Appendix: The Digital Gazetteer of Song China

By Ruth Mostern with Elijah Meeks

In 1958, Sinologist Hope Wright published a work entitled *An Alphabetical List of Geographical Names in Sung China*.¹ The *Alphabetical List* is an index to every jurisdiction in the Song spatial administrative hierarchy named in one or more of the following three Song texts: the *Song History* Geography Monograph, the 980 *Records of the Universal Realm in the Taiping Era* (*Taiping huanyu ji*), and the 1085 *Treatise on the Nine Territories in the Yuanfeng Reign* (*Yuanfeng jiuyu zhi*). The *Alphabetical List* consists of 4,009 headwords, including all Rank One circuits (*lu*), Rank Two prefectures (*fu, zhou, jun, and jian*), Rank Three counties (*xian*) and county-rank *jun* and *jian*, and Rank Four towns (*zhen*) and garrisons (*sai* and *bao*) that existed at any time during the Song dynasty, along with centers of state industry (mines, foundries, and commodity markets) located in prefectures, and information about the number of cantons (*xiang*) in each county, the resident (*zhu*) and guest (*ke*) population of each prefecture in 980 and 1085, the civil rank of each prefecture and county, the designation of counties that served as prefecture seats, the military-ceremonial designation, if any, of each prefecture, the latitude-longitude coordinate of each prefecture, and the distance of each county from the seat of its parent prefecture.²

¹ The *Alphabetical List*, originally published in Paris by the Centre de Recherches Historiques of the École Pratique des Hautes Études in 1956, was reprinted as a second-generation photocopy in Albany, New York in 1992 by the Journal of Song-Yuan Studies. It is currently out of print. Hope Wright died soon after the publication of this work, which is her only publication.

² Wright’s coordinates are expressed to zero decimal places. One degree of latitude ranges in size from 69 miles at the equator to 45 miles at a latitude of 45 degrees. This is therefore a highly imprecise measure of location. The military-ceremonial rankings used for Song prefectures were Military Training (*tuanlian*), Defense Commission (*fangyu*), and Military Commission (*junjiedu*).
Wright’s compilation is the most comprehensive print source for Song geography in any language, and it is the foundation for the data analysis and maps included in this book. It has three virtues that make it particularly valuable as the basis for a spatial database. First, it aspires to complete fidelity to the source material, incorporating all attributes about jurisdictions listed in the original Song sources. Second, it records bi-directional relationships among spatial entities. That is, all of the Rank One entries list the prefectures under the jurisdiction of each circuit, all of the Rank Two entries list the parent circuits and child counties of each prefecture, and all of the Rank Three entries list the parent prefectures and child towns and garrisons of each county. As a result, the data can be readily verified and cross-checked. Finally, with a degree of precision unique among works of this type, the *Alphabetical List* tracks all of the occasions when entities were promoted or demoted, founded or abolished, moved from the jurisdiction of one parent entity to another, conquered, or renamed. The care with which Wright documents that information is in large part the inspiration for this book project.

*My Digital Gazetteer of Song China*, a relational MySQL database and geographic information system (GIS) based primarily on Wright’s work, is the source for all of the maps and data analysis in this book. It is being made freely available for public download. The purpose of this appendix is, first, to evaluate the quality and usability of the information I have created and to reveal where remaining problems may lie. Second, it explains the mechanics of designing and populating the database in order to make my judgments visible and provide advice for others who wish to attempt similar work. Finally, it briefly locates the database and research in the field of historical GIS, the methodology for integrating spatial analysis and visualization with historical scholarship.
How Trustworthy is the Data?

Data quality for digital history is an important topic. The research for this book and database have required me to apply thousand-year-old information to tools which require commensurability between data items of diverse provenance, and which require prolix historical sources to be translated into quantifiable form. Like other digital historians, I am making demands of source material that its creators never envisioned. The challenge can be invigorating. As Willard McCarty has recently put it, “the twin computational requirements of complete explicitness and absolute consistency” open up a space for the scholar “to refine an inevitable mismatch between a representation and reality.”

Similarly, according to Steven Ramsay, databases are heirs to a long tradition of humanistic genres such as taxonomies and indexes that require scholars to make decisions that imply “a hermeneutics and a set of possible methodologies that are themselves worthy objects for study and reflection.”

Still, using historical information for quantitative analysis raises a number of problems. Essential data values may be missing if the sources in which they are preserved are not extant. Data may be regarded as uncertain if its prejudices, opaqueness or other characteristics make it difficult to interpret—a condition inherent, to some extent, in any kind of historical analysis. It may be inaccurate if it diverges from what one believes to be true, or imprecise if it is meaningful only when broadly measured. It may be ambiguous if

---

particular data values cannot be related to one another with certainty.\textsuperscript{5} All of these problems, and errors that result from them, can be introduced at any stage, beginning with Song land surveyors and extending to the interlocutors who collated their results, through generations of copyists, to the authors of reference works, and finally to efforts to generalize and interpret multiple data sources. All of these problems affect the Digital Gazetteer for Song Dynasty China.

Hope Wright’s Alphabetical List is based on only three Song sources—though it is true that they are canonical ones developed and held by the court. Some instances of renovation to the spatial landscape are not recorded in these three sources, particularly in the Southern Song. As Chapter Five documents, for instance, some rapid and Conditional changes to the landscape of wartime and postwar Huainan chronicled in the Drafted Documents are not recorded in the Song History, and Ming and Qing historical geographers complained that they could not write accurate Persistence and Transformation accounts for the early Southern Song. Similarly, recent scholarship has revealed that at least two instances of spatial change during the tumultuous New Policies reform era are not recorded in the Song History, and can only be identified from Shui Anli’s Handy Maps of Historical Geography.\textsuperscript{6}

The Song rank system itself is ambiguous. As Chapter One explains, \textit{jun} could function either as Rank Two entities, in which case they were a type of prefecture; or Rank


Three entities, in which case they were commensurate with counties. There is no way to assign them clearly to one category or the other, since Song practice did not clearly distinguish them. I have designated the 264 jun which had no dependent counties and which reported to prefectures as Rank Three entities. However, they could arguably be considered as Rank Two prefectures on the basis of the administrative rules that governed their personnel selection and dictated their capacity to field troops. This is an inherent ambiguity that the complex and redundant administrative system of the twelfth century could readily accommodate, but data analysis requires clear delineation. This is one reason why I am making my raw data available. Subsequent scholars are welcome to recalculate my results using different parameters if they wish.

For reasons like these, it should not be considered surprising that all of the contemporary scholars, who have worked with the same fairly limited range of Song geographical texts—even those who have worked solely with the Treatise on the Nine Territories—have come up with somewhat different results from one another. This is clearly visibly through a comparison of the results offered by authors including Ma Yuchen, Zhou Zhenhe, Kawakatsu Mori, Nie Chongqi and Winston Lo. Wright herself introduces additional errors. Her work is simply a typewritten list of names with inconsistent formatting. 68 counties named in the text of the entries for other entities do not have their own headwords, and had to be painstakingly identified by hand. However, I have ascertained that Wright’s work is the best base data for this project by comparison with other paper sources and by checking my digital facsimile of her data against other digital gazetteers. I have also conducted internal checks to reduce errors.
Among paper sources, the most useful reference work other than Hope Wright’s is Guo Li’an’s 郭黎安 Collected Interpretations of the Song History Geography Monograph (Songshi dilizhi huishi 宋史地理志匯釋). For every county and prefecture named in the Song History geography monograph, the Collected Interpretations presents the Song History text, citations to other Song documents about it, the name of the contemporary entity that most closely corresponds to the location of the Song one, its population according to the Treatise on the Nine Territories, its military-ceremonial rank (if any), and its population. However, it does not include towns, markets, or any other Rank Four entities; and since its basis is the Song History, it does not include any jurisdictions that are not listed in that single source. Its format, organized around extensive quotations, makes it extremely valuable for tracing the history of particular entities, but difficult to use for quickly identifying and recording change types.

Another important source, the Great Dictionary of Chinese Historical Place Names (Zhongguo lishi diminhg da cidian 中國歷史地名大辭典), edited by Shi Weile 史為樂, Deng Zixin 鄧自欣 and Zhu Lingling 朱玲玲, is comprehensive with regard to source and rank and includes the modern co-locations of historical entities, but with more than 50,000 entries including the names of physical features and religious sites as well as administrative entities, and covering 2,000 years of Chinese history, there is no feasible way to extract only the relevant material. In any event, it does not include comprehensive or dated Persistence and Transformation summaries. I have used all of these sources for reference,
but I have not systematically incorporated their information into the first release version of the database.

Three digital sources for Song geography have been essential for my work: first, the China Historical GIS (CHGIS), second, the database developed from Tan Qixiang’s *Historical Atlas of China (Zhongguo lishi dituji)* for the Chinese Civilization in Time and Space (CCTS) digital atlas, and finally, the Hartwell GIS.⁹ These are all extremely useful sources, but none of them supersedes the *Digital Gazetteer for Song Dynasty China* that I have developed. The CHGIS, launched in 2001, is an ongoing joint venture between Harvard University and the Fudan University Center for the Study of Chinese Historical Geography (*Lishi dili yanjiu zhongxin*) that is intended to provide continuous spatial change data in GIS format for the entire imperial era, from 222 BCE to 1911 CE. Currently in Version 4, it is the most careful work of historical scholarship among the digital projects. However, its coverage extends only to the county rank. In addition, the range of attributes that it records for each entity is more limited than those that Hope Wright has tracked. Most problematic for my research, the time-series data is being completed province by province, and the Version 4 Song data does not extend to the entire realm.¹⁰

---


⁹ The CHGIS and the Hartwell GIS are available online at http://fas.harvard.edu/~chgis. The public version of the CCTS consists of a series of map animations. Copyright restrictions prevent the public circulation of the data. I am grateful to CCTS director Fan I-Chun for providing me with a copy of the Song subset of the CCTS data for personal use.

¹⁰ Peter K. Bol and Ge Jianxiong. “China Historical GIS.” *Historical Geography* 33 (2005), 150-152.
Both the CCTS data and the Hartwell data include the jurisdictions that existed at two snapshot dates in the Northern and Southern Song: 1080 and 1200 for Hartwell, and 1111 and 1208 for CCTS. Since the central concern of my research has been to investigate the politics of spatial change over time, neither of these sources, which map the Song realm at only two dates, each separated by a century and divided by a major regime change, is adequate to my purposes. These two datasets also include very limited attributes information about each named place. Nevertheless, both of these datasets have been invaluable. Comparing them with the Hope Wright data from the same date has allowed me to identify discrepancies and validate the degree of overlap among the datasets. In addition, while neither of them exhibits exemplary precision in its assignment of spatial coordinates, the ability to utilize spatially referenced datasets has been extremely helpful.¹¹

I have collated my draft Digital Gazetteer database against the CHGIS, CCTS, and Hartwell sources. Given the differences in the structure and coverage of each database, it has been difficult to develop precise comparisons. Nevertheless, some generalizations are possible. There is only 28% overlap between the entities in the Digital Gazetteer and those in the Song subset of CHGIS Version 4; though this surprising comparison is largely a function of the incomplete coverage of the CHGIS. There is a 68% match with the Hartwell data, and a 71% match with the CCTS data. I performed the comparison using pinyin transliterations or Chinese characters, feature types, approximate geolocation (though precise spatial references varied between each source), and circuit designations where available. More systematic comparison between each of the available databases of Song political geography would involve statistically sampling each database, comparing

¹¹ Peter K. Bol, “An Overview of Work on an Historical GIS of China: An Introduction to Robert Hartwell’s Work.” China Historical GIS, 2000 (Online at <http://www.fas.harvard.edu/~chgis/>). Unfortunately, this paper does not discuss the source of the content of the database, only the spatial reference methodology.
entities to primary sources references, comparing the parent-unit relationships designated in each database, and heads-up visual comparison between datasets visualized as layers in a GIS.

Finally, just as geographers “ground truth” remote sensing data by collecting information about observable spatial phenomena through field work, I have “ground truthed” my database through close readings of primary sources and historical works about Song historical geography. The vast consolidation of counties and prefectures in Guangnan during the 970s, for instance, is reflected in both the Digital Gazetteer and the historical record. So is the founding of prefectures on the Tangut frontier in the early twelfth century, and the consolidation of counties and prefectures in North China during the New Policies reforms. And so on. In every case, nodes of regional spatial reorganization that are visible from the sources are also reflected in the data, and vice versa.

For these reasons, while I am sure that I have not succeeded in capturing every single instance of spatial change that occurred during the Song dynasty—indeed I am confident that Song commentators from the late eleventh century onwards believed that such an aspiration was unreasonable—I am confident that the database supersedes all previous sources for Song spatial history. Furthermore, I am certain that the conclusions and generalizations that I have drawn from the data are valid, even if the precise numbers I have introduced may diverge slightly from historical reality—such as it was—in particular cases.

From the Alphabetical List to the Gazetteer to the GIS
The print version of the *Alphabetical List* was converted into a Word document in China through a private subcontractor of the National Bureau of Statistics. While this introduced a certain amount of typographical error, it was well worth the effort to create a machine readable version, and the cost was moderate. It was trivial to migrate the formatted Word document into an Excel spreadsheet, and to convert Wade-Giles transliteration into pinyin.\(^{12}\) Digitizing the *Alphabetical List* using Optical Character Recognition (OCR) software was not an option because the quality of the Wright manuscript, a combination of typewriting and Chinese handwriting in third-generation facsimile, did not permit it.

The database that I have developed based on the *Alphabetical List* is a digital gazetteer. The term gazetteer generates some confusion in Chinese studies, since it is the English word most commonly used to gloss the Chinese local geographies known as *difangzhi*. As the expression is used by geographers, a gazetteer refers to a place name directory, like the list at the back of an atlas. In a networked computing environment, gazetteers refer to databases organized around named places and their locations, and they have become an essential to all spatial search infrastructure. A gazetteer is distinct from a geographic information system (GIS), although the two are often used together and frequently translated into one another. A gazetteer is a database about named places, while a GIS is a system for storing, analyzing and displaying georeferenced information. In a GIS about land cover or public health, for instance, data organized by spatially located

\(^{12}\) Thank you to Michael Fuller for the initial Word to Excel conversion. The Word to Excel conversion was performed within Microsoft Office. The Wade-Giles to Pinyin conversion used an online tool at MandarinTools (http://www.mandarintools.com/pyconverter.html), with error conversion in Excel. Thank you to Bao Shuming from the China Data Center at the University of Michigan for helping me to arrange digitization services. Thank you to Yu Lini and her staff for conducting the work.
named places might play a relatively limited role. By contrast, spatial information for historical scholarship is most commonly derived from texts. Unlike data digitized from maps or observed from sensors, surveys, or satellites, texts are saturated with place names. The critical task for historical geography is to determine how those names relate to entities that can be mapped, when those entities existed, what attributes are essential to understanding and classifying them, and how this information changed over time.\textsuperscript{13}

Historical GIS systems like the Vision of Britain/Great Britain Historical GIS (GBHGIS) and the CHGIS are therefore all designed around gazetteer architecture.\textsuperscript{14}

A gazetteer can be used, as I have done, to answer research questions about historical political geography \textit{per se}. It can also be utilized as one component of a data architecture. The GBHGIS gazetteer supports millions of data values for British demographic history, the entities in it are linked to the full text of early modern travel narratives for search and display, and it is the backbone for the extraordinary Vision of Britain through Time digital library. Future users of the \textit{Digital Gazetteer of Song Dynasty China} can integrate new data sources with it for research and visualization on the geography of social networks, commerce, transportation, artistic practice, or any other topic that would benefit from spatially explicit treatment.

The \textit{Digital Gazetteer} data model is highly indebted to the one developed by CHGIS, extended to accommodate the additional attribute information that was available from the \textit{Alphabetical List}. Each \textit{Alphabetical List} headword was initially designated as

\textsuperscript{13} The pathbreaking Alexandria Digital Library gazetteer standard defines the core elements of a gazetteer entry as including at least one name, at least one feature type, and at least one location. All of the elements in the ADL standard can be time-stamped and associated with source references. See Linda Hill, \textit{Georeferencing: The Geographic Associations of Information} (Cambridge: MIT Press, 2006), pps...The CHGIS data model, from which my gazetteer model is adapted, is designed in accordance with the ADL model and is fully interoperable with it. (CHGIS).

an entity. This entity can be understood as a unique and arbitrary numeric identifier that indexes the multiple names, ranks, attributes, parent-child relations, and change events associated with a particular place. That is to say: any place can be conceptualized as a collection of dynamic information. Since Song places experienced frequent changes in name, parent-child relations and other characteristics, the information about each entity is located in two related tables, each indexed to the entity table. A table of historical instances, and a table of attributes.

The historical instance table stores classical gazetteer data: place names, feature types, and spatial change information, such as when the place existed, how it came into existence (whether it existed at the founding of the regime, or was established, annexed, promoted, demoted, et cetera), and how the place ended. The attribute table contains information about civil and military ranks, population, subsidiary units not covered by the database model such as cantons, markets, or salt and coin works. It also records the sources that Hope Wright records for that information. Other subsidiary tables list change types, jurisdiction types, sources, attribute types and geolocation, along with other information that makes data analysis more straightforward. For instance, the table of political jurisdictions includes the transliterated and Chinese name of the jurisdiction and its rank in the spatial hierarchy. The drawback of separating this information into multiple tables is that it is impossible to simply peruse the entity table and see dependent jurisdictions, population or geographic location of a given place. To do so requires data processing. However, given that places change in fundamental ways over time, this method provides for the most accurate processing of regime-level changes as well as dated snapshots.
For example, Ningyuan 宁远 in Guangnan West circuit had three different names over the course of the Song, three different jurisdiction types, and was transferred multiple times. In 1150 it was known as Jiyang and was a Rank Two zhou, while in 1080 it was a Rank Four town known as Linchuan, and in 1050 it was a Rank Three county known as Ningyuan. Based on Hope Wright’s research, the Digital Gazetteer also includes its population according to each primary source she consulted (all different), as well as its prefecture rank and the number of cantons it governed during its time as a county.

Discovering and synthesizing this data required more effort than simply recording a single entry in the Alphabetical List. Her place name index referenced these variously named and ranked entities separately, with entries imperfectly cross-referenced. By contrast to an alphabetical print gazetteer, database software requires very syntax to designate linkages, which are not necessary in a print gazetteer entry. Gazetteers like the Alphabetical List can accommodate inconsistent language for designating relationships among entities. In some cases, Hope Wright has given a full account of entity data within a single entry, while in other cases relationships among named places after reconciling overlapping claims about a given political unit.

Once the database was complete, it was possible to run many kinds of queries, from producing snapshot maps to exploring complex patterns such as the changing ratio of towns to cantons over time and between regions. Entities experienced many changes that signaled the end of one historical instance and the beginning of another. For the purpose of this book, the most important were those that resulted in the appearance or disappearance of a given county or prefecture feature. Counties could be promoted into prefectures or demoted into towns, and they could either be merged or split. The
Alphabetical List does not systematically differentiate between similar types of events, and frequently designates the same event using different terms in cross-referenced entries. In order to achieve consistent results, related change subtypes have been analyzed together.

One way to analyze spatial change is to count the number of change events that occurred between two dates. Another is to compare the number of jurisdictions that existed at two different dates. The first is an event-based approach, and the second is an entity-based approach. The results that each method produces are rather different. The event-based approach reveals the amount of state activity directed toward the spatial landscape, while the entity-based approach measures the density of the state presence in different regions of the realm as it changes over time. I have used both approaches in this book as the particular question at hand dictates.

Also once the draft database was complete, it was possible to begin georeferencing the data. The difficulty and cost of mapping several thousand data points was alleviated by our access to Hartwell, CCTS and CHGIS geodata. In addition, the Alphabetical List the distance and bearing of counties from their prefectural seat. Unfortunately, the prefecture reference is at the scale of zero-decimal point latitude and longitude. To properly geolocate all counties, and even prefectures in some cases, the locations of entities for which no geospatial data was available were extrapolated based on the location of a parent entity. Both extrapolated geodata and references imported from the other datasets were tested for distance from their parent's center-point to eliminate outliers. Prefectures and counties that were placed, sometimes by Hope Wright herself, at latitudes and longitudes in the ocean, were moved to shore.
The geolocation of counties, and prefectures to a lesser extent, stands as the greatest weakness of the current database. With time and funding, individual locations can be more precisely plotted to achieve better spatial analytical accuracy. It would be feasible, though painstaking, to reference print gazetteers with modern co-locations and to use more sophisticated extrapolation algorithms. Town geolocation could be extrapolated from parent counties and plotted individually as well. For this version of the database, the goal was to be able to visualize generalized phenomena on maps and to perform spatial analysis, not to create an accurate model of each Song jurisdiction. Following CHGIS, we have represented all counties and prefectures as point locations.

The mapped results of the geolocated data are relationally accurate. There is at least a 97% match between county circuit affiliation according to the Alphabetical List, and their circuit location based on Hartwell's polygons (prefectures are 100% accurate). Outliers can easily be manually corrected. All hierarchical relationships are correct: database queries match the references to the historical record. The true gaps in the database reflect its status as a digital version of the Alphabetical List, which was based in turn on only three Song sources. In particular, data for the Shaanxi frontier lacks some entities and important changes that are depicted in the historical record.

**Spatial Analysis and Historical Research**

It is my hope that this discussion will contribute to more sophisticated digital mapping and spatial modeling for pre-industrial geographies in China and elsewhere as well. Paper maps tend to represent all places as if they were modern nation-states with
clear international and domestic borders. Yet that is not the way that their own rulers and subjects envisioned them. Maps and databases for regimes like the Song have to represent space with fidelity to historical logic, in the form of networks, relationships, and demography in addition to clear lines. Gazetteer databases like the *Digital Gazetteer of Song China* provide the potential for modeling spatial entities and spatial relationships according to hierarchical networks and other categories and concepts developed by people in the past; and it is my hope that spatial history incorporating quantitative data and utilizing spatial analysis based upon such information can raise and answer new questions about how people created, apprehended, and managed the landscapes in which they lived and the results of such activities.

With the advent of more accessible software tools, a corpus of texts and exemplars, and an active scholarly community, historical GIS—perhaps framed in a more ecumenical way as spatial history or, as J. B. “Jack” Owens terms it, geographically integrated history—is emerging as a distinctive field. As Owens has recently observed, “Historians are noticing GIS because they normally deal with processes in complex, dynamic, nonlinear systems and, therefore, demand a means to organize a large number of variables and identify those variables most likely implicated in the stability and transformation of such systems.” GIS and related technologies make it possible to integrate a wide range of human, environmental, and other spatial phenomena into a single system and analyze their convergences. It is only necessary that they can be located together in time and space. Using GIS, Owens has traced the development of a cohesive oligarchy in Southern Spain and its trading and smuggling networks. Geoff Cunfer has reopened the debate over the


causes of the North American dust bowl of the 1930s by disproving the spatial correlation between intensive agriculture and dust storms. David Bodenhamer and his collaborators have identified spatial and temporal trends in American Protestant history. Challenges remain. For instance, geographers and computer scientists need to improve handling of temporal change and uncertainty, and historians need to develop spatial content and become more comfortable with computers, geography, and quantitative data. Incorporating sudden ruptures and slow transformations into the same analysis and giving appropriate weight to each phenomenon is difficult. Most important, while spatial analysis can reveal geographical patterns that would otherwise have been undetectable, historians still need to use the traditional tools of their trade to describe and explain how and why distinctive spatial phenomena came into existence, and what effect they have had upon human activity. It is my hope that the book and gazetteer that I have completed can contribute to this collective endeavor.

---


19 Gregory and Ell, especially Chapter 6 and Chapter 9 on future challenges.