

# A Multilingual Gazetteer System for Integrating Spatial and Cultural Resources

## Summary of Activities and Findings

Principal Investigator: Lewis Lancaster,  
Director, Electronic Cultural Atlas Initiative, University of California, Berkeley.  
Project funded by NSF IM/ITR grant 0114019, September 2001 through August 2002.

### **Research and Education Activities\**

The major activities of the project were to design, develop, and test standards for entries in gazetteers, standards for designating geographical feature types, and standards for characterizing gazetteers themselves-at the level of complexity needed to provide effective support for computing in general, and GIS in particular, in history and the humanities.

### **Findings:**

A. Scholars in history and the humanities developing digital gazetteers have unique problems. These include frequent temporal change (in all of the characteristics of places-names, spatial footprints, and feature types); place names that are not only unstable, but are also multiple and often contested, representing different regimes, languages and cultures with interest in particular places, and information about places that is derived from texts and that is often incomplete, uncertain, or different sources which contradict one another. Oftentimes, a single place is contested between multiple political entities. These considerations need to be supported in gazetteer content standards and implementations.

B. Feature types for history and culture share the characteristics of all spatial information for history and culture. That is, feature types are unstable and frequently change over time, and they are often hard to attribute to places with certainty. In addition, many kinds of geographical features are inadequately described in existing standards, which have been derived from topographical maps, and maps of political and communications infrastructure.

C. Existing content standards can be adapted and expanded to meet the needs of cultural heritage managers, historians, and humanities GIS developers. However, even when standards are theoretically useable, there are no existing examples or guides to good practice that address the problems of structuring complex entries with multiple names, frequent temporal change, or highly uncertain information. Such examples and references must be developed and made available to developers in historical and cultural computing, who are eager to have access to resources of this type.

D. Historians need gazetteers that support place names in multiple languages that use a variety of scripts and have many co-existing transliteration systems and regional pronunciations. For instance, Chinese gazetteers may require support for characters; for pinyin, Wade-Giles and post office transliterations as well as variants, and for Mandarin, Cantonese, Taiwanese and other pronunciations. Korean gazetteers

require support for Hangul, Chinese, and Roman characters, and for several transliteration systems. Existing content and format standards for gazetteers need to be expanded to support multiple languages and multiple scripts.

E. Specialists require specialist feature type thesauri (to describe types of Buddhist temples, units in the medieval Chinese administrative system, or features of British canal archaeology). At the same time, to create the capacity to search multiple gazetteers by feature type, there must be some mechanism for mapping between specialist thesauri. No existing feature type thesaurus meets this need. A restricted generic feature type thesaurus can be used as a pivot language to reference specialist thesauri, and should be adopted as a basic common feature type thesaurus in an environment of networked gazetteers.

### **Training and Development:**

Since this project was devoted in large part to researching and developing standards, it was considered important to have regular workshops and training meetings. These provided opportunities for members of the historical and humanities GIS community to meet, evaluate existing standards, discuss their own needs and gazetteer development work, learn how to build better gazetteers, and to offer feedback about new standards and thesauri as they were drafted. The architecture for a gazetteer clearinghouse was also discussed at these workshops. Two meetings were held at Academia Sinica, Taiwan, hosted by the Academia Sinica Computing Center, where a multilingual gazetteer for Chinese history and culture is under development. The first meeting was held at the beginning of the grant year in August 2001, and the second was held at the end of the year in May 2002. In addition, the standards development team at UC Berkeley hosted a visit by Linda Hill, director of the gazetteer project at the Alexandria Digital Library, for an intensive walk-through of the gazetteer content standard that has been developed there. On the basis of this meeting, a number of changes have been made to the ADL standard in order to allow it to better meet the needs of gazetteer developers working on historical and cultural projects.

### **Outreach Activities:**

In addition to workshops dedicated to the gazetteer standards development project, there have been regular opportunities to present papers or lead discussions at other conferences in digital humanities and cultural heritage, historical GIS, and digital library conferences and workshops. Gazetteer standards and implementation working sessions have been held at each Electronic Cultural Atlas Initiative (ECAI) conference in the last two years: Sydney, Australia; Guadalajara, Mexico; Seoul, Korea, and Osaka, Japan. Other presentations and workshops have included: The UC Berkeley School of Information Management and Systems Seminar on Information Access; the Historical GIS workshop at Fudan University, Shanghai; the Cultural Heritage section of the Virtual Systems and Multimedia conference at UC Berkeley, and the Gazetteer Workshop sponsored by Networked Knowledge Organization Systems/Services at the Joint Conference on Digital Libraries in Portland.

In addition to presentations and workshops, drafts and informal reports have been regularly posted to the CGGR-L listserv, the major email discussion list for gazetteer developers and

implementers.

### **Websites**

<http://www.mip.berkeley.edu/ecai/gazetteer/>

<http://anther.mip.berkeley.edu/gazetteer/>

### **Description:**

<http://www.mip.berkeley.edu/ecai/gazetteer/> Draft Feature Type Thesauri.

<http://anther.mip.berkeley.edu/gazetteer/> Prototype multilingual gazetteer clearinghouse testbed

### **Other Specific Products**

Draft thesaurus of geographical feature types for gazetteer developers concerned with history and cultures.

### **Contributions**

This work provides the basis for design of cross-lingual searching of Chinese and English gazetteers. It improves the design for online gazetteer content, format and services. This aspect of the work is continuing under a National Library Leadership grant from the Institute for Museum and Library Services entitled *Going Places in the Catalog: Improved Geographical Access*.

This work provides an improved basis for structuring, documenting, describing and combining complex information about places with the characteristics found in most historical and humanities research. Places described by researchers in this domain tend to change over time, have many names, and are described in texts rather than through precise measurement methodologies such as GPS. This project has also produced a generic feature type thesaurus that describes all kinds of places including those with cultural and historical significance. It has developed an inventory of major gazetteer development projects for history and culture worldwide.

The work has included planning for incorporating a distributed gazetteer service into a time enabled spatial data clearinghouse for history and culture. Strategies for dealing with missing, ambiguous or inaccurate spatial information and spatial information derived from texts have been explored.

### **Contributions:**

There have been four significant influences on information resources:

- A. Development of a feature type thesaurus will facilitate feature-based searching of multiple networked gazetteers.
- B. Development of a revised gazetteer content standard that is optimized for the needs of developers of historical and cultural gazetteers
- C. Development of a multilingual, multi-script testbed for searching multiple gazetteers in different formats. This is a first step toward development of a multilingual, multi-script clearinghouse for gazetteers.

This project was inspired in large part by a commitment to assisting with cultural heritage management for the digital age. Improved gazetteer standards and an improved feature type thesaurus will assist cultural heritage managers such as UNESCO in describing threatened and/or historically significant places and sharing information about those places in a standardized manner on the internet. Steps toward creation of a gazetteer clearinghouse will enable cultural heritage managers to integrate information about sites with other contextual spatial information. The goal is to provide a textual-geospatial information service that allows multiple gazetteers to be searched at one time, and allows gazetteers to be used to georeference other toponym-rich digital resources in the clearinghouse. The ECAI Metadata Clearinghouse ([www.timemap.net/clearinghouse/html/index.html](http://www.timemap.net/clearinghouse/html/index.html)) is the basis for such an architecture.

**Acknowledgments:**

The following worked on this project: Michael Buckland, Ray Larson, Ruth Mostern, Ben Brose, Susan Stone, Jeanette Zerneke,

We are grateful for the advice and assistance of Merrick Berman, China Historical GIS project, Harvard University; Arthur Chen, Academia Sinica, Taiwan; Sophy Chen, Academia Sinica, Taiwan; Lawrence Crissman, ACASIAN project, Griffith University, Australia; Maggie Exon, School of Information and Library Studies, Curtin University, Perth, Australia; I-Chun Fan, Academia Sinica, Taiwan; Karen Frederickson, Polis center, Indiana University- Purdue University, Indianapolis; Linda Hill, Alexandria Digital Library, University of California, Santa Barbara, CA; Ian Johnson, Archaeological Computing Laboratory, University of Sydney, Australia; David Smith, Perseus Project, Tufts University; Humphrey Southall, Great Britain Historical GIS Project, Portsmouth University, U.K; Eric Yen, Academia Sinica, Taiwan.