



Screenshot of the Local Gazetteers project homepage on the website of the Max Planck Institute for the History of Science, 2016.

Material network analysis: An exemplary project on Chinese local gazetteers

Wilko von Hardenberg interviews Shih-Pei Chen

The Max Planck Institute for the History of Science (MPIWG) is very active in the promotion of projects that unlock the power of digital tools to innovate the ways in which the history of knowledge production is researched and understood. One project in particular, dedicated to Chinese local gazetteers, also appears to have particular potential to provide new ways to look at environmental issues in a historical perspective.

Shih-Pei Chen, digital content curator in the Institute's Department 3 Artefacts, Action, and Knowledge, directed by Dagmar Schäfer, has kindly agreed to discuss the project and how it is positioned in the broader field of digital humanities with *Ant Spider Bee* co-editor Wilko Graf von Hardenberg.

ASB: What is the overall aim of the Local Gazetteers project?

SPC: The project's aim is to turn an important genre of historical Chinese texts, the so-called local gazetteers (difangzhi 地方志), into new formats that will allow researchers to ask questions in new ways. These new ways sometimes concern the scale, sometimes the scope of research. We are particularly interested in "material network analysis," which we understand in relation to "social network analysis" as a way to follow the historical "ties" drawn between materials. Such ties sometimes were linguistically defined. They can have a spatial dimension or follow a geographical logic. We want to understand how terminology for specific materials traveled or not. Some ties were social: which materials spread with social change. The intellectual tying together is another facet: how people classified and organized materials in different ways. We look for associative patterns of thoughts in or across this large text corpus and ask questions of a quantitative and qualitative range that have not been researched yet, because they require a comprehensive view over all materials. In the Local Gazetteers project we transform the digitized text into an easy-to-work-with database, allowing historians to aggregate local knowledge from individual local gazetteers from different regions and different time periods and compile comprehensive datasets across historical China.

ASB: Could you briefly explain what a local gazetteer is in this context and tell us what time frame your project covers?

SPC: A local gazetteer is a summary term for regionally defined administrative synopses compiled by local officials or gentry since the tenth century. At least 8,000 local gazetteers from all over historical China are extant today, and some of them have been digitized as both images and searchable text. Their continuity and wide spatial existence make them a fantastic source of information for cross-regional and long-durée research.

At the same time each writing by itself provides a unique set of meticulous/detailed insights into a region's (a province, a prefecture, a city, a town) social, political, economic and religious characteristics, its topography and specialties. While not all the local gazetteers are organized in exactly the same manner, more or less they follow a similar structure and contain similar records. For example, almost every known local gazetteer provides information about local products, making them excellent sources for issues such as the regional distribution of material goods, trade relations, standardization of language, but certainly also on questions of environmental development or change.

The time-frame covered by this project ranges from the earliest local gazetteers that have been digitized (which, at the moment, date from the eleventh century) to the end of Republican China (1949). Most of the early local gazetteers were however lost, and most digitalized gazetteers date from the Qing Dynasty (1644–1911) and the Republican period (1912–1949).

ASB: What relevance do you think this project could have for practitioners of the environmental humanities? What kind of environmentally relevant information may be found in the sources you work on?

SPC: Local gazetteers contain detailed environmental information: for example, the flora and fauna of a region, rivers and mountains, hydraulics, weather records, natural disasters, and more. But they were local and thus are scattered without any central record keeping. Now we can collect them and aggregate them across broad geographical and chronological scales, and then compared or visualized to see comprehensive patterns. This could be very interesting for environmental historians.

ASB: In the project’s brief description on the Max Planck Institute website you state that you are “interested in exploring how the change of scales—by turning local records from individual gazetteers into a single global database—can reshape the study of historical China.” Could you briefly elaborate on this? This seems as an exceptionally important endeavor, in particular to understand environmental changes in a *longue durée* perspective.

SPC: Local gazetteers have been major sources for Chinese historians for decades. However, historians used to study individual local gazetteers through close-reading. This is due to two main reasons:

1. Chinese local gazetteers are physically preserved in different institutions in China, Japan, Taiwan, and Hong Kong and thus it’s often difficult to access different local gazetteers at the same time;
2. The core idea of local gazetteers was to collect regional knowledge for the purpose of local administration. This lack of a national dimension made it difficult to work comparatively on large batches of gazetteers. Only the advancement of digitalization programs has allowed blurring the boundaries of the knowledge contained within individual gazetteers.

What our project wants to contribute to the existing databases of local gazetteers are digital tools able to facilitate the aggregation of records kept in individual gazetteers. For this purpose we aim at producing a relational database collating records from all over China. In contrast to traditional full-text searchable databases that still require scholars to read and digest information by themselves, a relational database provides more structured results that can be easily mapped and analyzed by computers. Such a database can also support more research-oriented queries at the semantic rather than the textual level. For example, a query could be issued to retrieve all the grains recorded in available gazetteers, see the records from individual gazetteers immediately with their overall temporal and geospatial patterns, and analyze whether there are major changes in the patterns across regions and periods and what might be the causes.

ASB: Coming more specifically to the digital component of the project, we would like to know what tools you developed in-house and which instead you are getting off-the-shelf.

SPC: We are putting together three layers of digital tools in order to provide a full workflow from collecting and extracting to analyzing and visualizing data. The first layer of tools is a semi-automated tagging interface that helps the scholar to transform a section of text into a data table and thus records from individual gazetteers can be collected.

The large quantity of text makes it almost impossible to perform the transformation process in-house with limited human resources and diverse research interests. We want thus to set up a sharing platform where data collected by different scholars can be aggregated and contributors can get proper credit. We call this platform a “research data repository.” Right now we are trying out the [Dataverse Project](#), an open source software package developed by the Harvard Institute of Quantitative Social Sciences. We chose this platform because it already implements the idea of citing scholarly datasets as a way to promote data sharing in the academia. For data mapping and visualization we are using [PLATIN](#) (Place and Time Navigator), a tool developed in-house with funding from [TOPOI](#), a Berlin-based excellence cluster dedicated to the study of space and knowledge in antiquity.

ASB: And how do you proceed in practice? Could you tell our readers about the acquisition process of the sources, the encoding, and what final output you get?

SPC: The pivotal aspect of this project is that it aims at facilitating the transformation of texts, which are not immediately meaningful to computers, into structured data,

which computers can easily manipulate and analyze. Our major issue here is how to acquire proper “text mining” rights allowing our extraction interface to access copyrighted digital sources. MPIWG is working with the Berlin State Library on a pilot project in which the Library acquires a license with text mining rights from Chinese commercial vendors (through a network called CrossAsia set up by its East Asian Department) and MPIWG develops the digital tools to better use and exploit the sources.

Our extraction interface transforms the plain texts we get from the vendors into tabular datasets in CSV format. The tagged texts are also saved as XML files in order to keep track of all the changes made by each individual scholar. However, since these XMLs contain the original texts, they cannot be re-distributed on our research data repository. To avoid copyright violation the only thing available on the repository are spreadsheets.

ASB: Where does the project position itself within the digital humanities? Are there any projects that inspired you to start yours? How do you cooperate with the community beyond the Max Planck Institute?

Ours is a text-mining project, a branch of digital humanities that comes in two varieties. One is to use algorithms to mine hidden information in a large set of digital data without much human intervention. Examples of this approach include n-gram, named entity recognition, and topic modeling. The other approach is to rely on human interpretation to tell computer the meaning of (parts of) the texts before the computational analysis in order to improve its accuracy. For example, linguistic and literature scholars have been using TEI (Text Encoding Initiative) to mark up their texts before proceeding with the analysis. Our project also adopts the second approach, since the information recorded in local gazetteers are often very specialized and thus not easy to be retrieved by automatic algorithms.

The idea to allow more flexible and research-oriented queries of historical texts was inspired by the [China Biographical Database](#) project (CBDB), a relational database with biographical information about historical Chinese figures. Instead of including the biographical texts in its database, CBDB divides up each biographical text into different types of information such as names, places, methods of entering the government, postings, kinships, social relations, and stores them in different tables. In this way, CBDB allows to go beyond traditional person-specific queries and to look instead at whole groups of people with certain common backgrounds

(for example, the national exam passers in the whole Song dynasty). This way it becomes possible to research the patterns of their family or social attributes (for example, to see from which major regions Song exam passers came).

As regards collaboration with external partners, in 2015 we hosted a workshop on Chinese local gazetteers: we invited historians, computer scientists, and librarians to jointly explore what new questions can be generated and what new knowledge can be produced when enlarging the scale of analysis. In August 2016 MPIWG will host another workshop, in which eight invited scholars will test use our prototypes to explore their impact on actual, ongoing research projects. In the future, we hope to work with the Berlin State Library to make these tools available to a larger community of scholars within CrossAsia.

Links:

<https://www.mpiwg-berlin.mpg.de/>

https://www.mpiwg-berlin.mpg.de/en/research/projects/departmentSchaefer_SPC_MS_LocalMonographs

<https://www.mpiwg-berlin.mpg.de/en/users/schen>

<https://www.mpiwg-berlin.mpg.de/en/research/projects/departmentschaefer>

<http://dataverse.org/>

<http://platin.mpiwg-berlin.mpg.de/>

<http://www.topoi.org/>

<http://projects.iq.harvard.edu/cbdb/home>

<https://creativecommons.org/licenses/by/4.0/>

<http://www.antspiderbee.net/2016/04/19/material-network-analysis-an-exemplary-project-on-chinese-local-gazetteers/>

“Material network analysis: An exemplary project on Chinese local gazetteers,” an interview with Shih-Pei Chen by Wilko Graf von Hardenberg, is licensed under a [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. [An earlier version](http://www.antspiderbee.net/2016/04/19/material-network-analysis-an-exemplary-project-on-chinese-local-gazetteers/) was published 19 April 2016 in *Ant Spider Bee*.



Photo © 2016 Katrin Kleemann.

Flyover Country App, or What do airplanes and dinosaurs have in common?

Katrin Kleemann

On my flight home to Berlin from Munich last month (7 April 2016) I tested an app I recently read about. The app allows you to learn about places, points of interests, and geological features, such as lakes and mountains, when you fly over them. As you have to switch to airplane mode on the flight, the app allows you to download a flight path in advance, so you can access it when you are in the air.

The app Flyover Country was developed by Shane Loeffler, at the time a geology student at the University of Minnesota, with the help of a grant from the National Science Foundation. The app is fed by geological and paleontological databases. The content is linked to Wikipedia and the app also downloads the articles for you, so you can read them offline, but it's saved without the images—for reasons of memory space.