Markup meets GIS -
Visualizing the 'Biographies of Eminent Buddhist Monks'

Marcus Bingenheimer, Jen-Jou Hung, Simon Wiles
Library and Information Center, Dharma Drum Buddhist College, Taiwan

Abstract

This paper describes one of the projects conducted at the Library and Information Center at Dharma Drum Buddhist College, Taiwan.[1] Tracing the 'Biographies of Eminent Buddhist Monks' is a project with the aim of visualizing information contained in the collections known as Biographies of Eminent Monks (Gaoseng Zhuan 高僧傳) *in new and productive ways. This is achieved by linking texts encoded in TEI markup with authority databases and returning query results as KML.*

1. Aims and sources

The collections of hagio-biographies of eminent Buddhist monks are among the most important and widely used sources for the study of Chinese Buddhism. Comparable to the genre of Latin *vitae* in Christianity, these biographies offer a fascinating glance into the lives of religious professionals in China between c. 60 and 1600 CE. In contrast to similar genres in Europe or India, however, the Chinese hagio-biographies are not, in the main, legendary accounts. Rather, following the biographical *liezhuan* (列傳) tradition of Confucian historiography, they are replete with datable historical facts.

This project uses XML/TEI markup to encode the four largest of these collections (see Table 1, below).[2]

Collection Title	Author	Date	No. of Biographies[3]
Liang Gaoseng Zhuan (梁高僧傳)	Huijiao (慧皎)	519	257
Tang Gaoseng Zhuan (唐高僧傳)	Daoxuan (道宣)	665	482
Song Gaoseng Zhuan (宋高僧傳)	Zanning (贊寧)	988	531
Ming Gaoseng Zhuan (明高僧傳)	Ruxing (如惺)	1600	112

Table 1: The four major *Gaoseng Zhuan* (高僧傳) collections

Together these four collections contain more than 1300 hagio-biographies. They do, however, pose certain difficulties for the reader. In addition to the classical diction, the calender systems are not immediately interpretable, and many of the places and people mentioned in the text are hardly known today. Providing editions of these texts necessitates decoding all names and dates with the help of reference tools and encoding this information in the form of (often extensive) notes.[4] In recent years, however, it has become possible to visualize the texts in a way that is vastly superior to the apparatus of print editions. Prosopographical research, with its emphasis on large amounts of data and research into social networks, stands to benefit especially from these

---

2   There are other important biographical sources like the *Biographies of Eminent Nuns* (*Biqiuni Zhuan* 比丘尼傳 (T. 2063, dated 516)), or the two 18th century collections compiled by Peng Shaosheng 彭紹升 (1740-1796): *Biographies of Buddhist Laymen* (*Jushi Zhuan* 居士傳, CBETA/X. No. 1646) and *Biographies of Buddhist Lay-Women* (*Shan Nüren Zhuan* 善女人傳, CBETA/X. No. 1657).

3   Not including the considerable number of short biographical notices that are sometimes attached to longer *zhuan*, of which the *Liang Gaoseng Zhuan* alone contains more than 200.

4   For the *Gaoseng Zhuan* literature this task has never been attempted. Partial translations exist of approximately five percent of all biographies. For an example of a translation of a biographical account not included in the major collections that identifies all dates and gives coordinates for place names see Takakusu (1925) [2].

new techniques and approaches.

The aims of this project are:

1. To provide researchers with tools that can answer questions which cannot be asked of traditional reference tools, and to visualize these answers in a GIS-like manner. These might be query parameters such as: Show on a map who was active in Zhejiang province during the Tang dynasty (618-907), according to collection Y; or: Show all places the monk Xuanzang passed through on his way to India, according to collection Z.
2. To create an interface where researchers and the interested public can read the biographies and have immediate access to information about persons, places and dates. This is achieved by embedding Google Earth to create interactive maps and connecting person and place names to an authority database.
3. To provide an open-licensed corpus, markup from which can be used and modified by other researchers to interrogate the texts in new ways.

2. Markup and authority files

Markup is one of the core competencies at the DDBC Library and Information Center. Growing out of existing expertise in this area and a commitment to produce ever better digital editions of canonical texts, we have become specialists in assembling data in the form of marked-up corpora.

The main means by which information is encoded in the sources is through markup conforming to the TEI (Text Encoding Initiative) standard, a sophisticated XML standard which "specifies ways to encode machine-readable texts" (http://www.tei-c.org). TEI was chosen because with more than 500 elements it is the most comprehensive markup standard for textual data in the field of Humanities Computing. Further, it is well established and maintained, and guarantees a degree of long-term usability.[5]

The markup is realized in two layers. In the source material person and place names are marked up, as well as dates. In parallel with this, extensive authority databases for persons, places and dates are maintained, which provide unique identifiers for each distinct entity (http://authority.ddbc.edu.tw). This is necessary to disambiguate the plethora of homonyms and aliases of

names, as well as Chinese calendar dates expressed in reign periods and/or sexagenary terminology. This linking of the markup to database records allows for dynamic lookup of the information in the interface, as is discussed in the following section, and for the pooling of information between projects. The result of this first step is structured texts where all person, place and date references are identified and linked to database records by means of *key* or *xml:id* attributes (see Figure 1, below).

Figure 1: TEI markup used in the *Tracing the 'Biographies of Eminent Monks'* project

In a second step, associated information in each biography is grouped together. Here the markup moves to a different conceptual level and incorporates into the source text an understanding of the content which goes beyond the mere identification of names and dates. Whenever the text says that a person was at a certain place at a certain time a *nexus point* is created.[6] To facilitate this, the <linkGrp> mechanism of TEI is used, which collects names and dates from the main body of the text into simple events of the form: *X was at place Y at time Z* (see Figure 2, below).

```
<linkGrp>
    <ptr target="#A000883" type="person"/>
    <ptr target="#A002003" type="person"/>
    <ptr target="#A002004" type="person"/>
    <ptr target="#IN0000000E01AA" type="place"/>
    <ptr target="#d52560365257866" type="time"/>
</linkGrp>
```

Figure 2: Markup structure for nexus points

---

5 As part of our commitment to the TEI standard, we are responsible for the Chinese localization of the TEI standard. See Bingenheimer (2009) [1].

6 Of course this is still a very basic interpretative procedure and in theory more complex events could be described. This, however, would necessitate a taxonomy of events, which is not practicable within the personnel and time constraints of the project.

For GIS-like visualization, this kind of markup can be leveraged to create two layers of facts. Firstly, the *names and dates mentioned*; secondly, the *nexus points* asserting that one or more persons were at a certain place at a certain time.

Other forms of representation (tree-node etc.) might involve the visualization of social networks by analysis of those nexus points which include more than one person. The significance of visualization techniques such as sociograms in the successful analysis of social networks is well observed (cf. Wasserman and Faust, 1994 [4]), and certain techniques in this field may be especially applicable for use with these sources. In the absence of relational proximity data and other attributes, the assignment of gravity is problematic, but with the available datasets it may be possible to conduct fertile social network analysis based on degree, closeness and betweenness centralities, and to visualize the results of this research using social network analysis software.[7] Diachronic visualization techniques and the possibilities they introduce may also be interesting to explore.

3. Interface design for GIS-like visualization

In creating a visualization of a body of prosopographical literature, one obvious approach is to represent the geographical and temporal information about people in a GIS-like way. 'GIS-like' because there are considerable differences between this solution and a full-fledged GIS. Considerable efforts have been made to allow for diachronic visualization of historical information using GIS technology, and 'History GIS' has become an interesting sub-field for both GIS research and data-based approaches to the study of history.[8] The use of GIS, however, involves a substantial investment in sophisticated software[9] and expensive specialists. With the advent of GoogleEarth and GoogleMaps the display of geographic information, including a basic capacity for diachronic views, has come within the reach of non-specialists. This 'poor man's GIS' is based on data-transmission via Keyhole Markup Language (KML), an XML standard that typically encodes database output and can be read by GoogleEarth servers. Though it cannot compete with the analytical and modeling capabilities of full-fledged GIS platforms, this 'GIS-lite' approach gives digital content providers a method for visualizing geo-spacial information. In the case of this project, this makes it possible to view the biographies of eminent monks in new ways without resorting to expensive software. Moreover, KML, and its archive format KMZ, are in many ways open standards, and the output generated by the *Tracing the 'Biographies of Eminent Monks'* interface can be read by a number of other applications across diverse computing platforms.[10]

The system rests on three main components: an XML database, using eXist (http://www.exist-db.org/), which stores and retrieves the TEI source files; the MySQL database with the authority information on names and dates; and the interface, which integrates the data with the GoogleEarth plug-in in the browser and controls the KML output, and is built using the EXT JavaScript library (http://extjs.com/). EXT was chosen because it is one of the more advanced JS libraries available. It provides many vital interface functions 'out-of-the-box' and is used in for a number of other projects at the Library and Information Center. Figure 3, below, shows the organizational structure of the system.
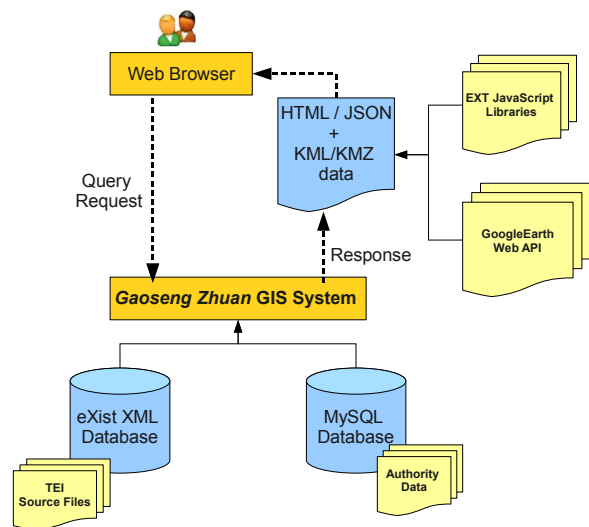

Figure 3: The design of the web interface for the *Tracing the 'Biographies of Eminent Monks'* project

---

7   A number of open source social network analysis packages are available at the present time, including igraph (http://cneurocvs.rmki.kfki.hu/igraph/) and SocNetV (http://socnetv.sourceforge.net/), which we are currently engaged in investigating.

8   For an up-to date bibliography on the subject see Gregory and Ell (2008) [3].

9   Although there are open source solutions like TimeMap (http://www.timemap.net/) and TimeLine (http://code.google.com/p/simile-widgets/wiki/Timeline) these usually provide only some of the functionality employed by GIS specialists.

10 KML was created in 2001. In addition to Google applications, KML is used by NASA WorldWind (since version 1.3.4) (http://worldwind.arc.nasa.gov/), ESRI ArcGIS (http://www.esri.com/) and others. KMZ provides the capability to include additional layers or icons with the data.

4. Visualization of textual content

As has been the case with all previous projects at the DDBC Library and Information Center, the data has been made available at a fairly early stage. As of April 2009 the interface, though still under development, is already functional for the complete *Liang* and *Tang Gaoseng Zhuan* collections. The other two collections will be made available between now and March 2010. At present the service offers two views:

A) An interface to read the biographies with on-click information on names and dates, which is retrieved dynamically from the Authority Databases (Figure 4).
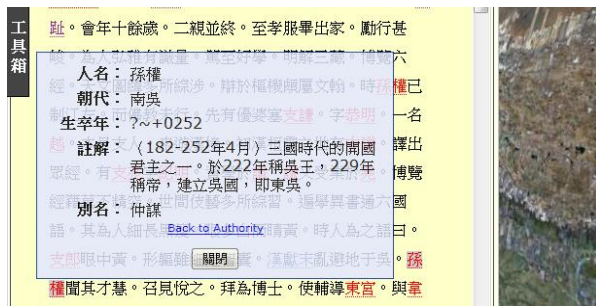

Figure 4: Pop-up information from the Authority Databases

Simultaneously, places and nexus points (a person or people, with a place and date) which appear in the text are realized in the GoogleEarth browser plug-in and are visualized on a map (see Figure 5, below).

Note that the link at the bottom of the tool tip references the information to a canonical source. This makes it possible for the user to verify the information for all nexus points. Experiments with overlaying this generated geographic information on historical maps of China are currently ongoing. These maps can be embedded in the KMZ zip format, and so the output can be self-contained, but they are difficult to obtain and license, and to calibrate and align correctly.

B) More complex queries can be made from the search dialog interface. It is here that the potential to extract information from the texts is significantly expanded, as researchers can download KML files that visualize answers to questions which could not even be asked of traditional reference tools, such as indices or comparative calendars. Users select one or more persons and places and specify a time period, and receive answers to their queries in form of a
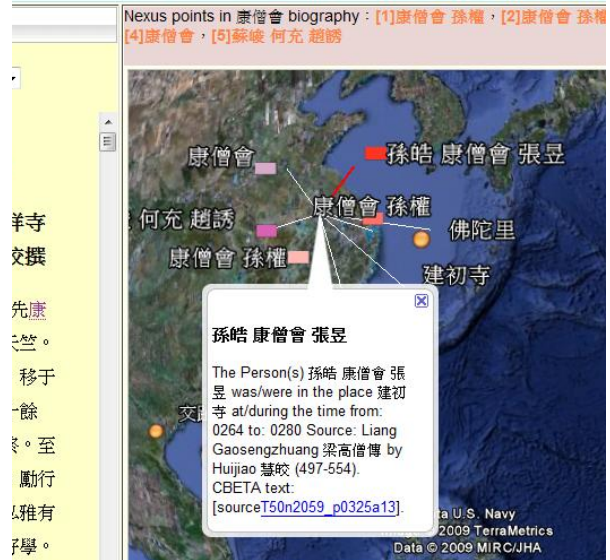

Figure 5: Visualization of a nexus point from the *Liang Gaoseng Zhuan*

KML file. Figure 6, below, for instance, shows the nexus points which represent Buddhist activity in Shandong Province from 400 to 500 CE according to the *Liang Gaoseng Zhuan*.
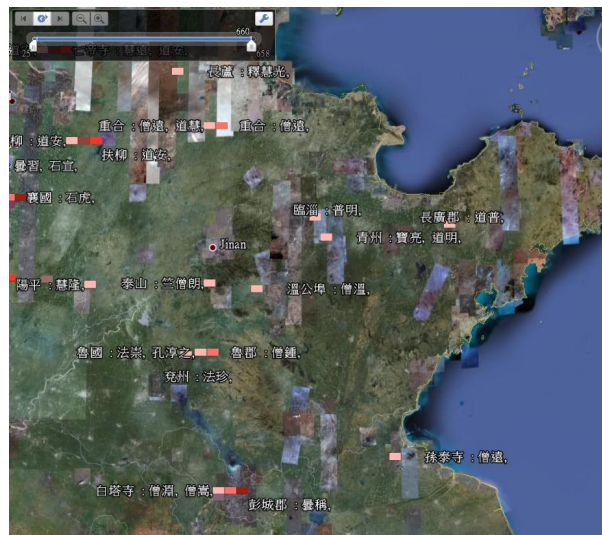

Figure 6: Buddhist activity in Shandong Province in the 5th century from the *Liang Gaoseng Zhuan*

Another example is a 3D visualization of the birthplaces of all monks in the *Liang Gaoseng Zhuan*. It allows researchers to identify the areas where most monks came from, which correspond roughly to the main centers of Buddhist activity. Though scholarly attention usually is focused on the section of the *Gaoseng Zhuan* concerning the translators, who are associated mainly with the south
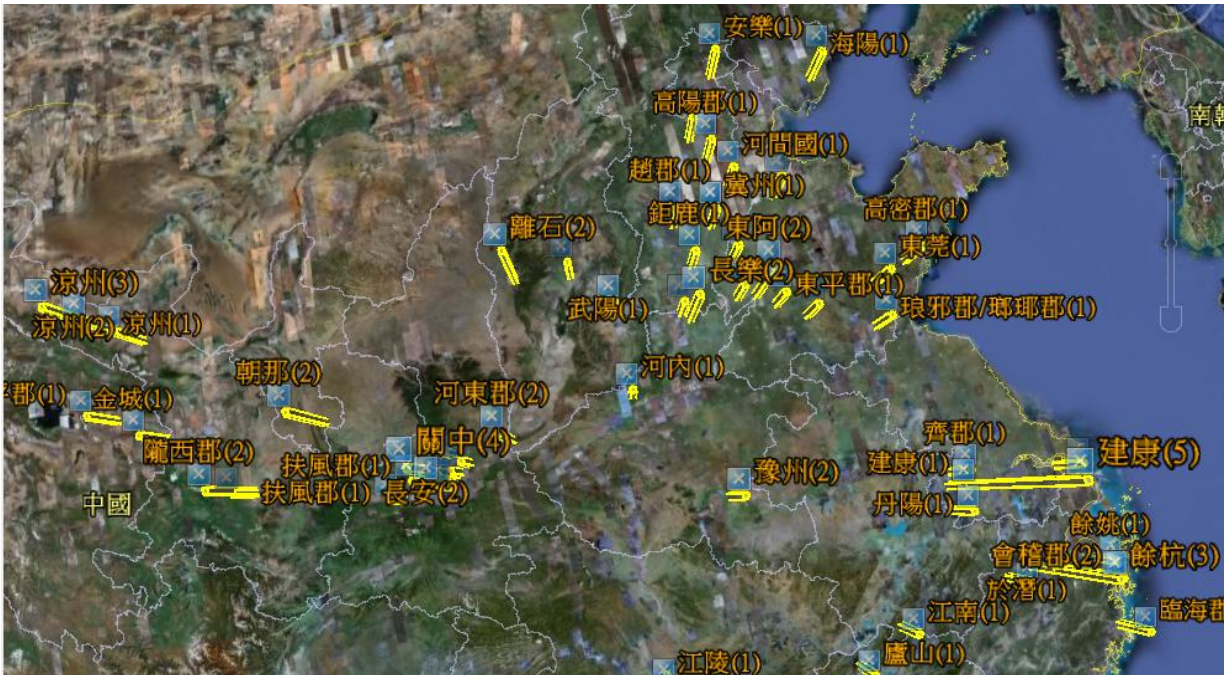
Figure 7: Visual representation of the birthplaces of monks from 60 to 500CE, as recorded in the *Liang Gaoseng Zhuan*. The heights of the bars represent the number of monks born in a given location.

of China, Figure 7, above, reminds us that the majority of monks born between 60 and 500 CE came from the north of the country.

As with most the Center's projects, source text is offered in raw TEI for free download. Care is also taken to make sure that KML files are available for use in the GoogleEarth client and in other applications, i.e. independently of the interface. An interface based on the GoogleEarth plug-in can not be considered a viable long-term delivery method, since long-term support is not guaranteed for the plug-in, and its technical capabilities stay one development cycle behind the GoogleEarth stand-alone client. It is also worth noting that the GoogleEarth plug-in is currently only available to users of Windows and Mac OS X from version 10.4 onwards.

5. Future development

From the perspective of visualization there are two areas into which we plan to make at least experimental forays with the present dataset. One is the spatial representation of social networks by visualizing the information contained in the nexus points. Individuals who the texts tell us were in the same place at a given time, i.e. are connected in a nexus point, are extremely likely to have known one another, and other person/location/time overlaps might be unearthed by querying the data in more involved and imaginative ways. These data can be visualized and represented as social networks and would be immediately useful to researchers trying to understand the rudimentary networks that existed within the Buddhist community and how they evolved over time.

A second field worthy of exploration is the use of historical maps in a KMZ file. This would allow users to view the data not only in the context of modern Chinese geography but also against the backdrop, quite literally, of the political and social geography of historical dynasties.

References

[1]   Bingenheimer, M. (a.k.a. Ma Dewei 馬德偉), ed. 2009 : *TEI shiyong zhinan - yunyong TEI chuli zhongwen wenxian* TEI 使用指南—運用 TEI 處理中文文獻 [Chinese TEI – A guide to using TEI with Chinese texts]. Taipei: Taiwan E-learning and Digital Archive Program. ISBN:978-986-01-8092-3.

[2]   Gregory, I. and Ell, P. (2008) *Historical GIS: Technologies, Methodologies, and Scholarship*, Cambridge: Cambridge University Press.

[3]   Takakusu Junjirō (1925) *Kanshin's (Chien-Chen's) Voyage to the East, A.D. 742-54*, London: Probsthain.

[4]   Wasserman, S. and Faust, K. (1994) *Social Network Analysis: Methods and Applications*, Cambridge: Cambridge University Press.