Key Issues in Compiling a Digital Gazeteer of China's Historical Religious Sites

Merrick Lex Berman
China Historical GIS
April 2004

Abstract: One way to compile a database of historical religious sites is to begin by digitizing descriptions of those sites found in widely used secondary sources. For example, *Da Qing yitong zhi* provides brief descriptive information about several thousand temples and monasteries. In turn, 2,407 of those temple descriptions were compiled into a single index, *Chugoku no jiin* . Therefore, by compiling each record listed in the 264 page *Chugoku no jiin* index, we can create a database of nearly all the temple descriptions that are scattered throughout several hundred volumes of *Da Qing yitong zhi.* This paper recounts the process of digitizing the *Chugoku no jiin*, including the georeferencing of the temple descriptions to known point locations in the China Historical GIS database, and considerations for how to carry on the next step of research. Thorny issues that still need to be dealt with include: reconciling conflicting information between different gazetteer sources, confirming more precise locations for the historical religious sites, representing locations of sites for which we have only general location descriptions, and extending the database to include important events related to each of the temples.

1.1 Selection of Sources

Naturally the selection of sources to be used for compiling a database is of paramount importance. We know the principle of "garbage in, garbage out," is true, and yet, what if the debate about which sections of our sources actually are "garbage" is far from settled? How can we make a first attempt to present the entire landscape as it has been depicted in various accounts over the course of several thousand years? In the case of Chinese texts, we benefit from centuries of scholarship and commentaries, which nonetheless, still manage to bury us in a deluge of possibilities.

For the purposes of creating a digital gazetteer of Chinese religious sites, the range of sources from the general to the specific is staggering. We have Imperially commissioned gazetteers, which sifted and extracted information from thousands of local sources, as well as the original local gazetteers from which they were compiled. We also have great numbers of modern works on specific localities, deities, temples, and religious practices. On top of these there are also archaeological reports, and local surveys to choose from. Ideally, we would first dedicate a certain number of years in the careful examination of all the sources mentioned to pick the wheat from the chaff. Our preliminary review of the literature indicates that the *Zhongguo fosizhi congkan,* [1] [*Fosizhi*] in 120 volumes would be perhaps the most comprehensive source to work from. The *Fosizhi* reprints, in fascimile, the contents of 172 temple gazetteers, each of which contain extensive descriptions of local temples, shrines, pavilions, ceremonial gates, and other sacred sites, along with numerous block printed illustrations of temple complexes. If anything, the *Fosizhi* is too extensive and detailed a source, and the prospect of redacting information from its 120 volumes is not practical without extensive funding and human resources.

As a realistic alternative, we decided to work from an abridged index, *Chugoku no jiin* ,[2] which is based on one of one of the best known Imperial gazetteers, the *Da Qing yitong zhi,* [*Yitongzhi*].[3] The *Yitongzhi* includes brief descriptions of several thousand important religious sites located across the length and breadth of the Qing empire. These temples have founding dates ranging from the 1st century CE to the late Qing Period. In this way we can begin with a framework that covers a great chunk of China, both in space and time, and which derives from a single source, albeit an abridged index of a secondary source.

1.2 Comparison of Index with Source

Let's briefly examine the benefits and drawbacks of working from the *Chugoku no jiin* compiled by Yajima Genryo [*Yajima Index*]. The primary benefit is that the *Yajima Index* consists of one slim 264 page volume, and the contents are specifically devoted to Buddhist temples and monasteries, with a separate entry for each site. In addition each entry has a temple serial number and site serial number. The use of serial numbers suggested that we could sort them quickly and from the outset we would be able to, at the very least, georeference them to the 252 sites listed in the *Yajima Index* introduction. In addition to the serial numbers, each entry included separate sub-entries for (a) location information, (b) establishment date, and (c) other notes on the history of the temple. Owing to the breakdown of each entry in the *Yajima Index* into clearly defined sub-sections, we knew that the entire index could be digitized directly into a spreadsheet with fields corresponding to each sub-section. This straightforward data entry process avoided the task of having to find, identify, and redact the entries from the classicial Chinese passages found in *Yitongzhi*.

The main drawbacks of the *Yajima Index* is that the original*Yitongzhi* entries were both slightly abridged and partially translated into Japanese. Examination of sample entries showed that the original sentences such as: *zai Jinhua Xian Tongji Qiao pang* were abridged to read: *Jinhua Xian Tongji Qiao.* Such abridgements, however did not seem to interfere with the critical information for identifying locations. Another drawback was that not all entries from *Yitongzhi* were included. The following chart compares the number of entries for particular prefectures in both the *Yitongzhi* and *Yajima* Index:

Prefecture	Yitongzhi	Yajima	% difference
Jinhua Fu	9	6	33
Songjiang Fu	19	12	37
Yunnan Fu	21	16	24

Figure 1: Total number of entries for selected prefectures

When comparing the entries, we found that none of them that ended with*guan* or *gong* in the *Yitongzhi* were included in the *Yajima Index*, and for the most part all of the other terms were included. If we eliminate the *guan* and *gong* entries the comparison can be updated:

Prefecture	Yitongzhi	Yajima	% difference
Jinhua Fu	6	6	0
Songjiang Fu	18	12	33
Yunnan Fu	18	16	11

Figure 2: Total number of entries for selected prefectures (exlcuding *guan* and *gong*)

Obviously, Yajima's selection process has yet to be fully examined or understood. Nevertheless, the expediency of being able to digitize directly from his index was estimated to require less than one semester of work by a part time research assistant. This was feasible. A comparable digitization of the*siguan* sections of the *Yitongzhi*, which take up 40 physical volumes, and more than 20,000 pages in facsimile editions, could not be calculated, owing to the necessity to examine each chapter separately to find the appropriate *siguan* section. Therefore, we decided to go ahead with the preliminary project of digitizing the *Yajima Index* and distribute the results to those wishing to carry the work forward.

1.3 Issues Uncovered During Digitization

Having embarked on the digitization of the*Yajima Index* , several other issues cropped up. First, we discovered that the temple serial numbers were not unique. There were cases in which the

same number was used twice, no doubt an accident in transcription. Similarly the site serial numbers were sometimes given for temples which clearly belonged to a different site, based on the location information notes. Another problem was found in the transcription of classical Chinese characters to some of their Japanese written forms, which caused some difficulties. Finally, some of the Chinese characters appearing in the *Yajima Index* may have been wrongly transcribed in the first place and have yet to be checked against the original entries in *Yitongzhi*.

Once the complete index had been digitized, new unique identification numbers were added for each temple, therefore the duplication of temple serial numbers does not pose much of a problem. As for the site serial numbers, they were no longer necessary once each temple record had been georeferenced to the closest known point location in the China Historical GIS [CHGIS] database, which will be explained in greater detail below. Regarding the transcription from classical Chinese to Japanese *kanji*, and then back to Chinese characters, it is not possible to be completely sure that errors have not been introduced without resorting to a complete cross-examination of the original entries in *Yitongzhi*. Since the time and resources required to do so are beyond the scope of the current work, we will just have to make note of the possibility of transcription errors and leave it at that for the time being. On a more positive note, all of the locations were matched to existing records in CHGIS, and only a handful of errors were discovered in Chinese characters for those 2,407 places.[4]

2.1 Georeferencing of Temple Locations

The primary goal of this digitization project was to create a prelimary digital gazetteer of important Chinese Buddhist temples throughout history. To accomplish this we had to georeference the textual descriptions of temple locations to actual locations on the ground. Using the example mentioned above for Falong Si: we had to assign a location to the temple which we are told is at "Jinhua Xian, Tongji Qiao." The *Yajima Index* also provides a site serial number, which corresponds to Jinhua Fu.

Our reference for locations was the CHGIS Version 2.0 database,[5] which contains some 30,000 historical instances of administrative units. Using the CHGIS Main table it is a simple matter to look up Jinhua Fu and find two instances, one valid from 1360 CE to 1470 CE, and another valid from 1471 CE to the end of the database coverage period 1911 CE. If we check the establishment date information in the *Yajima Index* entry for this temple, Falong Si, we see a note reading "Tang Period." Neither of the above instances overlap the Tang Period. Since the *Yitongzhi* entries quite often used location information current to the time it was edited, we fall back on whichever entry overlaps the year 1820 CE, the year in which the most comprehensive edition of *Yitongzhi* was published (*Jiaqing chongxiu Da Qing yitonzhi*). Therefore we select the second Jinhua Fu entry, Point ID # 32165, valid from 1360 to 1911 CE.

Next we look up all rows having the Point ID# 32165 in the Parent ID field of the CHGIS PartOf table, which shows us all of the administrative units immediately subordinate to Jinhua Fu. Among these rows is an entry for Jinhua Xian Point ID # 40708. Note that if we were to look up all occurences for Jinhua Xian Point ID # 40708 in the Child ID field of the same table, we would find that Jinhua Xian had the following parent records listed:

from 705 to 741 Jinhua Xian was part of: Wu Zhou
from 742 to 757 Jinhua Xian was part of: Dongyang Jun
from 758 to 1275 Jinhua Xian was part of: Wu Zhou
from 1276 to 1358 Jinhua Xian was part of: Wuzhou Lu
from 1359 to 1359 Jinhua Xian was part of: Ningyue Fu
from 1360 to 1911 Jinhua Xian was part of: Jinhua Fu

Here we can see that during the Tang Period, Jinhua Xian was part of several different superior, or parent, jurisdictions including: Wu Zhou, Dongyang Jun, then another instance of Wu Zhou. What this tells us is that the temple, Falong Si, may have been established in Jinhua Xian during the Tang Period, but Jinhua Xian itself was NOT part of Jinhua Fu at that time because Jinhua Fu

did not yet exist. Even so, the entry for this temple is found in the Jinhua Fu section of the *Yitongzhi*. Therefore we are given more confidence in assigning the location of the temple based on CHGIS records valid for the year 1820, which correspond to the sections in *Yitongzhi*.

Now that we have identified the correct location for Falong Si: Jinhua Xian, can we carry the process any farther? The rest of the location information note, "Tongji Qiao," is not an administrative unit. Although the CHGIS database contains over 20,000 records of towns and villages (current to 1820 and 1911), there are none matching the Chinese characters for Tongji Qiao. If there had been, we could also check CHGIS datasets to make sure that the point location of the town fell within Jinhua Fu and in the proximity of Jinhua Xian to be sure that it was the correct place. In this case, we can only be sure of the location for Falong Si down the to nearest County Seat: Jinhua Xian. Therefore we look up the coordinates for Jinhua Xian, Point ID #40708, and assign them to Falong Si. We also list Jinhua Xian as the "parent" of Falong Si, which, for the purposes of the present study, means: the closest known point identified in the CHGIS database that was used to georeference Falong Si.

The methodology just described should make clear that the x, y coordinates given for each temple in this digital gazetteer do not represent the actual location of the temple, but rather the closest known point found in the CHGIS database to that temple.

2.2 Visualizing Mutliple Temples Georeferenced to the Same Point Location

As described in the previous section, each temple record has been georefenced to the closest known point found in the CHGIS database. The resulting list of temples and the coordinates of their "parent" locations can then be imported into GIS. It should be noted that in some cases only one temple has been georeferenced to a particular point location, while in other cases dozens of temples were georeferenced to the same point location. If we open up our points layer in a GIS application, there is no way to distinguish points that represent a single temple, from those that represent a stack of many temples. The situation is shown in the following map:

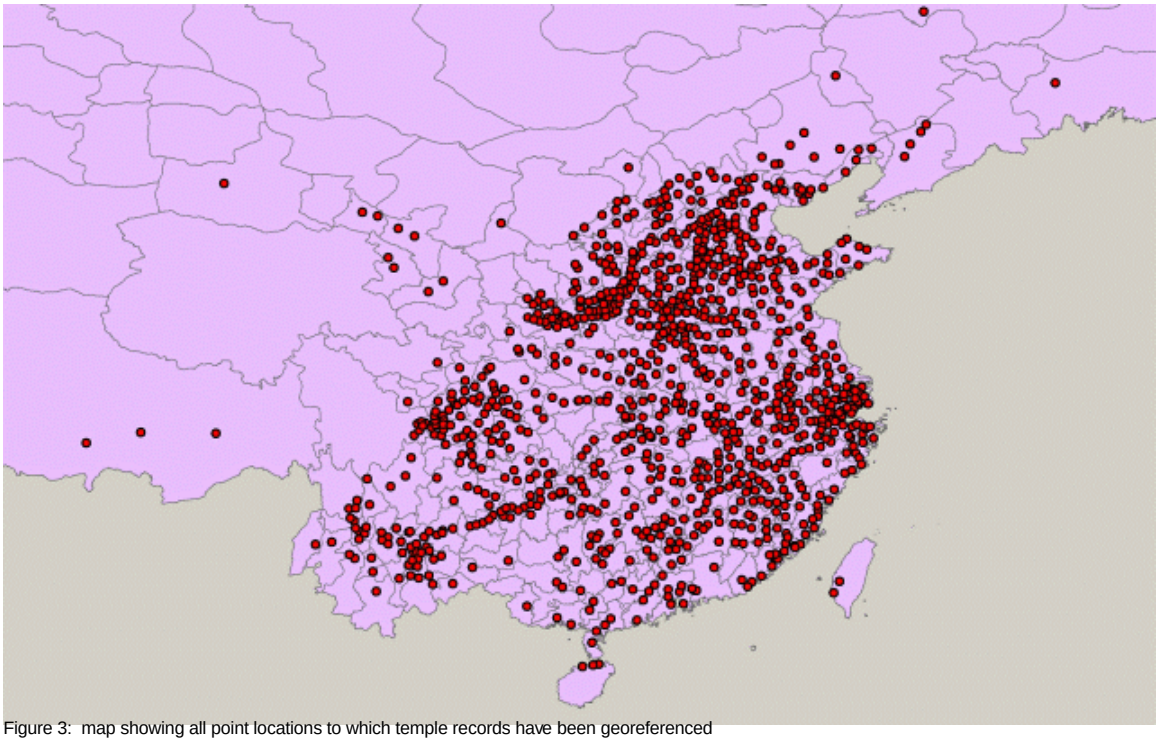

Figure 3: map showing all point locations to which temple records have been georeferenced

The visual impact of the map emphases the density of locations based on their proximity to one another, but does not reveal the density of the number of temples which were associated with each of these locations. The stacking of multiple points (in this case, representing unique temples,) is a phenomenon that also occurs in historical GIS time series. By time series we mean a set of historical spatial objects, each of which represents a single unique instance in time, and which taken as a group represent a series of changes over time. If these spatial objects are points existing together in a single data layer, there will be cases of changes in attributes that required unique instances at different times but which shared exactly the same point location. In these cases, the point objects will stack up on top of one another and appear to be only a single point symbol when displayed, just as our temples did in the map above.

To deal with this issue we decided to calculate an additional value for each record that counts the total number of temples which have the same "parent" or georeference location. In other words, we calculated the total number of temples that had Jinhua Xian as "parent," then added this value into a new field labeled count. This was done using the CHGIS relational database MySQL implementation, by running the a COUNT query on the temples' PartOf Table.[6]

What this does, essentially, is to count the number of unique occurrences of values in the parent_id field and save them into a new delimited text file. The resulting text file was then joined to the GIS layer using the Point ID # of the temple records as the key ID. The result is that for each temple record in the GIS layer we now have the field COUNT. The values in the COUNT field are then used to display the number of temples at each location, using larger point symbols for higher COUNT values.

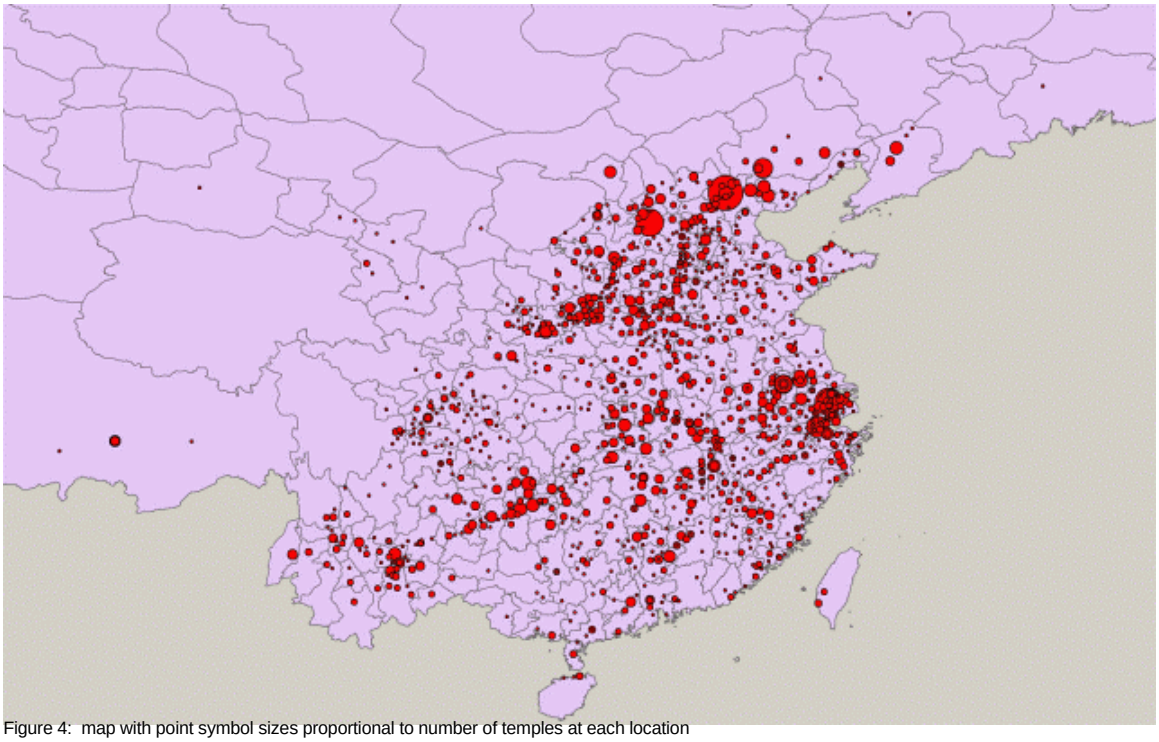

Figure 4: map with point symbol sizes proportional to number of temples at each location

In the second map we can see both proximity of point locations and the density of temples at each location based on the size of the point symbol. The larger the symbol, the more temples have been georeferenced to that point.
Having visualized the contents of the temples gazetteer, it is necessary to point out that the temples recorded here are a reflection of an editorial selection made by Imperial archivists. The inclusion or exclusion of temples in the *Yitongzhi* is clearly a subject that needs to be studied.

The striking lack of temple records for Qinghai, Eastern Tibet, and Central Tibet is noticeable immediately. One suspects that as a larger database of temples is assembled from local gazetteers and other sources, the current list of temples from the *Yitongzhi* will become more interesting because of which temples were left out of the list and which temples were deemed worthy of inclusion.

2.3 Another Model for Showing Spatial Distribution of Temples at the Local Level

When describing temple locations, entries in *Yitongzhi* often mention a direction from a particular place. For example: "Northeast of Such-and-Such County." In local gazetteers we find even greater detail, including both the distance (in Chinese *li*) and directional bearing from a given place. Here we might find: "12 *li* South of So-and-So County."

As part of his extensive research into the local history of Jinhua Fu, Peter Bol has gathered and digitized this information related to more than 600 temples and shrines. Bol has gone on to develop a means of calculating an approximate location of each temple site, based on the directional bearing (ie. South = 180 degrees, or Northeast 45 degrees), and the distance as-the-crow-flies measured in Chinese *li*. This calculated distribution model enables us to show the distribution of temple sites at the specified distances from a given georeferenced "parent" point location. Although the distribution of temple points is only approximate, this model provides another useful means of avoiding the stacking of multiple temple records on top of the "parent" point, and also has the advantage of providing a visual clue as to relative distance and directions of the temple locations as described in the local gazetteers.

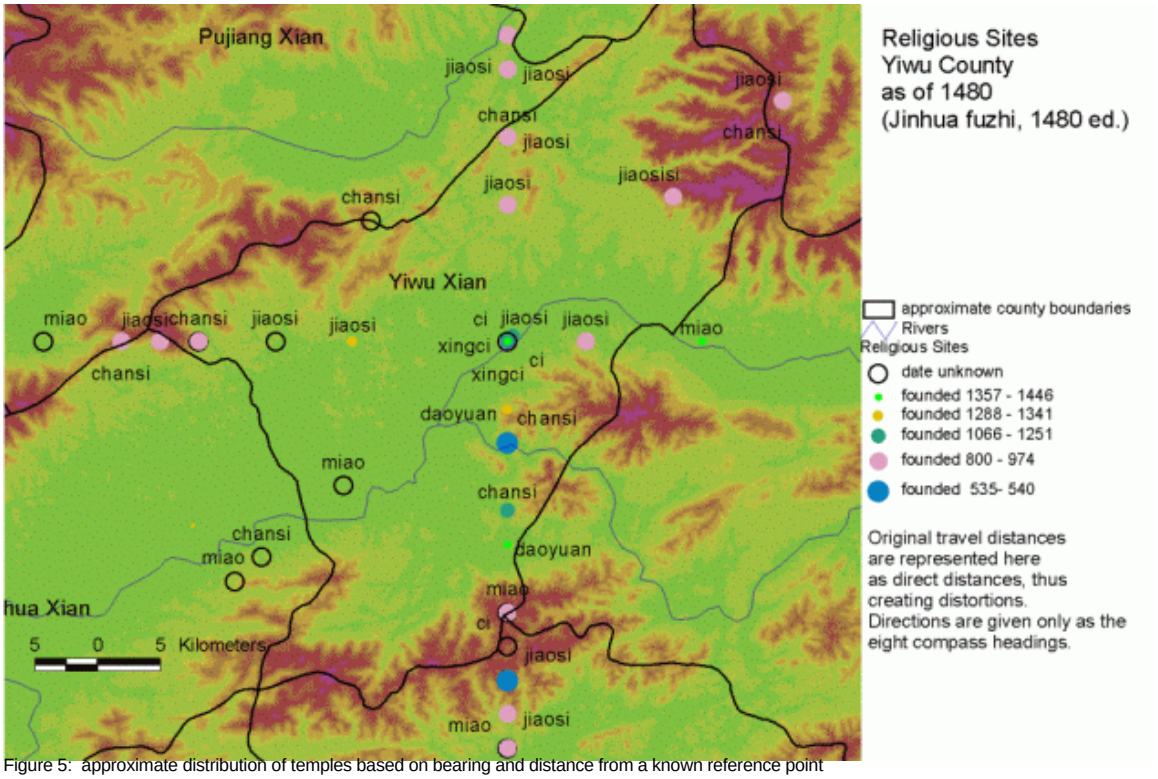

Figure 5: approximate distribution of temples based on bearing and distance from a known reference point

3.1 Dealing with Spatial Uncertainty

In the preceeding discussion, we have been dealing exclusively with georeferencing of places from textual sources. Currently there don't seem to be any standards for evaluating spatial accuracy of point locations derived from such sources, and the problem only becomes murkier as

we move farther back in time. This situation is further aggravated by the degree to which spatial objects in GIS convey the appearance of implicit accuracy. Here I would like to propose some criteria for assigning spatial accuracy to our historically documented point locations.

Let us begin with the most accurate category and work towards the wildest guess category. On the most accurate end of the scale we have ground surveys and parcel mapping, or archeological dig measurements. These data might be accurate in three dimensions to within a centimeter. However, in many cases, they are not synched to any real world projection, and even when they are given tic points, they prove to be difficult or even impossible to integrate with other GIS data that is only accurate to within 100 meters or more. Also on this level of accuracy we have handheld GPS readings, which vary in planimetric accuracy from 2 to 10 meters (or 10 to 40 meters when Selective Availability is engaged). It is safe to assume that there will be a far greater number of researchers who can begin to add GPS coordinates to their field observations, than will be able to do extensive site surveys. Therefore I would propose that we include "GPS coordinates" as one of the spatial accuracy categories.

After GPS readings and other measurements done on site, our point locations will be drawn from cartographic sources including paper maps, GIS data, DEM, and satellite images. Each of these will have their own scale, accuracy and resolution issues, but for sake of simplicity I would propose categories separating the vector data from the raster. For want of a better idea, I've called them "point location" for vector-derived data, and "grid cell value" for raster-derived data.

Places referred to in historical texts are often described in relation to points which we can identify. As in the case of the Buddhist Temple locations (described in section 2.1, above), such places are best directly linked with the known location, which I propose calling "georeferenced location." In this case the base location has a measurable degree of accuracy, but the historical place which has been linked to it does not.

Somewhat less accurate are point locations which have been derived based on bearings, distances, or other descriptive measurements which cannot be evaluated for accuracy, as described in section 2.3. These I would call "approximate location" for want of a better idea. In the same category I would include points derived from historical maps which have been rubber-sheeted to align with modern projected GIS data. A method has been developed to calculate the planimetric error for points taken from historical maps, but owing to both deviations and errors embedded in the original maps, as well as distortions introduced by the digitization and rubber-sheeting process, the resulting point locations remain approximate at best.[7]

We also have cases in which we might assign coordinates to a place based on textual descriptions of terrain, distinct natural features, or merely the fact that such a place is said to have existed between two other places which can be identified. In these cases we can neither georeference, nor approximate the real location, so for these I propose creating a category called "possible location." This category can include places which no longer exist and have not been found, places which are neither settlements nor natural features but are nonetheless described as part of events or incidents of travel, and also temporary encampments and pilgrimage stops.

Finally, on the wild guess end of the scale, I believe we should include a category called "theoretical location." In this category we can include: fictional places that might have been based on real world locations (where was Tao Yuan's Peach Flower Paradise?); or places in which unconfirmable events were said to have taken place (where was the cave where Hui Ke offered his arm to Bodhidharma?); or spatial analogs to philosphical concepts (constellations related to areas on earth?); and even non-geographic spatial constructs (mapping the internal meridians of the human body?).

In the process of assigning locations to places mentioned in classical texts, we are pushing the use of GIS into new directions. It is clearly too early to draft any standards regarding the handling

of our classical information, but as a means of starting the conversation, let us begin with seven categories for spatial accuracy of our derived points:

GPS location
point location
grid cell value
georeferenced location
approximate location
possible location
theoretical location

3.2 Dealing with Temporal Uncertainty

Any digital gazetteer of historical places will have to contain date information. These dates could be assigned a series of discrete dates describing events, such as founding dates, changes of name, restoration dates, closure dates, etc. We must also allow for the assignment of both begin dates and end dates to represent a span of time during which a place existed, or during which a phase of activity occurred.

Numerous factors have to be considered when transcribing dates from classical sources. Short of a lengthy discussion, which is really necessary to unravel all the details, here are some recommendations based on the work done for the CHGIS project.

First, CHGIS formally defined the criteria for creating a record in the database. Specifically these are: placename change, feature type change, spatial / locational change. These criteria are suited to the task of recording the establishment, changes in jurisdiction, and abolition of administrative units. For other subject matter, the criteria might be slightly different. For example, a name change of a temple might be recorded as an alternate name, rather than a unique record. Second, the original calendar date (ie classical lunar calendar, or reign period year) was preserved in the vernacular script in a Source Notes table. Third, the date resolution must be decided upon from the outset. For CHGIS, (a dataset with a temporal span of 2000 years,) we decided that one Common Era year would be the smallest resolution for dating records. This suited the sources for CHGIS, in which years were available in all cases, while months were traceable in a small percentage, and days only very rarely mentioned. Fifth, having set the temporal resolution to one year, we developed a set of date rules to describe the means of obtaining the year. These provide a scale of temporal uncertainty which can be used to compare records to one another and assist the user to interpret the original date information preserved in vernacular source notes. The following date rules have been used:

1. Year is set according to a pan-Dynastic period, such as "Qin Han," or " Song Yuan"
2. Year is set according to a Dynastic period, such as "Tang," or " Ming"
3. Year is set according to a Dynastic Title or Reign Period, such as "Shundi," or "Zhizheng"
4. Year is specified, such as "13th Year of the Kangxi Reign Period"
5. Season or Month is specified, such as "4th month of the Lunar year," or "autumn"
6. Date is specified, such as "jiachen day, 5th month, 14th Year of the Jiaqing Reign Period"
7. Uncertain, year is set by previous or subsequent known dates
9. Year is set according to a dated source (time slice data)

The date rule is to be applied separately to both the begin year and end year values. Let's look at an example, Wu Zhou:

Original Begin date text: "9th Year of the Kaihuang Reign Period, Sui Dynasty" (589 CE)
Begin Year value: 589 CE Begin Year Rule: 4
Original End date text: "1st Year of the Daye Reign Period name changed to Yue Zhou" (605 CE)
End Year value: 604 CE Begin Year Rule: 4

Note that the original text describes a particular year in the reign period, therefore date rule 4 applies. Also note that the year of the subsequent name change (605 CE) is to be used as the Begin Year for a separate entry: Yue Zhou. And the End Year value for Wu Zhou is set to the previous year (604 CE). This is done on purpose so that queries run on the database for particular years will achieve the least ambiguous results. If our temporal resolution had been set to months instead of years, the same process would apply. In other words if our historical text indicated the above change to have occurred in October of 605 CE, we would set the Begin Date of the Yue Zhou entry to October 605 CE and the End Date of the Wu Zhou entry to September of 605 CE. This is done specifically to avoid temporal overlap between records which have historical documentation to show that they were preceeding or subsquent to one another. Temporal overlaps are allowed for unrelated records, or for countervailing claims over the same territory by two political entities.[8]

4.1 Extending the Gazetteer to Include Event Based Records

In addition to setting up a temporal resolution and a formal logic for how dates are set, some thought should be given to extending the gazetteer or database to include events related to each historical place. This can be handled in various ways. The simplest way is to include all the related events or information in a free text description. However, this is not optimal if the database is intended to be used for frequent queries for particular dates or types of events. One way to prepare for such queries is to establish a relational table specifically for related events. In this way, event records can be added as needed and linked back to the original record.

To illustrate how this would work for our temples database, let's take the example of Liucun Si. The establishment date info for Liucun is uncertain Si. In the descriptive notes we find the following info: "Rebuilt during the Chenghua period of the Ming Dynasty. Repaired during the 11th year of the Shunzhi reign period, and repaired again during the 25th year of the Kangxi period." From the descriptive note we can establish that the temple existed some time before the Chenghua period (1465 CE) and was still in existence past the repairs in the 25th year of the Kangxi period (1686 CE). We can then set our Begin Year = 1465 with Rule = 7, and End Year = 1686 with Rule =7. Here Rule 7 means that the actual begin or end date is uncertain, and the date value entered is the earliest or last known instance of Liucun Si. Although it is clear that Liucun Si must have existed long before 1465, at least we are certain that it was in existence at that time. Similarly, even though Liucun Si must have been around long after its second repair in 1686, this is a reliable date, for which we can provide an accurate source citation. If we subsequently found more information in a different source, the begin or end dates are easy to update, and the new source citation can be added to the source notes.

In my view, historical digital gazetteers must establish the best practice of clearly citing specific texts and passages as a basis for each record, and provide quotes and excerpts if at all possible. Otherwise we will just be producing the same kind of unverifiable lists that appear in ordinary printed gazetteers and atlases, and we will be wasting the advantages of the database format.

Now if we were to repackage the information from the free text found in the descriptive note above, the minimal elements to be included in a related events table would include:

Main record ID, Event Date, Event Type, Event Description
Liucun Si ID, 1465, rebuilt, Rebuilt during the Chenghua period of the Ming Dynasty
Liucun Si ID, 1654, repaired, Repaired during the 11th year of the Shunzhi reign period
Liucun Si ID, 1686, repaired, Repaired during the 25th year of the Kangxi period

The main advantage of normalizing the data in this way is to allow sorting and searching by both date and event type. The secondary advantage is to allow the accumulation of additional event information over time without impacting the original records in the database.

4.2 Types of Events Specific to Buddhist Temples

If events related to Buddhist Temples are to be tracked in a database, as suggested above, we should make an attempt to establish a taxonomy of those events. To break the ice on this topic I begin with two general classes of events: structural and organizational. By Structural Events, we mean any event related to the building, damage, alteration, repair, expansion, or demolition of the physical property of the temple. We could start with the following basic categories: built, altered, repaired, rebuilt, expanded, damaged, destroyed, demolished, dismantled, moved.

Organizational events include anything related to the residents of the temple, their organization, reorganization, funding, factionalization, disbanding, ceremonies, dedications, and so on. A preliminary list may include: established, received gift, reorganized, abbot ordained, patron deity dedicated, disbanded. These are just sketchy ideas at the moment, but are worth considering if the objective is to develop a large, extensible database to absorb information over time that is related to each individual historical temple site.

5.1 Integration and Sharing of Data Among Digital Gazetteers

In addition to the present study of Buddhist temples mentioned in *Yitongzhi*, other datasets about or related to Buddhist temples in China have been developed. These include the Buddhist GIS project focused on the Jiangnan region,[9] the Digital Dictionary of Buddhism,[10] and the larger ECAI project which can hopefully serve as a central clearinghouse for the rest.[11] At the present time, the work of integrating these distributed datasets has yet to commence, which means that no discussion of such interoperability issues such as data models, data elements, control vocabularies, versioning, duplicate and ambiguous entries, continuing development or selection of sources has taken place. To get the ball rolling, this paper will offer some concluding thoughts on some of the issues that need to be dealt with for integration of existing gazetteer data and incorporating future developments.

Although the aforementioned projects were developed independently, without specific knowledge of the digitization methodologies used by one another, the general knowledge of creating digital gazetteers has improved so dramatically in the last few years, that we can assume that every contributor has already been thinking of core elements for the content, including: names, feature types, dates, descriptions, and footprints. Therefore, we can at least begin with some confidence that minimal intereporability is acheivable.

The data elements and classification of contents within each database are generally much more difficult to integrate. But at this very early stage of the project, I propose that a thesaurus of feature type classifications specific to religious sites in East Asia be developed. This would have to accommodate a variety of establishments. A preliminary list would have to include: relic pagodas, classic pagodas, monasteries, temples, shrines, pilgrimage sites, ceremonial gates, sacred sites, and so forth. This work can be undertaken by the umbrella project and contributed to by all interested parties. Some mechanism for soliciting input to the thesaurus should be set up and the current contents made available for review and comment during the editing process.

The problems inherent in versioning and continued development are contingent on the status of contributing gazetteers. For the present study, the CHGIS Buddhist Temple dataset can be considered complete in its present form. Therefore, the entire list can be absorbed into any larger draft gazetteer and cross-checked with existing records. For other projects, where incremental entries are made on a continuous basis, it might make more sense to integrate those records on-the-fly as part of a distributed query sent to different servers from a single interface.

The integration of slowly growing datasets by means of such a distributed query clearly adds to the problem of ambiguity between records. Even within one database there is plenty of ambiguity. For example there are nineteen temples called “Longquan Si” [Dragon Spring Temple]

in the CHGIS database alone. These are all unique temples at different locations, but when we attempt to integrate them with "Longquan Si" entries in other gazetteers how can we do so effectively without creating additional confusion?

One suggestion that has been made is to actually merge two records for the same "Longquan Si" into one record. This would require the verification that those two temple records are, in fact, referring to the same establishment at the same place. To accomplish this will require considerable time for cross-checking the sources used to digitize the original gazetteer entries. In the case of CHGIS, these are included in the source notes. However, in other gazetteers, the original vernacular quotes from which the gazetteer entry was derived may not have been digitized, or even more problematic, may not have been documented with exact citation information that would lead us to the original passage in the original text. In short, my argument is that it may be impossible to merge records from different sources with any degree of confidence, unless we reduplicate work that has already been done, and may have taken months or years to complete.

Another option is to absorb all of the records from the different gazetteers into one master database. Naturally, this database must include a field indicating which gazetteer each record came from, so when querying the data it will always be filter by source gazetteer. And to deal with the ambiguity, I would like to suggest creation of a relational table where a any specific record can be given a particular type of association with any other record. In other words, "Longquan Si" in gazetteer A could be defined as <u>may be related to</u> "Longquan Si" in gazetteer B. Or it could be defined as <u>is the same as</u> should there be enough evidence to do so. In my view, the clear advantage of using this method is that incremental changes can be made at any time while preserving the original contents of all the contributing gazetteers intact. Using this method, no original records are removed owing to having been merged with any others, while at the same time the relationship between any two records can be redefined quite easily.

Although the suggestions presented here are just preliminary steps in the process of putting together a larger integrated database of religious sites in East Asia, I hope they have provided some food for thought. Similarly, I hope the issues presented in the previous sections will assist those who are dealing with historical materials in the production of geospatial datasets.

Notes:

1. Bai, Huawen. *Zhongguo fo si zhi cong kan* [Collected Records of Chinese Buddhist Temples]. Yangzhou Shi : Jiangsu guang ling gu ji ke yin she, 1996.

2. Yajima Genryo, *Chugoku no jiin* [Temples of China]. Sendai : Tohoku Daigaku Fuzoku Toshokan, 1966.

3. There are numerous editions of *Da Qing Yitongzhi*, but the *Yajima Index* does not specify the exact edition that was used. Since the *Yajima Index* was published in 1966, and the introduction notes that the draft was compiled 25 years earlier (circa 1941), it is presumed that the *Yitongzhi* edition used must have been one of several major editions (or reprints of them) which were available at that time. These include the *Jiaqing chongxiu Da Qing yitongzhi* (1820), the *Qianlong Da Qing yitongzhi* (1744), or the reprint of *Yitongzhi* found in *Siku Quanshu* (1782). A comparison of the contents of these three sources was done. The *siguan* sections for the sample prefectures: Jinhua Fu, Songjiang Fu, and Yunnan Fu, were all identical. Lacking any other means to verify the original edition used for the *Yajima Index*, we leave the matter open for further inquiry.

4. All of the temple name characters and location information characters were double-checked. In case of any anomalies found in the use of Japanese Kanji variations, these were verified using Jim Breen's Japanese-English Dictionary [http://www.csse.monash.edu.au/~jwb/wwwjdic.html] and replaced with the original Chinese characters. Characters in the other entries (for establishment information and other notes) are more likely to contain any misreadings, if any exist.

5. *China Historical GIS, Version 2.0: Time Series.* Cambridge: Harvard Yenching Inst, 2003. Available for free download here: http://www.fas.harvard.edu/~chgis/data/chgis/downloads/v2/

6. SELECT * COUNT(parent_id) FROM part_of GROUP BY parent_id INTO OUTFILE 'count.txt';

7. For example, a planimetric accuracy calculation was made by comparing scanned and georeferenced Russian maps of Central Asia (dating to the late 19th century) with a DCW basemap, and found a planimetric error of up to 3000 meters. See Merzliakova, Irina and Alexei Karimov. *From Tibet to Altai: Western China in Russian Historical Maps*. Presented at Computer Assisted Archaeology – ECAI Conference (Vienna), 2003: p13.

8. See also a more detailed discussion of temporal relationships: Berman, M.L. *A Data Model for Historical GIS: The CHGIS Time Series*. CHGIS: Oct 2003: p4-7. http://www.fas.harvard.edu/~chgis/data/chgis/downloads/v2/docs/v2_chgis_data_model.pdf

9. Wu, Jiang. BGIS, Searchable Relational Database of Buddhist Monasteries in China. University of Arizona, Tucson. http://bgis.coh.arizona.edu/

10. Muller, A.C. *Digital Dictionary of Buddhism*. http://www.acmuller.net/ddb/

11. *Religious Atlas of China and the Himalayas*. Electronic Cultural Atlas Initiative. (pending)