Reconstructing Local Materialities from Chinese Local Gazetteers

運用中國方志重建地方上的物質文化與認同

Shih-Pei Chen 陳詩沛, Martina Siebert 馬君蘭
Department III “Artefacts, Action and Knowledge”
Max Planck Institute for the History of Science (MPIWG)

2015 DADH
What are the Chinese local gazetteers (*difangzhi* 地方志)?

- A genre in historical China for recording **local knowledge**

- **Geography** 地理
  - maps, borders, landscape, subordinates

- **Infrastructure** 營建/建置
  - Building of the city and the local government offices, schools, temples, post stations, bridges...

- **Census and taxes** 戶口/田賦
  - Population, taxes from crops, civil services, financial status

- **Local products** 物產
  - grains, fruits, plants, drugs, animals, birds, fishes

- **People** 人物
  - Famous historical figures, local officials, degree holders, celebrity, ...

- **Weather, disaster, literature, ... and more.**
Characteristics of the local gazetteers

• **Long time span**: from 10th century onto now

• **Width in coverage**: Almost every well-populated region has its own local gazetteers. Mostly for administrative units: counties (xian 縣), prefectures (zhou州, fu府), and provinces.

• **Consistent in topics**

• **Vast in number**: 8,000+ titles extant today (accord. to 中研院中國大陸各省地方志書目查詢系統). Various commercial and open access digitization and full text projects (Beitu 北京圖書館, Erudition 愛如生, THDL 台灣歷史數位圖書館, etc.)

• **Lists of things**: a database waiting to be born
Geographical distribution of 2,000 digitized local gazetteers

2,000 local gazetteers in the first batch of Erudition “Zhongguo Fangzhi Ku” database. Background map: 1820 China with prefecture boundaries from China Historical GIS.
Reconstructing local materiality

- Local gazetteers were *tools of the state* (“owner’s manual”) *AND tools of local identity building.*

- Want to understand how “local materialities” were constructed through the compilation of Chinese local gazetteers by closely examining the chapters on *local products* (*wuchan* 物産 or *tuchan* 土產)
Section of “local products (wuchan)”

Flowers: plum, camellia, crab apple …

Bamboo varieties: “cat” bamboo, spotted bamboo …

Example pages on local products: Fujian tongzhi (Qing, 1737)
Research questions

• How products listed varied over time, how products were categorized into groups, regional variation of orders and categories of products, etc. What do they reflect in local identity building?

• How was local materiality and environmental diversity received in, validated by, and configured to the mainstream view?

• How did the re-enactment and literary preservation of landscapes and materials (in this special literature that each official could consult) affect the identity of a local region, its perseverance, or technical and scientific development?

• => requires looking at, comparing, analyzing records across multiple local gazetteers on larger scales
A technical approach for our research questions

• We first need to have an effective way of **compiling a national/global database** of local records for historical China that are collected **from individual local gazetteers**.
  – => A tagging interface that helps to **transform texts into data records** from which one can trace the sources

• So that **local records** from different sources can be viewed, compared, analyzed, and visualized together **as a whole**.
  – => A **research data repository** where historians can share and publish their data on local gazetteers and get academic credits.
  – => Analytical, geospatial, and visualization **digital tools** can be applied to the data collected
1) A tagging interface for transforming texts into data
Transforming texts into data

Texts  ➔  Data

Plain, unstructured  ➔  Texts with semantic structures that allows computer to recognize and process

Full text search  ➔  Computer assisted analysis (larger scale)

• Texts are not enough for enabling computer-assisted visualization/analysis, especially for larger scaled analysis
Our approach for transforming text into data

• A user interface that helps to transform digital texts and its built-in structures into data tables
  – You begin with tagging / marking up texts. You end up with a data table.

• Speeding up: automatic tools that helps to capture observable writing patterns for important information. E.g. name of item, places, category of item.

• This tool is first developed by the China Biographical Database project (CBDB) and further developed by MPIWG.
Defining topics & tags

- Topic: Local products (物產)
  - Main tag: product name 物產名稱
  - Descriptive tags: category 屬, production place 產地, alternative name 別名, sub product 種類, notes 註解, citation...

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Tag</th>
<th>Color</th>
<th>Save</th>
<th>Delete</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>物產名稱</td>
<td>product_name</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>屬</td>
<td>category</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>無意義</td>
<td>null</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>物產產地</td>
<td>product_place</td>
<td>blue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>別名</td>
<td>othername</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>種類</td>
<td>sub_product</td>
<td>green</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>種類名稱</td>
<td>sub_product_name</td>
<td>red</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>注解</td>
<td>note</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>待確認名稱!</td>
<td>product_name_check</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example Local Products chapter

Sampled text we acquired by copy & paste in order to develop this tool as an experiment.
To speed up tagging: Smart Regex

- Smart Regex: an automatic tool to capture observed writing patterns
  - E.g., product names are often surrounded by spaces

Credit: William Pang and CBDB
Step 1. Break the text into records (rows)
Right click to manually tag text as “category 屬” or as other tags
Step 2. Appending “category” to each row

Categories: veggies, leaf veggies
Step 3. Tag further information that you want to capture

2015 DADH
Step 4. Export as table

<table>
<thead>
<tr>
<th>Product name</th>
<th>Category</th>
<th>Alternative name</th>
<th>Desc</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015 DADH</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The proposed workflow

1. Define a topic
2. Define tags
3. Create regexes
4. Semi-automatic tagging
5. Export as tables

Examine result, modify regular expressions, and modify tagging result
2) A Research Data Repository

2015 DADH
To compile a national database from individual local gazetteers
Collaborative data collection

- A national database on local gazetteers records requires collaboration among scholars
- A **research data repository**: a place to aggregate scholarly produced research data. **Scholars can publish their data and get academic credits.**
- The Dataverse Network (DVN): an open source software

http://thedata.org
Our instance: LGDataverse (built on Dataverse)

A scholar can create datasets, each consists of data collected for a research study
3) Digital tools for visualizing and analyzing collected data
Text to data: enabling computerized analysis

- Texts
- Data tables

2015 DADH
LGMap for map visualization & analysis
(built on PLATIN GeoBrowser)

Blue: 532 local products collected from Quchou Fuzhi 衢州府志 (清)
Orange: distribution of 2000 local gazetteers in Zhongguo Fangzhi Ku

Blue: 532 local products collected from Quchou Fuzhi 衢州府志(清)
Orange: distribution of 2000 local gazetteers in Zhongguo Fangzhi Ku
Table view

<table>
<thead>
<tr>
<th>Book ID</th>
<th>Section ID</th>
<th>Title</th>
<th>Place</th>
<th>Page</th>
<th>Name</th>
<th>Check</th>
<th>Quote</th>
<th>Commercial Aspect</th>
<th>Processing/Manufacturing Issues</th>
<th>Reference Additional Places</th>
<th>Other</th>
<th>Name</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td>17036</td>
<td>医州府志</td>
<td>医州府</td>
<td>1482</td>
<td>北瓜</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
</tr>
<tr>
<td>48</td>
<td>17036</td>
<td>衡州府志</td>
<td>衡州府</td>
<td>1481</td>
<td>冬瓜</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
</tr>
<tr>
<td>49</td>
<td>17036</td>
<td>衡州府志</td>
<td>衡州府</td>
<td>1481</td>
<td>黄瓜</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
</tr>
</tbody>
</table>

2015 DADH
Analysis tools provided in PLATIN

User-defined pie charts based on “category” 屬

Temporal distribution & animation
Wrapping up

The LG project provides a set of tools to:

• Transform texts into data: in order to **compile a big database** from the local gazetteers in a efficient manner

  An computer assisted tagging interface

• Once texts become data => provide a research data repository for scholars to **publish data and get academic credits**

  An research data repository

• Once texts become data, computers can easily visualize and analyze data for us in a **larger scale**

  One click-away digital tools
Team members

- **Researchers:**
  - Dagmar Schäfer 薛鳳
  - Martina Siebert 馬君蘭

- **IT/DH specialists**
  - Shih-Pei Chen
  - Zoe Hong 洪明彤
  - Jorge Urzúa
Challenges and Opportunities for Sinologists, librarians, and information technologists

- Sinologists / humanists 漢學、人文學者:
  - New mind set for developing **digital methodologies** for effectively using digital research materials in order to advance humanities research

- Information technologists 資訊學者:
  - To **bridge** humanities research questions and computing algorithms by their knowledge in computers

- Librarians 圖書館員:
  - To acquire digital sources, especially the **text mining right**
  - To preserve the results of digital projects (both data and tools)