the entities. In the case of grains, each record might include the usual time of harvest, the visual features of its outlook, alternative names, or unique functions. In general, lists are the most common form of presenting information in local gazetteers, aside from narratives. The task of collecting structural data from digitized textual lists calls for proper tools to ensure the efficiency of the curation process.

2.3 Comprehensiveness of local gazetteers

A group of scholars at Beijing Astronomical Observatory surveyed libraries and museums to identify extant historical local gazetteers, and located 8,000 distinct titles stored in libraries, museums, public and private institutions world wide (Zhuang et al., 1985). Thirty-five years later, one can already find large electronic databases that deliver historical local gazetteers as scanned images and searchable digital texts. Erudition's Database for Chinese Local Gazetteers currently contains 4,000 titles of gazetteers and the number

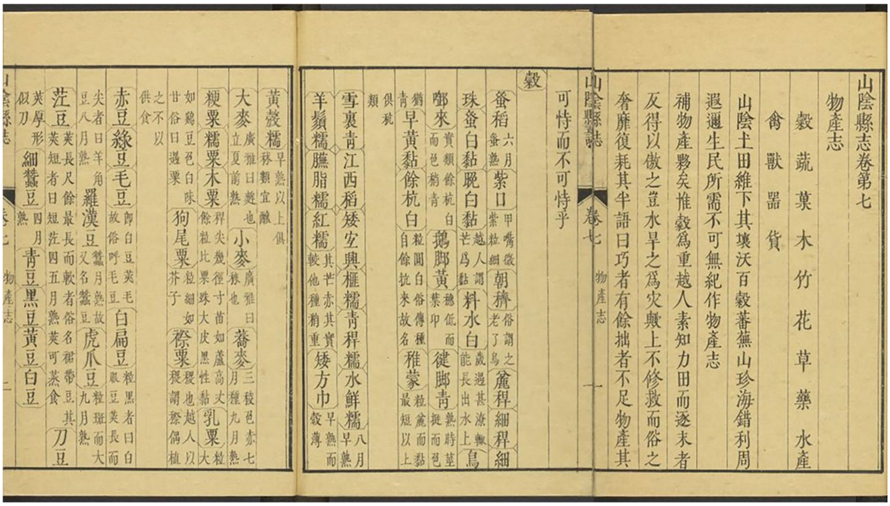
山陰縣志卷第七
物產志
穀 蔬 菓 木 竹 花 草 藥 水產
禽 獸 器 貨
山陰土田維下其壤沃百穀蕃蕪山珍海錯利周
退邇生民所需不可無紀作物產志
補物產夥矣惟穀爲重越人素知力田而逐末者
反得以傲之豈水旱之爲災興上不修救而俗之
奢靡復耗其半諺曰巧者有餘拙者不足物產其
可恃而不可恃乎

Fig. 2 First page of the Local Products chapter in Shangyin County Gazetteer, 1671. (Image source: Harvard-Yenching Library.)

is increasing. Diaolong's Fulltext Database for Chinese and Japanese Ancient Books contains also about 2,100 titles. EastView's Comprehensive Chinese Local Gazetteers Database contains about 7,000 titles. While these databases must have high overlaps in their contents, conservatively speaking they should cover more than 5,000 distinct titles, which is more than half of the extant ones. Based on our estimation, the number of digitized gazetteers should cover more than ¼ of historical gazetteers that were ever compiled in historical periods (by 1949).[5] This high proportion implies that scholars can now use the digitized gazetteers to represent the whole genre with sufficient statistical confidence: if we run a full text search in either of the three databases, the overall distribution of result should roughly reflect the actual distribution in all the historical gazetteers, extant and lost.

The fact that each local gazetteer is itself a data-base consisting of useful local data (the data nature), and that such data are categorized and almost consistently included in the majority of gazetteers (the structural nature) makes the genre a conceptual database for collecting historical data on consistent themes. The comprehensiveness of digital gazetteers to represent the whole genre with statistical confidence further enables such possibility and allows scholarly interpretation based on the digital collection.

3 Research methodology: Treating a genre as a database

Thus, we propose the research methodology of treating the genre of local gazetteers as a conceptual database (since it doesn't exist) and using a large-enough digital subcollection as a probe to represent the genre. A testimony of this approach was carried

[5] We estimated the total number of historical gazetteers ever compiled in history should be between 13,000 to 20,000 titles. Note that many of those have been lost.

out by a Chinese geographer, Chen Cheng-siang.[6] In the 1960's he conducted a project of creating a map showing locust plagues. He manually flipped through 3,000 titles of local gazetteers stored in Hong Kong, Japan, and Taiwan to collect temples devoted to locusts and their locations. Based on this dataset, he produced a map that demonstrated the geographical distribution of locust temples as an approximation of where in historical China suffered from locust plagues (Chen, 1983). His work was replicated by one of the authors in 2016 via digital means: we ran full-text searches with Chen's keywords for the various names of such temples in a digital collection of 2,000 gazetteers and produced another map (Chen, 2016). The two maps have very similar geographical distributions although they are based on two different sets of local gazetteers, which testifies our assumption that a large-enough digital set is capable of reflecting historical phenomena that are generally captured by this genre. The same assumption is also employed by Moretti and Michel et al. to derive their interpretations.

What Chen Cheng-siang has demonstrated for Chinese historical studies is twofold. First, Chinese local gazetteers, when treated as one unit, can be great sources of collecting historical datasets across time and space. This approach is now supported by the large amount of full-text digitized gazetteers. Second, by displaying the locust temples on a map, Chen demonstrates how visualization as a condensed form of his dataset can be useful for observing overall patterns, just as Moretti, Michel et al., and many other digital humanists have pointed out.[7]

Furthermore, in this paper we propose that such overall patterns, produced by visualization techniques based on counting quantities in a dataset, should also serve as entry points to lead the scholars to dig into large collections for details, rather than being the artefacts of study. In LoGaRT, we link statistic and visualization tools to every search result so that our users can immediately see the overall temporal, geospatial, and field-dependent distributions of the search result. We combine such "distant reading" with close reading by allowing users to easily go back and forth between the overall distributions and the individual records in the search results, so that a user can "zoom in" to a geographical location on the map or a peak on a temporal chart to check why the data is so. This close linkage of the two reading modes in LoGaRT allows our scholars to be able to validate each data point while seeing the whole picture, and thus it creates a deep understanding for the scholar of how a dataset is composed of beyond the overview.

3.1 From methods to tools

LoGaRT is a suite of digital research tools that we built to introduce the collective perspective for studying local gazetteers to Chinese historians. This collective approach contrasts with traditional ways of using local gazetteers, in which scholars

[6] Another important endeavor was done by the Central Weather Bureau (Zhong yang qi xiang ju qi xiang ke xue yan jiu yuan, 1981) in Beijing to collect historical climate data. In 1981, the bureau published a map collection consisting of 510 yearly maps of dryness and wetness zones in China from 1470 to 1979. The data were collected from 2200 local gazetteers as well as from other historical genres.

[7] See the above-mentioned China Biographical Database and Mapping the Republic of Letters as some of the examples.

are only interested in reading particular sections of gazetteers from certain regions and time periods to find local information. With LoGaRT, we encourage scholars to take this collective perspective to ask large-scale questions about historical China that require the whole genre to answer. Based on the large amount of digital gazetteers linked behind LoGaRT and the proposed collective perspective, LoGaRT offers evidence from digital gazetteers in response to scholarly questions that reflect phenomena embedded in this genre.

More specifically, LoGaRT helps scholars to pose questions to a collection of digital gazetteers. The questions are generally posed in the form of queries, whether keyword search in digital texts, search in bibliographical records, or the two combined with selections by the scholars. The result of a question is considered a dataset, that is the joint response to the question answered by all the gazetteers in the collection and should be treated as a whole. That said, we think any search result in LoGaRT has a collective meaning, rather than being simply an aggregation of unrelated records. Based on this philosophy, we provide quantitative, statistical, and visual analysis tools to help scholars to understand the resulting datasets from a collective perspective. Scholars can immediately obtain the temporal and geospatial distributions of the resulting dataset, as well as distributions based on any categorical metadata fields, such as dynasties, provinces, or section headings. Seeing the distributions along different dimensions immediately after a search can inform scholars of important overall features of her question reflected by the genre. On the other hand, the distributions also offer a crucial gateway for scholars to examine the dataset in depth: each of the temporal, regional, and categorical dimensions divides the data into smaller clusters, each of which features data points of specific attributes – records produced at a specific time, in a specific region, or classified under specific sections. The scholar can first choose to see the overall distributions, looking from afar, and when spotting unusual patterns she can switch focus to a specific cluster of the dataset, read the records in depth to understand why they are there, and then jump back to the overview to continue her reading of the data. During this process, scholars can also quickly zoom in to another cluster of the same dimension, switch to another dimension, or zoom in to records lying at the intersection of two dimensions. In this way, we provide scholars with multiple perspectives and flexible orders of reading through a dataset in relation to one's question toward the genre.

In order to support the above methodology in a software, we designed the following user workflow to enable every scholars to apply the same methodology with one's specific questions toward local gazetteers (see also Fig. 3):

1. Collecting data across a collection of digital gazetteers (by search or text-tagging)
2. Obtaining the corresponding dataset
3. Applying built-in visual and statistical tools to obtain the overview of a resulting dataset, that is, the temporal and geospatial distributions of the data and also distributions along other categorical dimensions based on metadata.
4. Switching between overview (distributions in various visual representations) and record-view, and among dimensions to understand the data in depth.
5. After digesting the result, adjusting one's query to locate an improved dataset.

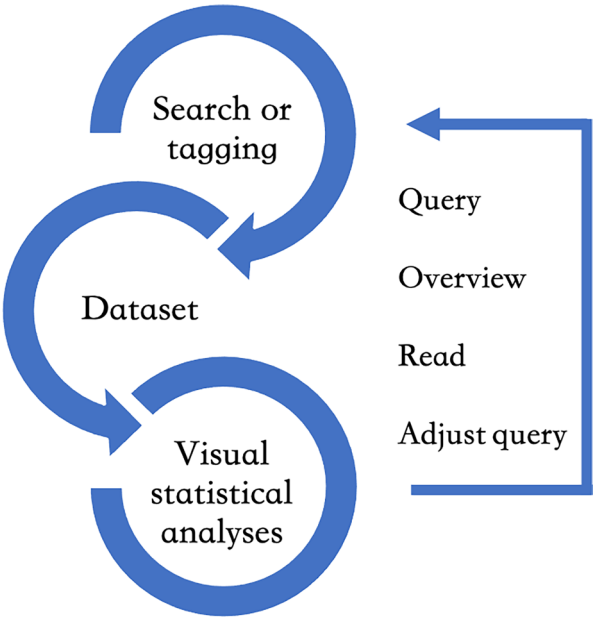

Fig. 3 The workflow of approaching local gazetteers with a collective perspective implemented in LoGaRT

The search function of LoGaRT is provided at four different levels: book bibliography (local gazetteer titles), content structure (chapters and sections), page texts (digital texts), and visual materials (maps and illustrations). A search at any of these levels results in a dataset at that level, and LoGaRT provides visual and statistical tools for users to immediately get an overview of the dataset for its composition.

In addition to search, LoGaRT provides another tool for scholars to collect thematic data in list form: LoGaRT's Tagging Interface assists scholars to convert textual lists to computer-operable data via text-tagging. Scholars can define a thematic "topic" for tagging and its corresponding "tags" (fields, attributes), and associate a text segment with a tag. Once tagging is done on a text, LoGaRT provides an export function to covert the tagged text into a data table, which can then be used for further visual and statistical analyses.

3.2 LoGaRT: a technical overview

After logging in to LoGaRT, a user lands at LGService, the portal where all the functions are presented. As mentioned above, LoGaRT has four levels of data that can be collected through search – Books, Sections, Page Texts, and Pages with Images, each of which is assigned a dedicated working tab on LoGaRT's web interface. A user needs to first choose a data level to work with, and run browse or search accordingly. After browsing or searching at any of the four tabs, a traditional Data Record View shows all the matching records at that data level (see Fig. 4).

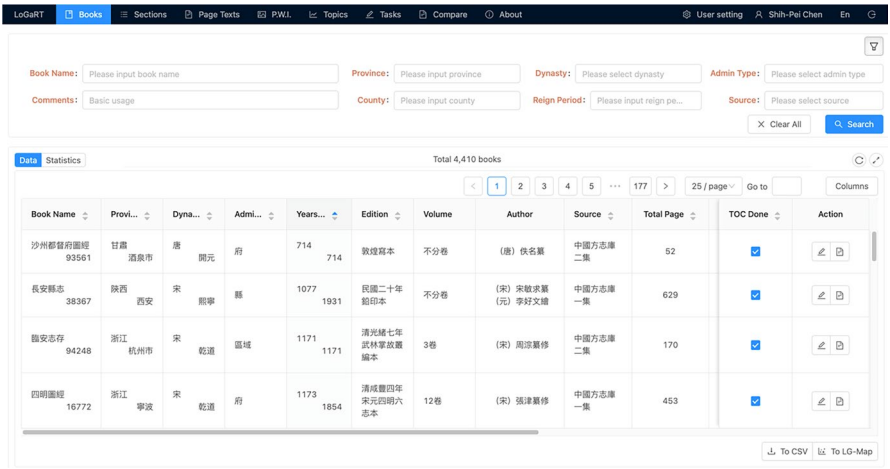

Fig. 4 The Data Record View of browsing all the available books in LoGaRT

The collective analysis of search results is realized by providing actions that take the whole search result into account. For every search at any level, users can immediately obtain the overview of the search result via three visual statistical tools. Currently, LoGaRT implements its own "Statistics View" module that provides categorized statistics based on the metadata of data (Fig. 5), and links to two external visualization tools: LGMap and CHMap. LGMap is customized from the open-source software PLATIN, provides a set of visual analytical tools, including an interactive mapping service, a timeline, and a user-defined pie chart, enabling the user to explore the geospatial distribution of a dataset flexibly and interactively (Fig. 6) (Büttner & Kruse, 2014). In PLATIN, users are able to filter the data by selecting a region on the map, a segment on

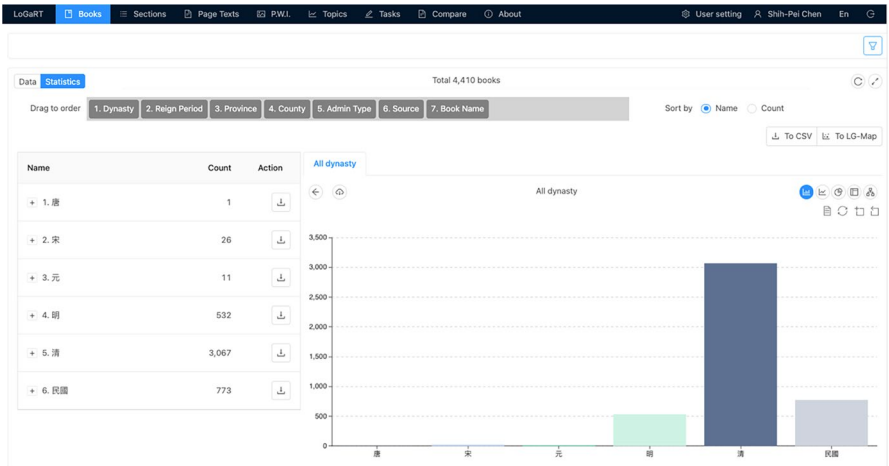

Fig. 5 The Statistics View of a search result in LoGaRT

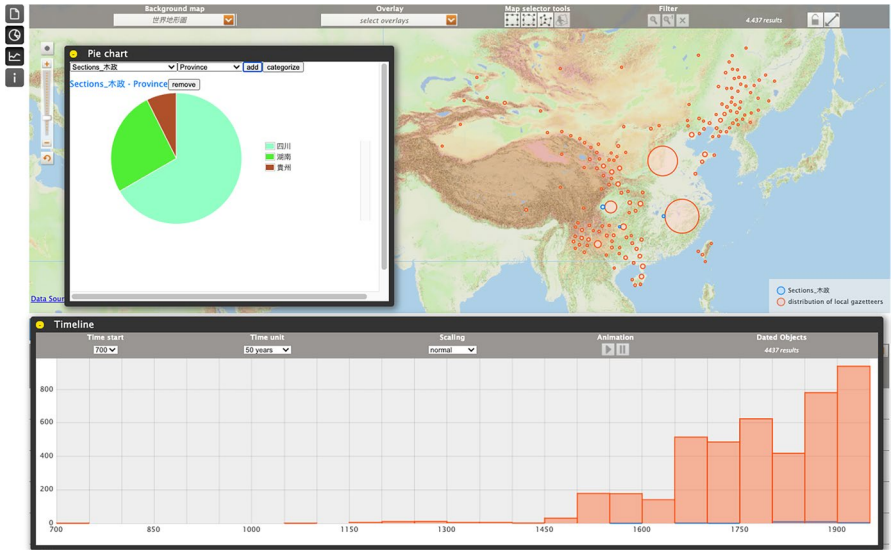

Fig. 6 LGMap displays geospatial and temporal distributions of a search result, with a user-defined pie chart

the timeline, and a piece on the pie chart. All the other visual components then update themselves to show the distribution of the selection. CHMap is another independent WebGIS tool which we link only to LoGaRT's Pages with Images to display visual materials on top of third-party Chinese historical maps (Fig. 7) (Lin et al., 2019).[8]

Ideally, LoGaRT can link to further visualization tools. We chose the above tools due to their interactive design: such design makes them extremely useful to help the user explore the dataset – the user can see the overall distribution along different dimensions, dive into a specific cluster to read the records in detail, and can flexibly combine different dimensions to locate records with more specific features. LoGaRT also provides an export function "Save as CSV" that allows the user to download the resulting data and bring them to other applications for further functionalities, such as to create a professional printable map with desktop GIS software.

3.3 Collective analysis of search results

As mentioned, any search result in LoGaRT can be analyzed collectively with built-in statistical and visual tools. It is worth addressing that both statistical and visual tools are quantitative by nature: in such tools, each data point in a dataset is regarded as "equal

[8] CHMap is a joint project between the MPIWG and the Department of History at Shanghai Jiao-Tong University. This project digitizes and geo-references 4,088 maps produced with western cartographic technology by the Republican China and Japan governments in the early twentieth century. They cover a wide geography of proper China and provide large scale outlooks (1:50,000) to China's populated regions. CHMap was built to bring maps from different hosting institutions to be viewed together (Lin et al., 2019).

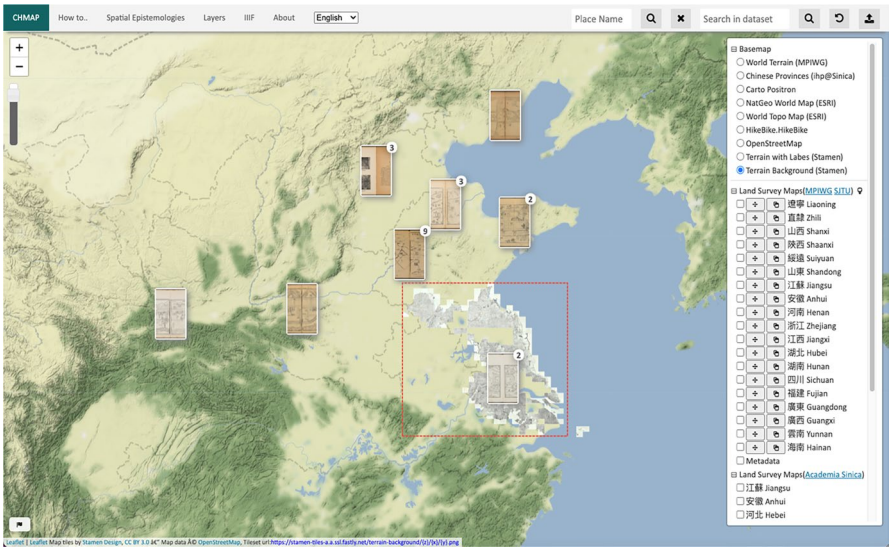

Fig. 7 CHMap displays thumbnails of visual materials from LoGaRT on top of third party open maps

weight" and is assigned of value "1". This is the underlying assumption that visualization tools are based on, so that they are able to count and add up the quantity of data records that possess a certain feature in the dataset. It also implies that such tools must be used for data records at the same granularity level to derive meaningful interpretations. This is exactly why we distinguish the four levels of search rather than mixing them together as "one box search". Each of the four levels of data in LoGaRT also represents a meaningful type of entity that should be viewed collectively.

A Book-level record in LoGaRT represents a local gazetteer. At the moment, LoGaRT is linked with 4,410 titles of digital gazetteers. Among which, 4,000 titles are sourced from a commercial database and 410 titles are open access. One can immediately obtain the overviews of both collections using LoGaRT's statistics and visual tools simply by not giving any search criteria. For example, Fig. 8, generated by LGMap, shows the timelines of both collections, indicating how many gazetteers were published at which year for both the collections. Such an overview helps to understand the overall temporal distribution of

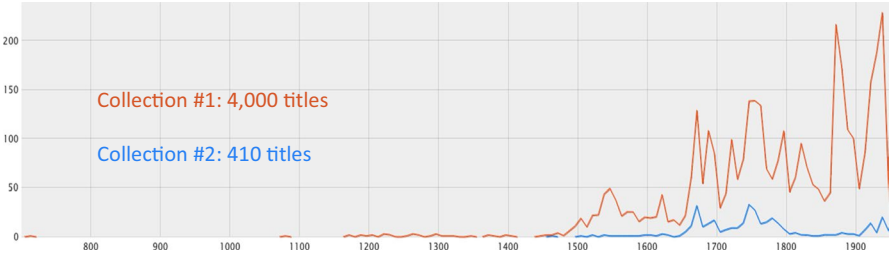

Fig. 8 Temporal distributions of the two digital gazetteers collections linked with LoGaRT in comparison with each other

a collection – where data are richer, which is crucial when interpreting the timeline of a search result so as not be misled by biased data. Such overall temporal distribution itself can also be an artefact of study – to understand why there are peaks and valley at specific times, and further to dig into records from those times to see why it is so. Similarly, geographical distribution of a collection also gives hints to where and why certain regions contribute many or few gazetteers to the collection. Figure 9 allows us to understand the geographical distribution of the whole collection of 4410 titles. When combined with timeline, we can see when gazetteers were produced where.

A Section-level record in LoGaRT represents a record in the Table of Contents. It denotes the section heading, the start and end page numbers, and its hierarchical level in the Table of Contents. The MPIWG is embarking on creating three levels of Table of Contents for all the 4410 titles in collaboration with the Nanjing University of Information Science and Technology. Such work enables us to pull the same thematic section from different gazetteers to be compared (via Section Search), moving beyond the original orders of reading imposed by the gazetteers, and making possible thematic data collection across gazetteers. Similar to Book Search, after a Section Search, one not only sees the Data Record View which lists the sections matching the search criteria but is also able to flexibly choose one of the visual statistical tools to obtain distributions of the search result based on various dimensions. For example, Fig. 10 is a timeline chart showing sections with headings "*xingye* 星野 (star allocation)" and "*fenye* 分野 (field allocation)" over time, the purple and orange lines respectively. The two lines are also presented together with a third green line for the temporal distribution of the collection, serving as

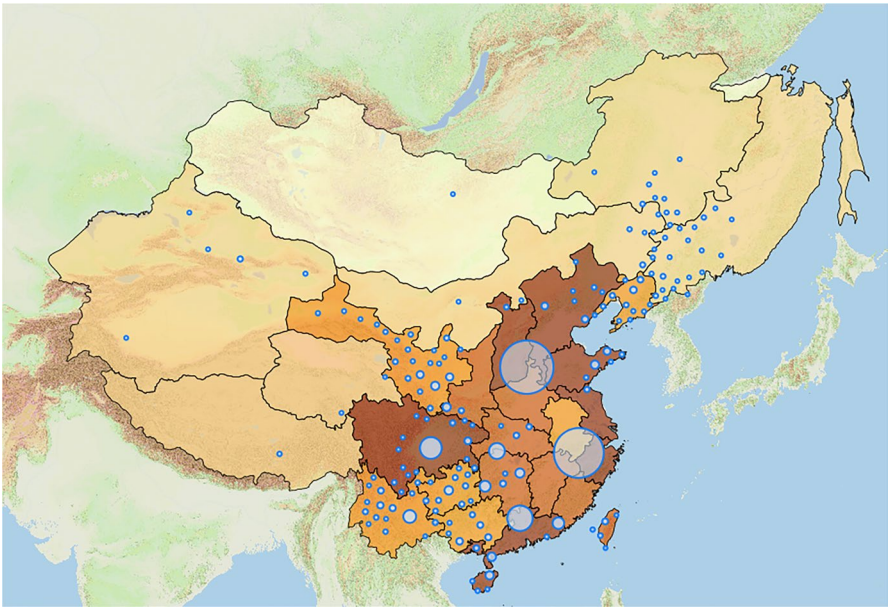

Fig. 9 Graphically distribution of the 4,410 local gazetteers. Background map: 1820 provincial boundaries from China Historical GIS, with colors denoting the number of gazetteers of each province: the darker the color, the higher the amount

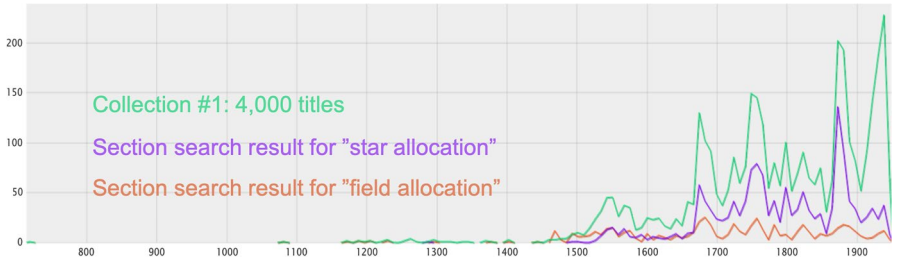

Fig. 10 Temporal distributions of section search results for "xingye (star allocation)" and "fenye (field allocation)" in comparison with the targeting collection

the baseline. Based on comparison of the three, we can see that both "star allocation" and "field allocation" are popular section headings in the imperial period, since they almost add up to occupy roughly 80% of the gazetteers during the imperial times. However, in the Republican era (1911–49) the usage of both headings decreased dramatically, hinting that the practice of referring to astrological objects to identify a locality on earth was somehow discarded. Such phenomenon has not been fully discussed in scholarship and requires further research.

Page Text Search allows scholars to find occurrences of particular keywords on certain pages. It is equivalent to a full text search, while in LoGaRT we use "page" as the atomic unit for counting matches and for summing up visual statistics. Figure 11 shows a recreation of Chen Cheng-siang's map of locust temples with LoGaRT and LGMap. The ability of PLATIN to compare different datasets by color coding helps us to identify the

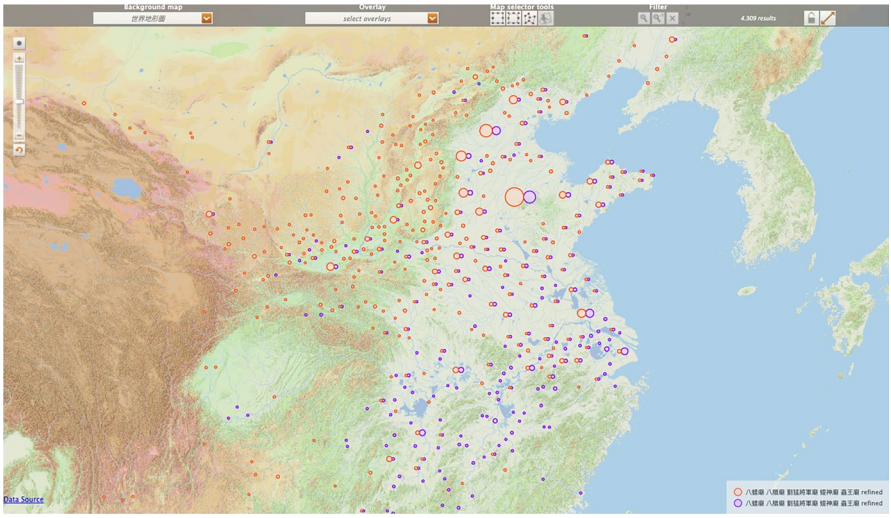

Fig. 11 Geographical distributions of two temples devoted to locusts: orange dots for ba zha miao (Temples of the Eight Farming Gods); purple dots for liu meng jiang jun miao (Temples of General Liu the Fierce)

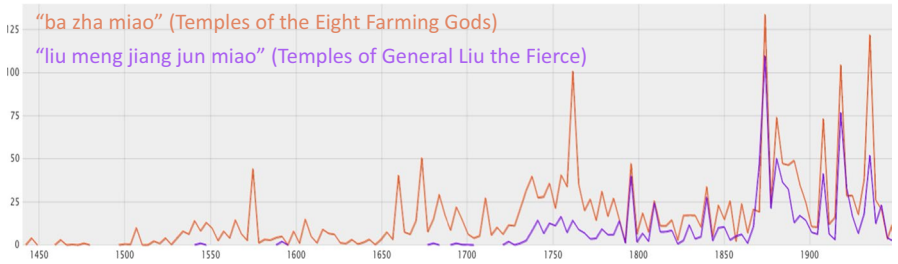

Fig. 12 Temporal distributions for two temples devoted to locusts: orange line for ba zha miao (Temples of the Eight Farming Gods); purple line for liu meng jiang jun miao (Temples of General Liu the Fierce)

phenomenon that the two major temples devoted to locusts in fact each has a different geospatial focus – *ba zha miao* 八蜡廟 (Temples for the Eight Farming Gods) were mainly distributed in the Yellow River basin, while *liu meng jiang jun miao* 劉猛將軍廟 (Temples for General Liu the Fierce) were mainly located in the southern regions (Fig. 11). The temporal distributions also show that the latter appear in local gazetteers during much later time periods (Fig. 12), hinting that the latter was only introduced to local societies at a later stage. This phenomenon was not identified in Chen's article that accompanies his map, and we were able to identify it through the interactive user interface of PLATIN that creates an experimental "playground" for manipulating data (Chen, 2016).[9]

Although local gazetteers are mainly a textual genre, many of them contain visual materials to accompany their textual content: maps showing the administrative unit of focus and its geographical relations to other units in the broader region, city layout maps, star maps (that often go with the section on star allocation), illustrations of mountains and rivers, building complexes, and so on. In addition to images that give geographical information, others related to rituals, instruments, objects, and even photographs (mainly in the Republican era) can be found. We used semi-automatic methods to identify 72,943 pages with images from millions of pages. In addition to tools for scholars to collect, annotate, and analyze images, we defined 21 tags for scholars to categorize an image. Based on pre-assigned tags, book metadata, text appearing on the same page, or browsing page by page, scholars can search through the image collection and collect those of interest. A thumbnail-based Carousel view is designed specifically to allow easy browsing of images. The same set of visual statistical tools are provided as well to help users see overviews and identify patterns. In addition to LGMap, scholars can also send their collections of images to CHMap, which then displays the images together with third-party open map collections. Figure 7 shows an example of 23 images displayed in CHMap, where scholars are able to call up maps from other collections of the same region and compare the two types of maps and their different depictions of space (Fig. 13), which might indicate their unique epistemologies toward living space (Lin et al., 2020).

[9] Please note that the keyword search function for both Section Search and Page Text Search relies on users to combine different terms that are with similar meanings. In other words, LoGaRT so far does not built in any technical mechanisms, such as controlled vocabulary for known personal names and place names or categorizing / normalizing similar terms, to deal with possible synonyms for searching.

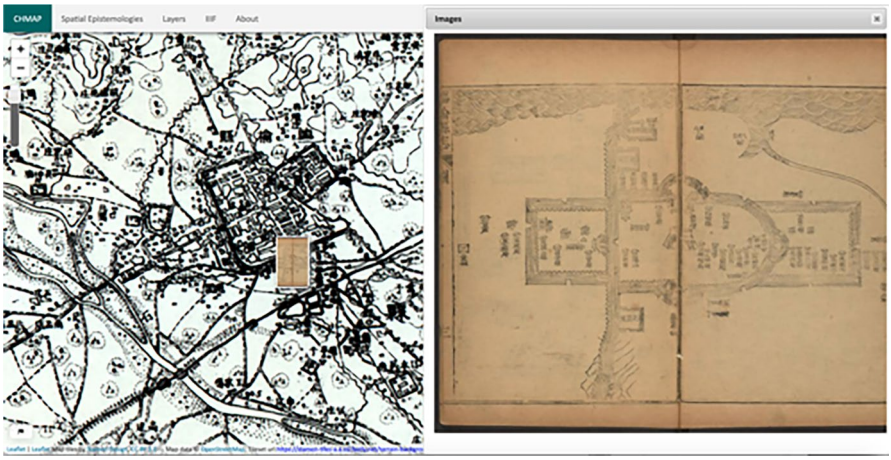

Fig. 13 Side-by-side comparison of a local gazetteer map and a Republican military map displayed in CHMap

3.4 Text tagging

As described earlier, the local gazetteers genre is structural not only because of its consistent way of organizing contents but also due to its general use of lists to represent local data. Such lists are data tables by nature, but are hard for computers to process and analyze. We developed a Tagging Interface to allow scholars to convert such textual lists into a computer-operable format. This interface was first used in the China Biographical Database project to tag lists of officials mentioned in local gazetteers, and has been further developed at MPIWG to extend such tagging facility to allow any types of lists (Peng et al., 2018).

Text tagging has been a well-established method in digital literary research. TEI (Text Encoding Initiative) is often used as the underlying schema to tag literary works, and literary scholars often use existing software to tag TEI.[10] In Digital Sinology, MARKUS is a popular tool developed to tag entities specific to Chinese history, such as personal names, official names, and Chinese dates.[11] The major difference between these and the tagging interface in LoGaRT is that it focuses on assisting users in converting textual lists into data tables like spreadsheets. LoGaRT also provides

[10] Such software includes oXygen and Catma. oXygen is a general-purpose XML text markup software for TEI markup, which automatically validates documents against TEI schema document structure at https://www.oxygenxml.com/. CATMA is Computer Aided Textual Markup and Analysis, developed by the University of Hamburg, Germany, for literary analysis research. See: https://catma.de/webarchive/catma-4.0/home.html.

[11] MARKUS is a web interface markup In addition to manual marking, it also has a semi-automatic marking function, which can automatically mark Chinese historical names, place names, year names, and official names according to its linked dictionaries, and also provides a keyword search function at https://dh.chinese-empires.eu/markus/beta/

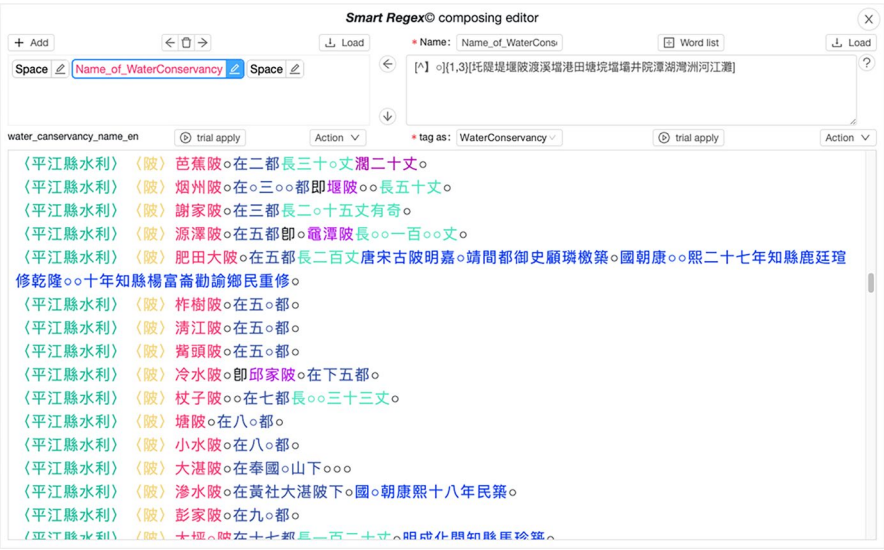

Fig. 14 A segment of tagged text displayed in the SmartRegex Editor

a semi-automatic tagging helper called SmartRegex to detect and tag occurrences of user specified textual patterns via regular expressions.[12] Figure 14 shows a scholarly tagged version of a list of water conservancy constructions recorded in the Yuezhou Prefectural Gazetteer of 1746, where each record is within the same line, with its corresponding data points (features, attributes) tagged in different colors that are predefined by the scholar. A tagged text can then be exported as a spreadsheet and loaded to other software for further analysis. (See Fig. 15.)

4 Case studies

As a voluminous genre spanning more than eight centuries, local gazetteers have served as the basis for research in historical China in wide-ranging topics. Copied, re-edited, and collected throughout the centuries, local gazetteers enacted the social, political, and material composition of a place in written format: the landscape, history, flora, fauna, the taxes and products of a region, the temples and schools, officials and celebrities, local festivities and customs, weather records, and disasters were all documented within. Since 2014, the Local Gazetteers Working Group at the MPIWG has been organizing research that aims to create new forms of digital historical analysis for this genre. Through the development of LoGaRT, the Working Group has also attracted scholars whose individual research interests collectively reflect the eclectic themes documented in the local gazetteers, and whose research needs feed into iterative technical refinements and new, unanticipated

[12] Regular expression is a computer language for expressing specific combination and repetitions of characters and symbols. See: https://en.wikipedia.org/wiki/Regular_expression.

Book Name	Section Name	Place Name	Page No.	County	WaterConservancy	Location	Type	Destruction/Construction	Alternative name	Lengh	Width	Height
岳州府志	水利志	岳州府	722	平江縣水利	芭蕉陂	在二都	陂			長三十 丈	濶二十丈	
岳州府志	水利志	岳州府	722	平江縣水利	烟州陂	在 三 都	陂		煙陂	長五十丈		
岳州府志	水利志	岳州府	722	平江縣水利	謝家陂	在三都	陂			長二 十五丈有奇		
岳州府志	水利志	岳州府	722	平江縣水利	源澤陂	在五都	陂		龜潭陂	長 一百 丈		
岳州府志	水利志	岳州府	722	平江縣水利	肥田大陂	在五都	陂	唐宋古陂明嘉 靖間都御史顧璘檄築;國朝康 熙二十七年知縣鹿廷琯修;乾隆 十年知縣楊富寄勸諭鄉民重修		長二百丈		
岳州府志	水利志	岳州府	722	平江縣水利	柞樹陂	在五 都	陂					
岳州府志	水利志	岳州府	722	平江縣水利	清江陂	在五 都	陂					
岳州府志	水利志	岳州府	722	平江縣水利	鵝頭陂	在五 都	陂					
岳州府志	水利志	岳州府	722	平江縣水利	冷水陂	在下五都	陂		邱家陂			
岳州府志	水利志	岳州府	722	平江縣水利	枝子陂	在七都	陂			長 三十三丈		
岳州府志	水利志	岳州府	722	平江縣水利	塘陂	在八 都	陂					
岳州府志	水利志	岳州府	722	平江縣水利	小水陂	在八 都	陂					
岳州府志	水利志	岳州府	722	平江縣水利	大湛陂	在奉國 山下	陂					
岳州府志	水利志	岳州府	722	平江縣水利	渗水陂	在黃社大湛陂	陂	國 朝康熙十八年民築				
岳州府志	水利志	岳州府	722	平江縣水利	彭家陂	在九 都	陂					
岳州府志	水利志	岳州府	722	平江縣水利	大坪 陂	在十七都	陂	明成化間知縣馬珍築		長一百二十丈		
岳州府志	水利志	岳州府	723	平江縣水利	長善陂	在長 善村	陂					
岳州府志	水利志	岳州府	723	平江縣水利	喻家陂	在二十一都	陂			長三十五丈		

Fig. 15 Exported tagged text as a data table

features. LoGaRT allows researchers to flexibly combine search and tagging to collect historical data across time and space, to quantitatively analyze the data to observe patterns, and to derive deep understanding through switching between distant and closed views. We discovered that the quick turn-around time for conducting a search, seeing the overall patterns, reading representative snippets, and adjusting search parameters has helped scholars to quickly identify phenomena that merits further study or to adjust their hypotheses toward the genre. In this section, we present some case studies of our scholars to demonstrate the possible research applications of the proposed research methodology for collections beyond local gazetteers.

Scholars have used LoGaRT to conduct their research in primarily three ways. First, many have used full-text search to identify key terms as a proxy for the historical presence or absence of certain objects or phenomena. Cao Ling (2017) has collected 53 alternative names used for the crop maize to run a Page Text search in LoGaRT. Reading through the materials and tagging them with her scholarly categories, she provides an approximation of introduction, planting, and transmission of maize in historical China. Li Fupeng and Cao Ling (2017) use similar methods to trace the circulation of a professional manual on mulberry trees and silk production in the local gazetteers to understand how this genre contributes to the circulation of such technological knowledge. Grace Fong is collecting records on "filial daughters" and analyzes how this category of women was established and served in local politics. Snyder-Reinke (2019) looked at the descriptions of infant burials and columbaria in local gazetteers as evidence for shifts in societal attitudes toward infant mortality in historical China. Huiyi Wu looks at the mentions of "westerners" in local gazetteers and discovered that the majority came from the section on Star Allocation of local regions, showing the effort of local officials to use translated western astrology works in the creation of local identity.

Second, scholars have used LoGaRT's Section Search function to isolate the same thematic section across multiple local gazetteers, which has enabled them to extract data about a specific category across geographic regions and/or time periods. In some cases, the presence or absence of certain thematic sections in specific gazetteers was also indicative of large-scale patterns or historical changes. Ian Matthew Miller, for instance, whose research looks at the long-term interplay between social and environmental change, used information in the forestry sections to chart the development of

administrative structures that have governed the logging industry and its associated taxation (Miller, 2020). Based on the Table of Contents of 4,410 titles that were digitized by the MPIWG, Pingtzu Chu and Shih-Pei Chen are now analyzing the data categories employed by the genre to unwrap the underlying knowledge systems.

Third, using LoGaRT's tagging interface, scholars have tagged texts from multiple local gazetteers to collate and compile stand-alone datasets that serve as a curated "digital concordance" for re-use and further investigations. Joseph Dennis—whose research focuses on the history of Chinese print culture, law, and society—tagged any information related to book holdings in school libraries to produce a concordance that maps the circulation of books and knowledge across formal spaces of learning. A first publication derived from this dataset on the role of donations in building local school book collections is just published, in which Dennis demonstrates how this dataset can be used to study knowledge circulation in historical China (Dennis, 2020). Gregory Scott (2020) tagged and analyzed data on reconstruction of Buddhist sites for other purposes. Qun Che collects water conservancy constructions in in the middle stream of the Yangtze River and analyzed their types and function to recover the change of landscape in that region over three-hundred years. Schäfer et al., (2020) also tagged records of mulberry tree related disasters to discuss the nature of local knowledge represented in this genre. Griet Vankeerberghen (2021), when tagging a premodern geographical text with LoGaRT's Tagging Interface, also noted that the tagging process and the mindset of identifying structure behind the text provides new perspectives and thus new understanding compared to close reading.

The possibility of looking at all the visual materials appeared in local gazetteers opens up fascinating research possibilities to study the purpose of such images and their relationships to the texts. A pilot group in 2018 and an extended group of fourteen scholars in 2021 joined the "Visual Materials in Local Gazetteers" working group to use LoGaRT to look at the images in our collection and to explore the roles of visual materials, as opposed to texts, in describing and transmitting local knowledge. They each brought a research project that worked with a specific type of images in the local gazetteers. For example, Anne Gerritsen collected images of rituals and dances from the collection and found that such images often came in sets and they appeared only in specific regions. Jiajing Zhang analyzed the use of latitude and longitude lines used in maps of late Imperial gazetteers to discuss their relations with western cartography (Chen et al., 2020). Shellen Wu and Peter Lavelle looked at 20th century maps to identify the spread of scientific institutions, such as agricultural institutions, in Republican China.[13]

5 Discussion and conclusion

In this paper, we present our research methodology of treating the genre of Chinese local gazetteers as a conceptual database. We describe the genre and its characteristics: being data-centric, structural, and comprehensive. All of those features contribute to designing the software LoGaRT, that allows scholars to collect historical data, to test research

[13] As some of the research results are in process of being published, please refer to the Working Group website for a complete list of the participated scholars and their research projects: https://www.mpiwg-berlin.mpg.de/event/visual-materials-local-gazetteers-0.

hypothesis, and to obtain overall systematic knowledge from the genre, as approximated by a large digital collection of 4410 gazetteers. Following the methodology, we designed a user workflow in LoGaRT that leads scholars to pay attention to the collective meaning of any obtained dataset, whether through searching or text-tagging, by incorporating visual statistical tools that provide a multi-dimensional overview of the data. Such tools also serve as reading guides that divide the dataset into featured clusters for in-depth reading. This process helps scholars to quickly adjust one's hypotheses based on the fast consumption of query results and thus also helps to identify research questions that are worth pursuing. The tagging interface provides scholars a feasible way to collect large-scale thematic data in textual list form across gazetteers and turn them into computer-operable data, and therefore turns this genre into a scholarly enhanced database.

We have also reported on the scholarly projects that are derived from LoGaRT and their research breakthroughs in Chinese history. These successful projects have demonstrated that LoGaRT and its underlying methodology is capable of supporting various kind of research rather than answering only a few research questions.

For interested readers who wish to use LoGaRT, the access is bound with licenses. Currently, LoGaRT links to two digital collections: one open access collection containing 410 titles, and the other commercial collection with 4000 titles.[14] While scholars can register for a free account to use the open access set, the larger collection stays protected by a commercial license and can only be used by MPIWG affiliates.[15] On the other hand, LoGaRT as a software can be freely used by other institutions, and other digital collections of gazetteers can be loaded when proper licenses are in place. In practice, it is achieved by installing an empty shell of LoGaRT on a server that belongs to an institute with interest. Subscribed materials by the institute have to be prepared and processed to match the format required by LoGaRT and then loaded.[16] It is worth noting that it is dependent on law and regional consensus and is still up for debate that whether every usual database license already grants right to the subscriber for processing the contents, including uploading them to third party software. Otherwise, institutions would have to obtain a special license that explicitly grant such right in order to load their subscribed contents to LoGaRT.

As more and more scholars are using LoGaRT to curate historical datasets, we are now exploring the possibility of publishing such datasets so that they can be used independently from their sources while keeping referential linkages to trace back.[17] We argue that such curated datasets should earn their curators proper

[14] The open access collection in LoGaRT are digitized by the Max Planck Institute for the History of Science from the Chinese Rare Book Collection of Harvard-Yenching Library with funding from the Max Planck Society and the Chiang Ching-kuo Foundation's Database Grant (project number DB001-U-17). The commercial collection is Beijing Erudition's Zhongguo Fangzhi Ku. The access is made available to Max Planck Institute affiliates via Berlin State Library's CrossAsia service (https://crossasia.org/).

[15] Free LoGaRT accounts can be registered at: https://logart.mpiwg-berlin.mpg.de/LGServices2. For accessing the commercial collection through LoGaRT, interested scholars can apply to become a visiting scholar of MPIWG. In the meantime, we are working with Berlin State Library to find legal solutions to extend this access to all German research institutes.

[16] LoGaRT uses MySQL as its backend storage, in which metadata and texts are stored. LoGaRT defines its own data schema for storing the above contents.

[17] We are exploring this new publication format as Digital Concordances. See: https://www.mpiwg-berlin.mpg.de/news/brill-and-max-planck-institute-history-science-sign-agreement-digital-concordances.

academic credit since it takes serious efforts and scholarly knowhow to digest the raw texts, to design a tagging framework that is able to accommodate the implicit knowledge structure behind the texts, and to find missing data points by searching through other sources, among other tedious tasks, to produce reusable datasets that will benefit the field.

We envision that the proposed research methodology can be extended to other textual collections of any regions and languages *if* a digital sub-collection is comprehensive in representing / sampling the underlying genre. In such cases, scholars should be able to use the digital sub-collection to collect evidence that can be projected to the underlying genre, and thus providing tools to help the user to collect data and to lead the user to pay attention to the collective meaning of every search result with statistical and visual aids should be as useful as we demonstrated with the local gazetteers. Please note that such genres don't have to possess strong inner structures for this methodology to be useful. Also, when interpreting such projections, one should note that every genre has its own limitations – aspects of the underlying historical conditions that the genre cannot reflect – due to its production process. Therefore, even with digital tools that accelerate the exploration process, it depends more than ever the scholar's craft to properly interpret such computational results.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The data that support the findings of this study are composed of two collections: an open access one and a commercial one. The open access collection was generated during this study and is available to the public under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International license (CC BY-NC-SA 4.0) via the RISE Catalog: https://rise.mpiwg-berlin.mpg.de/resources. The commercial collection used in this study is made available to Max Planck Institute affiliates via Berlin State Library's CrossAsia service (https://crossasia.org/).

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Ba, Z. (2004). *Fang zhi xue xin lun*. Xue lin chu ban she.

Büttner, J., & Kruse, S. (2014). *PLATIN (Place and Time Navigator) - A tool for the interactive visualization of geospatial and temporal data*. Web. http://platin.mpiwg-berlin.mpg.de/. Accessed 12 Sept 2022.

Cang, X. (2013). *Fang zhi xue tong lun (zeng ding ben)*. Hua dong shi fan da xue.

Chen, C.-S. (1983). *Zhongguo wen hua di li (Cultural geography of China)*. San lian shu dian.

Chen, S-P. (2011). *Information Technology for Historical Document Analysis* [Ph.D. dissertation, National Taiwan University]. Airity Library. https://www.airitilibrary.com/Publication/alDetailedMesh1?DocID=U0001-0208201117414200. Accessed 12 Sept 2022.

Chen, S.-P. (2016). Remapping Locust Temples of Historical China and the Use of GIS. *Review of Religion and Chinese Society, 3*(2), 149–163.

Chen, S.-P., Hammond, K., Gerritsen, A., Wu, S., & Zhang, J. (2020). Local gazetteers research tools: Overview and research application. *Journal of Chinese History, 4*(2), 544–558.

Dennis, J. (2015). *Writing, publishing, and reading local gazetteers in imperial China, 1100–1700*. Harvard University Asia Center.

Dennis, J. (2020). The role of donations in building local school book collections in the Ming dynasty. *Ming Qing Yanjiu, 24*(1), 46–66.

Edelstein, D., Findlen, P., Ceserani, G., Winterer, C., & Coleman, N. (2017). Historical research in a digital age: Reflections from the Mapping the Republic of Letters project. *The American Historical Review, 122*(2), 400–424. https://doi.org/10.1093/ahr/122.2.400

Harvard University, Academia Sinica, and Peking University. (2021). *China Biographical Database*. Web. https://projects.iq.harvard.edu/cbdb. Accessed 12 Sept 2022.

Heng, Z. (Ed.). (2012). *Di fang zhi zhi shi zu zhi ji nei rong wa jue yan jiu - yi fang zhi wu chan guang dong wei li (Research on the organization of knowledge and content excavation of local gazetteers: taking the example of local products in Guangdong)*. Anhui Normal University.

Hinrichs, E., & Krauwer, S. (2014). The CLARIN research infrastructure: resources and tools for eHumanities scholars. In N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)* (pp. 1525–31). European Language Resources Association.

Huang, W., Ba, Z., Yao, J., & Chu, G. (1993). *Fang zhi xue*. Fudan University Press.

Jockers, M. L. (2013). *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press.

Lai, X. (1983). *Fang zhi xue gai lun*. Fujian ren min chu ban she.

Lin, N., Che, Q., & Chen, S.-P. (2019). *CHMap: Land Survey Maps of China*. Web. https://chmap.mpiwg-berlin.mpg.de/. Accessed 12 Sept 2022.

Lin, N., Chen, S.-P., Wang, S., & Yeh, C. (2020). Displaying spatial epistemologies on Web GIS: Using visual materials from the Chinese Local Gazetteers as an example. *International Journal of Humanities and Arts Computing, 14*(1–2), 81–97.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., Alden, & Lieberman, E. (2011). Quantitative analysis of culture using millions of digitized books. *Science, 331*(6014), 176–182.

Miller, I. M. (2020). *Fir and Empire: The Transformation of Forests in Early Modern China*. University of Washington Press.

Moretti, F. (2000). Conjectures on World Literature. *New Left Review, 1*, 54–68.

Moretti, F. (2005). *Graphs, maps, trees: Abstract models for a literary history*. Verso.

Peng, W., Cheng, H., & Chen, S. P. (2018). From full text to table: Semi-automatic capture of official information in local zhijiaoji. *Digital Archives and Digital Humanities, 1*(1), 79–125.

Schäfer, D., Chen, S.-P., & Che, Q. (2020). What is local knowledge? Digital humanities and Yuan dynasty disasters in imperial China's local gazetteers. *Journal of Chinese History, 4*(2), 391–429. https://doi.org/10.1017/jch.2020.31

Scott, G. A. (2020). The Post-Taiping Reconstruction". In G. A. Scott (Ed.), *Building the Buddhist revival: reconstructing monasteries in modern China* (pp. 43–88). Oxford University Press.

Sinclair, S., & Rockwell, G. (2016). *Voyant tools*. Web. http://voyant-tools.org/. Accessed 12 Sept 2022.

Snyder-Reinke, Jeffery. (2019). Cradle to grave: baby towers and the politics of infant burial in Qing China. In Thomas Mullaney (Ed.), *The Chinese deathscape: grave reform in modern China.* Stanford University Press.

Tu, H.-C. (2011). *Taiwan History Digital Library*. Web. http://thdl.ntu.edu.tw/. Accessed 12 Sept 2022.

Vankeerberghen, G. (2021). Writing memories: The sanfu huangtu on western Han Chang'an. *Monumenta Serica, 69*(2), 353–386. https://doi.org/10.1080/02549948.2021.1989778

Zhong yang qi xiang ju qi xiang ke xue yan jiu yuan. (1981). *Zhongguo jin wu bai nian han lao fen bu tu ji* [Map]. Di tu chu ban she.

Zhuang, W., Zhu, S., Feng, B., Wang, S., & Zhongguo ke xue yuan. Beijing tian wen tai (Beijing Astronomical Observatory). (1985). *Zhongguo di fang zhi lian he mu lu*. Zhonghua shu ju.