
Treating a Genre as a Database: the Chinese Local Gazetteers, the LG Tools, and Research Based on This New Digital Methodology

Shih-Pei Chen

schen@mpiwg-berlin.mpg.de

Max Planck Institute for the History of Science, Germany

Qun Che

qche@mpiwg-berlin.mpg.de

Max Planck Institute for the History of Science, Germany

Ling Cao

caoling1013@163.com

Nanjing University of Information Science & Technology
China

Dagmar Schäfer

dschaefer@mpiwg-berlin.mpg.de

Max Planck Institute for the History of Science, Germany

Hongsu Wang

hongsuwang@fas.harvard.edu

Harvard University, United States of America

This panel discusses how digitization and digital tools help to bring new insights to a well studied genre, in this case the Chinese Local Gazetteers, by supporting research inquiries that treat the whole genre as a conceptual "database" to answer especially large scale questions that take into account gazetteers from multiple geographic regions and within long time spans.

The Chinese Local Gazetteers is a long established genre of writing in China since the twelve century for recording local knowledge about a region. Local gentry and officials compiled information about a region, ranging from landscape, flora and fauna, officials and celebrities to temples and schools, local culture and customs, and taxes and census, and kept them in this genre. Despite that Local Gazetteers have been major sources for scholars to find specific information about a place, it turns out to be very difficult if a scholar wishes to study the gazetteers on larger scales due to the vast amount of information contained within.

Thanks to the increasing recognition of digitizing historical sources, as of 2016 nearly half of the extant 8,000 titles of local gazetteers (one title can mean one to dozens of physical volumes) has been digitized as searchable full texts and provided access through various databases from public and private sectors. However, despite of the large amount of local gazetteers available electronically, these databases mainly provide features that replicate how physical books are used: users are able to flip a gazetteer page by page; even when reading the returns from a full text search, a scholar still needs to click the returns one by one in order to read the texts. Such principle of treating the gazetteers individually highly restricts the possibility of researching large sets of digital gazetteers. Historians need better tools to work with the large amount of digitized gazetteers to employ new forms of digital research methodologies and to fully benefit from the digital formats. One important methodology that we address here is the possibility to treat the whole genre as the body of inquiry in order to research global and large scale phenomena that are across individual gazetteers, geographical regions, and time spans.

Under this context, Department III "Artefacts, Actions, and Knowledge" of the Max Planck Institute for the History of Science (MPIWG) embarks on a digital project to build digital tools specifically for this genre that allow scholars to adopt this methodology of treating the available set of digitized gazetteers as a conceptual "database" to post inquiries. The digital tools, which we call **the LG Tools**, include a full text search facility, an Extraction Interface to collect data in the form of lists, a research repository to store and publish collected data, and an interactive mapping and analysis platform that are linked to data collected via full text search or the Extraction Interface in order to visualize the data geographically. We have presented the LG Tools last year in this conference. The abstract, which contains detailed description for the tools, can be found [online](#).

In this panel, we would like to shift the focus from the tools themselves to the types of research that can be derived from the LG Tools and from the methodology of treating the whole genre as a database. We invite four scholars in Chinese history to demonstrate their research projects and to discuss what advancement the LG Tools have brought to their research. The topics of their research range from history of science and technology, environmental history, to social and intellectual histories. The four historians and their research projects are described below.

Qun Che

Qun Che uses the LG Tools to trace the transformation of the geographic landform of the Dongting Lake, a flood basin of the Yangtze River, by looking at the construction of “Yuan” recorded in the local gazetteers during the Ming and Qing dynasties (1368-1911). Yuan is a type of water conservancy that literally refers to fields surrounded by embankments to prevent river flooding. Che first finds all the occurrences of Yuan constructions in the available digital set of local gazetteers by full text search. The resulting dataset is then sent to LGMap, the interactive mapping and analysis interface among LG Tools. By using LGMap, Che is able to clarify the Yuan constructions in the Dongting Lake region in different time periods and by doing so to trace the flooding and deposition process of the lake to understand how the lake has been shaped from the past to today. Figure 1 shows the LGMap interface for analyzing the periods and locations of the Yuan constructions near the lake.

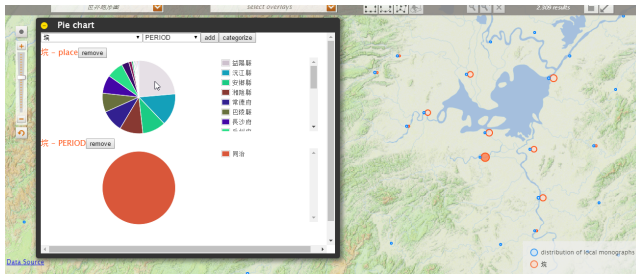


Figure 1. Yuan constructions recorded in the available digital set of Local Gazetteers, visualized in LGMap

Ling Cao

Ling Cao collects records in the local gazetteers describing the growing and spreading of maize in China during the Ming and Qing dynasties in order to understand through which geographical routes maize was introduced to and spread in China. She further relates the regions of maize growing and their scales to the occurrence rate of natural disasters of the typical maize regions to see the influence of maize over the ecological environment. From the records she collects, Cao is able to raise the hypothesis that the introduction and growing of maize in the mountain areas might cause the destruction of forests and result in serious soil erosion which in turn led to blockage in downstream rivers and floods.

Dagmar Schäfer

Dagmar Schäfer researches on the relationship between material availability and the development of local knowledge organization in the local gazetteers.

While not all local gazetteers share the same classification schemes or even topic, local officials recorded local expertise and material specialties in the gazetteers. In which sections did local gazetteer compilers place such information? When and how did their descriptions vary (such as details, terminology, etc.)? The full text function of the LG Tools and the section titles that MPIWG invested to type up give hints on the above questions. The contribution also aims at discussing the possibility of tracing and analyzing networks of materials: how did officials relate materials to another; can scholars identify clusters among the materials. The contribution also posts questions on how social network analysis can be applied and what are the advantages of geographical mapping.

Hongsu Wang

Hongsu Wang will demonstrate how the LG Tools especially the Extraction Interface helps the China Biographical Database project (CBDB) to collect data about local officials from local gazetteers. CBDB is a freely accessible relational database with biographical information about more than 360,000 individuals in historical China. In addition to the large dataset, CBDB embeds functions to analyze the individuals via statistical, social network, and spatial analysis. CBDB turned to local gazetteers when it tried to increase its collection on individuals that were mainly based on central government records. The individuals listed in the sections of “local officials” in most gazetteers are ideal for this purpose since most of them don’t appear in imperial government records. By using the Extraction Interface, which provides a regular expressions editor to help capture regularly written information such as lists, student assistants at CBDB were able to collect 250,000 officials from 291 local gazetteers using just 420 man-hours.

Organization of the panel

This panel will include five speakers. To give the audience a background for discussion, Shih-Pei Chen (MPIWG) will begin by giving an introduction to the genre of Chinese Local Gazetteers and its characteristics as well as the LG Tools (15 minutes). Then, the other four panelists will each present their research projects and show how they get new findings through using the LG Tools (each speaker 10-12 minutes). We will then have a discussion session (roughly 30 minutes) for the panelists to talk about how the existence of the LG Tools has changed the objects of study and/or the approach to these objects and to discuss

the conceptual frameworks and assumptions challenged by the use of digitized sources. We will also address the point of how the LG Tools help the panelists to develop the idea of a regarding the genre as a “conceptual database”. The audience is invited to join in the discussion session for conversations on the types of advancement these tools bring to the study of history and humanities, whether similar tools can be set up for digital resources in other humanities disciplines, and the copyright issues encountered for building such tools.

All the panelists listed above have agreed to attend the conference shall the proposed panel be accepted.