The China Historical Geographic Information System (CHGIS)
Choices Faced, Lessons Learned

Peter K. Bol
Center for Geographic Analysis, Harvard University
July 2007

For the Conference on Historical Maps and GIS
Nagoya University
August 23-24, 2007

In this paper I consider, in retrospect, the choices that the China Historical GIS project has faced. I hope that these choices, and the lessons we have learned, may prove to be of value to others undertaking an historical GIS project.

In brief, CHGIS (currently in Version 4) is a spatial database covering the period 222 BCE to 1911 CE. This database include the administrative hierarchy from the regime capital to the county (*xian* 縣) level. At present a varying number of towns (*cunzhen* 村鎮) within the county are included for two points in time: 1820 and 1911. Counties and towns are given as point files. Prefectures and higher order units are given as polygons for the territory administered and point files for the administrative capitals. County polygons are also available for 1911. The core datasets are "time-series" which include all administrative changes dated to the year. The 1820 and 1911 datasets are "slices-in-time" and pertain to the administrative structure and major towns around those years. The project currently aims to cover the eighteen core provinces, home to 90% of the population, with the exception of the dynastic coverage of the Qing empire in 1820.

From the user's perspective an historical GIS may appear to be digital maps. However, from a designer's perspective a GIS is always a database in the first place.

1. Individual effort versus collaborative scholarship

The immediate inspiration for CHGIS was the work of one man, the late Robert Hartwell, who had been building a historical GIS (hGIS) for Chinese history on his own in conjunction with his prosopographical database (China Historical Software, Inc.), now known as the China Biographical Database. Hartwell did this on his own, creating datasets for Tang in 741, Northern Song, Liao, and Xi Xia in 1080, Southern Song, Jin, and Xi Xia in 1200, Yuan in 1290, and Ming in 1391. However in order to do this he used county boundaries as of 1990, which he obtained from Lawrence Crissman, G. William Skinner, and William Lavely, to identify the counties at various points in the past, merging them, dividing them, and renaming them as the case demanded. [1] I had first come to learn about the use of GIS to study China as a member from 1992 to 1997 of the now defunct Joint Committee for Chinese Studies (of the ACLS and SSRC), which dispersed some research grants and fellowships. Hartwell's efforts came after and were inspired by the development of GIS for China by others: G. William Skinner (UC Davis)[2], who had led the way in regional system analysis of China's history and had turned to GIS, William Lavely (University of Washington) who was then building the China in Time

[1] Hartwell set out his vision in Robert M Hartwell, "A computer-based comprehensive analysis of medieval Chinese social and economic history," in Victor H Mair and Yongquan Liu, eds. Characters and computers (Amsterdam; Washington, D.C.: IOS, 1991: 89-121 The Hartwell dataset are available for free download from CHGIS. For a discussion Hartwell's method and its limitations see Peter K. Bol, " Introduction to Robert Hartwell's Historical GIS of China," CHGIS website: http://www.fas.harvard.edu/~chgis/ and Bol, "The Hartwell Historical GIS of China," Proceedings of the 2000 Annual Meeting, Pacific Neighborhood Consortium/Electronic Cultural Atlas Initiative.

[2] For Skinner's projects see http://han.skinner.ucdavis.edu/.

and Space (CITAS) database,[3] and Lawrence Crissman (Griffith University) who has developed a large number of GIS datasets for Eurasian countries for license.[4] Crissman had collaborated with Skinner and did much to get Hartwell started. Later, through the Electronic Cultural Atlas Initiative[5] and its director Lewis Lancaster, I had the opportunity to meet Fun I-chun of Academia Sinica who had learned about the potential of GIS from Skinner and was planning to create an historical GIS based on the eight volume Atlas of China Historical Maps *(Zhongguo lishji ditu ji*) edited by Tan Qixiang,[6] and Ge Jianxiong, then director of the Tan's Institute (now Center) for the Study of China's Historical Geography.[7] After Hartwell and Marianne Colson Hartwell died in 1996 and 1997 his estate, including his projects, went to the Harvard Yenching Institute. It was then that I was faced with the prospect of ensuring that Hartwell's unfinished GIS would become available to scholars.

The Hartwell GIS may have been a viable solution for his purposes, and it was based on a method that made it possible for a single individual to proceed, but when we went to enlist the help of the historical geographers at Fudan University in completing the work we came to conclude that it was not satisfactory as a foundation for the application of spatial analysis to China's history. This led to a revamping of the research and publication strategy. But more importantly it became clear that a truly adequate historical GIS for China would have to be a collaborative project, involving many researchers in several countries. To do this we needed funding and to secure funding, it seemed to me, we needed the support of all those who had an interest in the application of GIS to China's history. This made sense—broad agreement that this was important and that we should build a system that would work for all—was likely to persuade foundations that this was a valuable investment. This proved to be the case. The Luce Foundation provided a grant of $750,000 (2001-2004). However, Fudan's original assumption that the work could be finished within three years has turned out to be a gross underestimation; we now expect to finish by then end of 2008. Continuation has been made possible by two grants from the National Endowment for the Humanities totaling $556,000 (2005-2008), supplemented by $50,000 in matching funds from the Harvard Yenching Institute and Harvard's Asia Center. International management and advisory committees have been ways of keeping the field informed and ensuring that its interests are represented.[8]

Even with adequate funding the project would not have been possible without the devotion and expertise of a full time project manager at Harvard, Merrick Lex Berman,[9]

---

[3] Datsets currently available from CHGIS http://www.fas.harvard.edu/~chgis/ . For information about the project and datasets (as of 1997) see http://citas.csde.washington.edu/.

[4] Available through the Australian Consortium for the Asian Spatial Information and Analysis Network (ACASIAN) at http://www.asian.gu.edu.au/.

[5] http://www.ecai.org/

[6] Tan Qixiang 谭其驤, ed. Zhongguo li shi di tu ji 中國歷史地圖集. 8 vols. Shanghai: Di tu chu ban she, 1982.

[7] The Institute has since become the Center for the Study of China's Historical Geography. Ge has become the director of the Fudan University Library and has been succeed by Man Zhimin. Its website is at http://yugong.fudan.edu.cn/default.asp.

[8] The Management Committee: Bol, Crissman, Ge, Skinner (replaced by Mark Henderson), Tang Xiaofeng (Peking University); Lavely has attended management meetings as well. For the Advisory Committee see http://www.fas.harvard.edu/~chgis/

[9] Berman's many papers on CHGIS are available at CHGIS http://www.fas.harvard.edu/~chgis/ . See also his "Boundaries or Networks in Historical GIS: Concepts of Measuring Space and Administrative

and a Fudan project manager, Man Zhimin, the vice-director and now director of the Center. And of course, for reasons that will be elaborated on below, it would not have been possible without the research of Fudan's senior historical geographers, led by Zou Yilin, and their assistants.

We had the advantage of being able to learn from contemporary China GIS work by Skinner, Lavely, and Crissman and to obtain an initial database design by Crissman, and although it was much changed during the course of work it showed how a system might work. [10] We also benefited from the earliest and most elaborate historical GIS, that of Great Britain, and several other European projects. Unfortunately, Academia Sinica withdrew from the proposed collaboration and has gone on to develop the Chinese Civilization in Time and Space system based on the Tan Atlas, but does not provide GIS layers for download.[11]

My work as director of the project has been to facilitate cooperation between managers, institutions, and scholars, without having the expertise or knowledge of any of those contributing to the work, and to bring the perspective on an interested end user.

2. A research-driven GIS versus an infrastructural GIS

The distinction between a research-driven GIS versus an infrastructural GIS is similar to the difference between the material one borrows from a library for a specific research project and the library itself. An infrastructural hGIS is thus like a library: it aims to serve many different research questions. Thus rather than being a vehicle for analyzing a particular body of data (e.g. all known census data) it is meant to provide the common basis, the infrastructure, that many different researchers will use in the spatial analysis of their own data.

CHGIS aims to include all places for which data exists, recognizing that over time the amount and kind of data available changes. To illustrate this, suppose one wanted to trace the growth and distribution of population over several hundred years based historical record with population data collected by county or prefecture. One will almost always find that the names of the administrative units were changed, some boundaries were adjusted, and new units were created by division or merger. In order to analyze the population data one would need a GIS that accurately depicted the administrative units as they changed over time. In the case of CHGIS, the base GIS is the record of changes in the administrative units and the changes in the points, lines, and polygons that digitally map them. Because it is meant to be a base hGIS, CHGIS does not provide the population records, that is the task of the user; instead it provides the base on which population records can be analyzed. However, because different research projects use the common base hGIS all the datasets created in the course of the research will be compatible and shareable. A demographer may thus share population datasets with the institutional historian who has a dataset of official appointments, the intellectual historian with a dataset of academies, the economic historian with a dataset of tax receipts, and the historian of religion with a dataset of religious sites.

Geography in Chinese History," Historical Geography Vol. 33, Special Issue: Emerging Trends in Historical GIS.

[10] Crissman, "CHGIS Draft Database Design and Geocoding System," CHGIS, Dec. 2000; This and the current database design are available at CHGIS http://www.fas.harvard.edu/~chgis/.

[11] Viewable at http://ccts.sinica.edu.tw/

CHGIS is establishing a common "base" GIS for Chinese history: a single, common basis which can be used to represent, analyze, and share all Chinese historical data with temporal and spatial attributes To accomplish this we trace all changes in administrative units from the founding a centralized bureaucratic empire in 222 BCE (for database reason we do not use the year 0, thus years BCE are off by one year) to the end of the imperial period in 1911 CE. In later stages the project can be extended into the present and further into the past. There were several levels in the imperial administrative system. At the highest level there were the dynastic states. In some periods there was a single dynastic state such as the Former Han (202 BCE-8 CE) or Tang (618-907). In some periods there were multiple states as in the Northern and Southern dynasties period (386-589). And in some periods foreign states invaded and held territory such as the Liao (905-1125) or Qing (1644-1911). We include all states that held some territory in the general Chinese cultural area. At the lowest level were counties (*xian*). The number of counties fluctuated but usually settled at around 1200 for the core provinces. Between the state and the county level there were intermediate administrative units. The most important and constant intermediate level unit was the prefecture (*jun, zhou, or fu* 郡州府), of which there were usually about 300. The prefecture oversaw the county administrations and reported directly to the center. Between the prefectural and the dynastic state levels there was a changing array of units such as inspection circuits and, eventually, provinces.

There is an argument to made that a hGIS project should be guided by research questions in the first place rather than an infrastructural agenda. When, for example, a research agenda only requires knowing the places for specific period and region to fit a specific data resource, the academic leadership may not be interested in building an infrastructural hGIS that covers a longer period. A national historical GIS is an expensive project requiring commitments of scholarly resources and funding for many years. Nevertheless those who undertake more focused research-based projects ought to think in terms of how the datasets they create can contribute to the building of a larger hGIS.

In creating an infrastructural historical base GIS three issues must be kept in mind.

1. It must provide coverage with sufficient detail and scale to suit different kinds of research agendas. Some questions may be national in scope, requiring only a map of provinces. Other questions may call for regional or local analysis, requiring maps that give information at the county level. This means that the GIS be detailed enough that users can zoom in or out to the level that is most useful to analyze their data.
2. It must cover this extended territory through history. Human and natural geography are not constant. Rivers change course, new settlements appear, and the boundaries of administrative units change. An historical GIS must track something that changes, define when and how it changes, and maintain continuity through change.
3. It must ensure that the level of detail it provides and the changes it traces are appropriate to the kinds of information that researchers draw from the historical sources. In other words, it must adopt as its base something that fits what users need.

3. Goals of an historical GIS: spatially-enabled historiography or historical geography?

There is a difference between seeing an hGIS as a tool for spatially enabling the study of history in the first place (which is the CHGIS view) and seeing an hGIS as a digital historical geography (which is how a geographer might see it).[12] The value of GIS in historical studies is that it enables the user to perform a thorough spatial analysis of empirical data, something that was costly and labor intensive when spatial analysis depended on drawing paper maps. Many of the advantages of using a GIS are practical – they allow historians to do draw maps on their own at minimal cost. But historians have begun to use GIS to go beyond mapping their data to analyze the spatial relationships in the data with ease and exactness and to see connections and explanations that earlier scholarship had missed.[13]

How we define the goals of historical GIS (hGIS) has consequences for the project. Clearly an hGIS involves historical geographical research because knowing where administrative units and named places were and when they were valid is fundamental to spatial analysis. But if our goal is to enable spatial-temporal research then our approach to building the GIS will be guided by the historical data available to us, rather than by a goal of representing the historical geography of the past to the maximum extent (which is in itself a truly valuable enterprise, but a different one). Thus rather than trying show the administrative structure of a regime to its fullest possible extent (although we do this for 1820) CHGIS aims to maximize detail for the eighteen core provinces where the bulk of the population resided, because this is where the historical data is richest.

The structure of CHGIS reflects China's changing administrative structure because in past and present data was collected according to this structure. The system of local administrative units was overseen by officials sent from and reporting to the capital. The capital bureaucracy collected data from local units that was essential to maintaining state access to labor and production, and some of this data entered the permanent historical record. In addition the central government mandated the conducting and reporting of land surveys and household registration, held censuses and compiled national administrative geographies. This information was compiled and reported by administrative unit. Although there were regional variations, historical legacies, and special exceptions, the administrative structure is known for most periods and well populated regions.

In addition, for the last thousand years of Chinese history in particular CHGIS aims to include cities and towns outside of the county and prefectural capitals. The need to include settlements that were not themselves administrative capitals has to do with

---

[12] See the exchange between Karl E. Ryavek, "Manchu Empire or China Historical GIS? Re-mapping the China/Inner Asia Frontier in the Qing Period CHGIS," Inner Asia 6, and Peter K. Bol, "Historical Geography or a Spatially-Enabled Historiography?: Reply to Ryavec," Inner Asia 7 (2005): 249-56.

[13] For example Goeff Cunfer has been able to challenge Worster's argument that the "dust bowl" of the 1920s and 30s was the result of over-plowing by combining land-use and meteorological data with dust storm records and archaeological data to show that dust storms were primarily a consequence of drought and did not result from plowing. Amy Hillier's construction of a GIS that combined real estate grading maps with the exact location of bank loans has allowed her to challenge the claim that banks refused mortgages in "red-lined" districts. (These and other recent examples are collected in Past Time, Past Place: GIS for History Ed. Anne Kelly Knowles. Redlands, CA: ESRI Press, 2002).

fundamental changes that took place during the middle period of Chinese history. By the end of the eleventh century the population had nearly doubled to about 100 million. The state apparatus did not expand with the increase in population and it was no longer able either to control the ownership and distribution of land or to limit commerce to official markets. The extraordinary increase in commercial activity outside of state control brought with it unprecedented urbanization and new commercial networks. Once it was possible for market towns to grow larger than the county or prefectural seat under whose administration they fell, we need to recognize both economic networks and administrative structures.

The work of tracing the development of non-administrative towns is made possible, first for the south and later for the north, by the appearance in the twelfth century of a new kind of historical source: the local gazetteer (*difang zhi*). The combination of a free market in land, increases in the private wealth of non-official households, and the spread of printing and educational institutions resulted in the emergence of highly literate local elites who saw the writing of local history as a means of commemorating their locale and their families. Local histories added far more place-specific information to the historical record: settlements in the county, religious establishments, schools, population, villages and towns, tax quotas, arable land, crops, names and dates of civil service degree holders, etc. Many included maps. Local gazetteers were produced through a collaboration between the local administration and local scholars and printed. Although we have only about thirty extant original editions of gazetteers from the twelfth to mid-fourteenth century, later editions included the contents of earlier editions. By the seventeenth century all places were required to compile local gazetteers in a standard format. There are over 8000 extant titles for the pre-1911 period. In the 1980s and 1990s all county and prefectural units were once again ordered to compile local gazetteers in a new format. The amount of spatiotemporal data available for the study of Chinese history thus made a quantum leap forward.

However, historians are just beginning to learn how to use GIS software and spatial analysis. In presentations to historians we have seen that many expect CHGIS to provide them with data about the places in the database, not just the places. In fact the CHGIS database is a resource for understanding changes in the administrative structure over time, but this holds interest for only a limited number of historians. We have thus begun to explore ways in which the CHGIS project can provide datasets that are based on the CHGIS base. Examples include the “China W” dataset from Yue Zumou, G. Wm. Skinner, and Mark Henderson, the 1820 Buddhist temple sites from M. L. Berman, and the forthcoming Song period demographic datasets from Wu Songdi and Ruth Mostern.

4. A historical gazetteer versus a historical GIS

We need to make a distinction between a GIS and a gazetteer. A GIS has two parts: a database of places with their attributes (a “gazetteer”) and a database of GIS layers (the points, lines, and polygons) which are located in space according to coordinates, usually longitude and latitude (hereafter x/y coordinates). In order to create a GIS there has to be a gazetteer, but when our first concern is a GIS we often see the gazetteer as a secondary issue. By this I mean that in creating a GIS, particularly when we begin from a paper map, we record only those places for which we have x/y

coordinates. This is how CHGIS began, but we could have treated gazetteer creation as a separate enterprise—in retrospect I think we should have.

A gazetteer is a record of named places; a historical gazetteer is a record of named places that existed in the past. Places, in past and present, have "attributes;" that is, different kinds of data that pertain to that place. For example, the attributes of the place named Jinhua xian could include the prefecture it belonged to, the date it was established, its "feature type" (administrative seat, rather than a temple or market), and so on. From this perspective the x/y coordinates are two among many possible attributes. X/y coordinates are necessary if one wants to represent the place in space relative to other places, but they are not necessary to create the gazetteer. Because GIS began as a way to analyze space today, it was not apparent that there was a problem in determining the x/y coordinates since these could be extracted from detailed paper maps.

However, in historical studies we deal with places which we know of from the written historical record but whose precise location has yet to be determined. Thus if we limit ourselves to places with known x/y coordinates we inevitably leave out many historical places, although their x/y coordinates might be determined in the course of future research. In the CHGIS case we began from the idea of a GIS and thus made defining the x/y coordinates of places the basis for building the gazetteer.

The conclusion I reach is that historians need divide the work of historical gazetteer building from hGIS building. An historical gazetteer can include all sorts of attributes, but a base gazetteer would at minimum provide attributes that help us locate a place relative to other places. These would include the following general categories, each of which can be divided into more specific attributes (e.g. "Name" can be subdivided into Vernacular name, Romanized name, Alternate names, etc.)

Name
Subordinate to/Superior to (or Includes/Included in)
Preceded by/Followed by
Time period
Feature type
Data source

This approach would allow us to extract much of the data from historical geographies. Building the hGIS would then be based on the gazetteer.

It should be noted that "place" is not so easy to define. A place can be considered as an object that can change its name yet keep its location, but it also can keep its name and change its location, while keeping the other attributes of its "identity," such its feature type and relationship to the administrative hierarchy, A "place" is a matrix of variables that change independently—name, location or footprint, feature type, and relation to administrative hierarchy. In the CHGIS database we fragment the changing matrix along the only track that is linear and constant: time. Time moves in a single direction, so it can be easily searched backwards and forwards. All the other variables are non-linear, in fact analog factors, in that they can suddenly change in a totally unpredictable manner.[14]

[14] I thank M. L. Berman for this paragraph.

5. Points versus Polygons

Maps today depict administrative entities as territorial units with clearly marked boundaries. In the United States, administrative units are likely to have names different from the town in which the administrative seat of government is located. But this has not been the typical assumption in China's history. From early times the administrative seat, rather than a boundaried territory, identified the administrative unit. To illustrate, if the state of Massachusetts were in China, its administrative seat would be called Massachusetts rather than Boston. In earlier historical sources, the state would be identified on a map not by bounded space but by the point location of its administrative seat. In China, the national capital was the seat of the court (chao 朝), the center of power and wealth, and for much of history, a planned city. It was the "pivot of the four quarters," standing against the uncultivated wilds (*ye* 野), just as the light of the "son of heaven" in the central plain was in the midst of the forces and deities in the darkness of the surrounding mountains.[15] In almost all extant Chinese maps prior to the eighteenth century the depiction of administrative units, which obviously had jurisdiction over surrounding villages, is by named points rather than bounded territories. This is true for national maps and for local maps. Surviving county and prefectural maps from the thirteenth century represent prefectures and counties as central places. Even the subcounty administrative units that were apparently territorial units without administrative seats of their own in earlier times (the "township" or "canton" *xiang* 鄉) are given as labels rather than territories. However, beginning in the eleventh century geographies and local gazetteers typically listed the distance from the administrative seat to the border in the eight compass headings (N, NE, E, etc.) and from the border to the next administrative seat in that direction. In many cases these were apparently along roads, but in some cases mountain ranges and rivers are noted. Thus although maps did not draw boundaries local administrators clearly had an idea of how far their jurisdiction extended.

This brings us to the problem. Should an hGIS that includes periods and areas for which there are no reliable maps devote resources to identifying boundaries (polygons) or not? The CHGIS project has decided to draw boundaries for prefectures through history and for 1911, based on late Qing and early Republican maps, it draws boundaries for counties. These boundaries are the basis for prefectural boundaries, and based on references in the historical record (for we lack maps with boundaries for earlier periods) to changes in the jurisdiction counties and prefectures, adjustments to the boundaries are made accordingly. There are three questions we need to ask about boundaries. First, given that the starting point is modern maps, how reliable are the boundaries as we move farther back in time (and is there a point in time when administrative capitals are outposts, so to speak, in the wilderness, which only have practical jurisdiction of nearby settlements and thus do not need boundaries)? Second, what useful analytic purposes do boundaries serve? Third, given the investment of research time in the definition of boundaries, is this a good investment. If not, what is the alternative?

[15] My thinking about this owes much to Sarah Jo-Shao Wang, "Out of control: The Place of Shanshui (Mountains and Rivers) in the Geographical Discourse of Early Imperial China" (PhD thesis, Ann Arbor, University of Michigan, 1999); and Paul Wheatley, *The Pivot of the Four Quarters: A Preliminary Enquiry into the Origins and Character of the Ancient Chinese City* (Chicago: Aldine, 1971).

It seems to me that as we go further back in time all boundaries become ever more approximate, and thus their depictions is ever more likely to create a false certainty. The one analytic value of boundaries is that GIS software measures area accurately and can thus be used for density analysis. However, there is a viable alternative to boundaries: GIS software can create "Thiessen polygons." A Thiessen polygon is generated by assuming that a boundary is equidistant between known points; this is a heuristic device, to be sure, but its accuracy increases as the number of points increases. Thus the more towns we can locate within a county, the better we can approximate county boundaries, and so too for the prefecture. Thiessen polygons are one solution to density analysis, but not the only one.

The density of human activity is directly related to the topography of the landscape—people generally do not settle thickly on the slopes of mountains. Thus density analysis ought not to be based on the area of an administrative unit but on its topography. Here sophisticated GIS software can be used to model the distribution of the population across the landscape by taking into account known settlements, measuring slope, comparing soil quality, simulating rivers, and so on.

The conclusion I reach from this is that, at least from an analytic perspective, a hGIS ought to devote its time to maximizing the number of points (i.e. known cities, towns, and villages) for which there is attribute data, rather than try to create boundaries. In fact, by maximizing the coverage of features in the landscape we will have better means of identifying where people in the past thought the boundaries were.[16]

6. Slices-in-time versus time-series

When we began CHGIS our goal was to create a hGIS with slices-in-time; that is, to give the administrative structure and major towns as they existed at particular moments in the past. We thought we could rely on the same data sources used for the slices-in-time in Tan's Historical Atlas of China and add several new layers through new research. For example, Tan's Atlas has national coverage for 1582 and 1820; we planned to add at least one time-slice between those years.

A continuous time-series is more true to historical realities—administrative changes took place irregularly and historical data was often collected on an occasional basis—and thus the ideal historical GIS would be a continuous time-series that recorded

[16] But unless polygons can be tracked with accuracy, it may be better to do without them entirely. An administrative unit can be both an administrative polygon and its respective administrative seat, but the two can change independently of each other. Our way of handling this has introduced inconsistencies in the database system. M. L. Berman explains it as follows: The administrative hierarchy is tied to the ID of the SEAT (PointID), not the ID of the AREA (PolygonID). This was done because there are Parents that have no defined boundary but all items in the CHGIS do have defined seat locations. However, for any particular historical instance CHGIS lists both the PointID and PolygonID. This introduces an inconsistency because a new instance is only created when there is a change in Placename, Feature Type, Polygon Change (for polygons), or Point Change (for points). The problem is that the Seat could move during the period of time when the Polygon remained the same. The instance should have been split into two unique instances, one for each unique Seat location. A more consistent approach would be to keep track of two separate objects: a Point object for the administrative Seat, and a Polygon object for the jurisdictional area. There would be a one-to-many relationship that could be captured in a separate table (such a Seats Table). By trying to keep the jurisdiction and seat together as a single named instance results in a logical error when the Seat changes but there is no unique instance created for the change and no new instance is created for Polygon instances until the jurisdiction area changes.

all changes. This would make it possible to generate slices-in-time for any year over a 2100 year period. We did not at first adopt this approach because we assumed that it required excessive research time. However, the researchers found that even if they aimed to add new slices in time they would have to trace back through administrative changes, region by region. In short, in the effort to produce new slices in time a body of research was being created that in its level of detail and accuracy went far beyond anything that was hitherto available. By adopting a times-series model these new findings could be preserved and made accessible to all. The disadvantage of the shift to a continuous time-series is that national coverage from 222 BCE to 1911 is now dependent on finishing regional coverages; the advantage is that regional coverages can be released incrementally.

The adoption of a time-series approach came after completion of the 1820 slice-in-time. The 1820 GIS layers, based on the Tan Atlas, were done for CHGIS Version 1 in order to give users a sense of the scope of the project. However, it is the 1911slice-in-time that is the basis for the time-series. As the data in the 1820 layers is absorbed into the time-series the polygons drawn for 1820 are superceded, as a comparison between the 1820 layer and query of the time-series for 1820 shows.

The time-series approach had an unplanned benefit. Because researchers now had to rely on multiple historical sources and consider each administrative unit on a case-by-case basis, it was possible to collect their citations of primary and secondary sources and notes on their reasoning in making judgments about time and place and include them in the CHGIS database.

However, it is also true that a time-series requires significant research effort. There were tens of thousands of administrative changes. These changes tend to be clustered during dynastic foundings and periods of reform, but often older patterns reemerge after a period of intense reform. We can take Wuzhou as an example

758 to 1275: 婺州 Wu Zhou
1276 to 1358: 婺州路 Wuzhou Lu
1359 to 1359: 宁越府 Ningyue Fu
1360 to 1911: 金华府 Jinhua Fu

During this entire period there was only one known change in the prefectural boundary (in 1471). We can argue that this is more information than needed but in fact recording changes in names is just as important as changes in territory. An infrastructural hGIS must also serves as the authority file for places and locations. From this perspective it ought to include any and all changes.

7. Mono-linguality versus multi-linguality

Who is the hGIS for? A focused, research-driven hGIS or spatial dataset is typically in the language of the researchers and/or in the language of the historical sources on which the dataset is based. But an infrastructural hGIS has a larger and longer-term audience in mind. CHGIS was designed for a world audience and seeks to maximize accessibility. For this reason CHGIS is largely bilingual, using Chinese and English, except that research notes on boundaries and points are left in the original Chinese. Even within this framework it is more than bilingual: Chinese characters are given in both *fantizi* (the authoritative form in CHGIS) and *jiantizi*, there is *pinyin* romanization for Chinese characters, and there are English translations for entries that define the nature of

each record (e.g. 更名, name change). And in fact CHGIS does not limit itself to this. The "alternate names" field for places allows us to use any vernacular (Japanese, Russian, Mongolian, Manchu, Uighur) that is included in the Unicode (UTF-8).

It seems to me that the argument for multi-linguality is strong no matter what researchers perceive to be their audience at the moment. An hGIS is a way of making spatiotemporal data available to the world—we want people who do not read Chinese to be able to do analytic work on Chinese history—and to the future. Eventually I think we shall see the various national historical GIS come together into something approaching a world historical GIS. The research driven datasets created now will, I hope, be created with the future in mind.

8. Website versus datasets

When CHGIS began we envisioned providing users with GIS layers in two standard file formats through a simple website. Thanks to the work of Lex Berman the website has been developed into an online historical gazetteer, allowing users to search various datasets for China, acquire the information about them in the database, see their location on various kinds of maps, and (for CHGIS datasets) see their positions relative to superior and subordinate units and their predecessors and successors. The vast majority of visitors go to the website to use the online gazetteer rather than download datasets.

In contrast to the GBHGIS, the American NHGIS, and German HGIS, CHGIS does not maintain the map server capabilities that would allow users to select data layers, query the data by year, and create a map of their own choosing. Our stated goal has been to encourage scholars to use the CHGIS base to create datasets of their own and to share them with other scholars. In retrospect, it might have been a good idea to plan and budget for a system that would allow users to view data and to share data—to create a "geodatabase" which the CHGIS layers would be part of. We did not plan for this, however, because our goal was to go out of business once the CHGIS database was created, and to distribute datasets free of charge to libraries and users with the expectation that they would be preserved and continue to be made available.

Today I am not sure that any project should ever go out of business, just as we do not expect a university library to close shop. The world of spatial data will continue to grow, the use of GIS and hGIS will continue to expand, and the number of expert users will continue to increase. The website as gazetteer, download site, and map server ought to be seen as a permanent fixture of the scholarly landscape. Accomplishing this requires planning for the long term. It also requires that universities, either through the library or some other institutional mechanism, find ways to keep websites operating after a funded project has come to an end.