Persistence and Scale in Historical Gazetteers

Merrick Lex Berman - CHGIS, Center for Geographic Analysis
what is a gazetteer?

index of records defining tangible or abstract geographic features

- one or more names
- feature types / classes
- locations / footprints
- unique identifier
- relationships
- time periods
- supplemental information

ideal scenario:

- gazetteer entries serve as georeferences to footprints
- footprints relate to intelligible named features
- intermediary geographic references are persistent
attributes of placenames:

> official names / variants
  (Los Angeles / L.A.)
  (Köl n / Cologne)

> endonyms / exonyms
  (London / Londres)
  (Beijing / Peking)
  (Qomulangma / Everest)

> metonyms (synonyms)
  (Chicago / Windy City / Second City)

> vernacular pronunciations
  (Emei / Omei)
  (Peking / pa’k gyan)

> vernacular scripts
  (北京 / Beijing)
  (珠穆朗玛 / Qomulangma)
ideal scenario:

DISPLAY NAME > Qomulangma [chi] exo
ALTERNATE NAME > Jomo glangma [tib] endo
ALTERNATE NAME > Everest [eng] exo
ALTERNATE NAME > Sagarmatha [san] exo

excessive (though perhaps necessary) scenario:

<table>
<thead>
<tr>
<th>name</th>
<th>primary</th>
<th>variant</th>
<th>endo-exo</th>
<th>synonym</th>
<th>language</th>
<th>vernac</th>
<th>encoding</th>
<th>rom-system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beijing</td>
<td>Y</td>
<td>N</td>
<td>endo</td>
<td>N</td>
<td>chi</td>
<td>N</td>
<td>ascii</td>
<td>Pinyin</td>
</tr>
<tr>
<td>北京</td>
<td>N</td>
<td>N</td>
<td>endo</td>
<td>N</td>
<td>chi</td>
<td>Y</td>
<td>utf-8</td>
<td></td>
</tr>
<tr>
<td>Peking</td>
<td>N</td>
<td>Y</td>
<td>exo</td>
<td>N</td>
<td>eng</td>
<td>N</td>
<td>ascii</td>
<td>Non-Standard</td>
</tr>
<tr>
<td>Pekin</td>
<td>N</td>
<td>Y</td>
<td>exo</td>
<td>N</td>
<td>fre</td>
<td>N</td>
<td>ascii</td>
<td>Non-Standard</td>
</tr>
<tr>
<td>Pei-p'ing</td>
<td>N</td>
<td>Y</td>
<td>endo</td>
<td>N</td>
<td>chi</td>
<td>N</td>
<td>ascii</td>
<td>Wade-Giles</td>
</tr>
</tbody>
</table>
- persistence -

are we dealing with one entity that changes over time?
or a sequence of discrete historical instances?

relate unique instances to one another, creating a sequence

<table>
<thead>
<tr>
<th>name</th>
<th>vern</th>
<th>beg-year</th>
<th>end-year</th>
<th>relation</th>
<th>primary</th>
<th>lang</th>
<th>rom-system</th>
<th>id</th>
<th>prec_id</th>
<th>sub_id</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nanjing</td>
<td>南京</td>
<td>938</td>
<td>1152</td>
<td>Liao</td>
<td>Y</td>
<td>chi</td>
<td>Pinyin</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhongdu</td>
<td>中都</td>
<td>1153</td>
<td>1215</td>
<td>Jin</td>
<td>Y</td>
<td>chi</td>
<td>Pinyin</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dadu</td>
<td>大都</td>
<td>1267</td>
<td>1402</td>
<td>Yuan</td>
<td>Y</td>
<td>chi</td>
<td>Pinyin</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Khanbalik</td>
<td>胡鼻力</td>
<td>1267</td>
<td>1402</td>
<td>Yuan</td>
<td>N</td>
<td>mon</td>
<td>Non-standard</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cambaluc</td>
<td>胡鼻力</td>
<td>1267</td>
<td>1402</td>
<td>Yuan</td>
<td>N</td>
<td>ita</td>
<td>Non-standard</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beijing</td>
<td>北京</td>
<td>1403</td>
<td>1420</td>
<td>Ming</td>
<td>Y</td>
<td>chi</td>
<td>Pinyin</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jingshi</td>
<td>京师</td>
<td>1421</td>
<td>1644</td>
<td>Ming</td>
<td>N</td>
<td>chi</td>
<td>Pinyin</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shuntianfu</td>
<td>順天府</td>
<td>1421</td>
<td>1644</td>
<td>Ming</td>
<td>Y</td>
<td>chi</td>
<td>Pinyin</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beijing</td>
<td>北京</td>
<td>1644</td>
<td>1928</td>
<td>Qing</td>
<td>Y</td>
<td>chi</td>
<td>Pinyin</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beijing</td>
<td>北京</td>
<td>1929</td>
<td>1949</td>
<td>Republic</td>
<td>Y</td>
<td>chi</td>
<td>Wade-Giles</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beijing</td>
<td>北京</td>
<td>1949</td>
<td>present</td>
<td>PRC</td>
<td>Y</td>
<td>chi</td>
<td>Pinyin</td>
<td>11</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What does the Library of Congress recommend?

taken from the “LOC Pinyin Conversion Project – FAQ”

Ganzhou Fu (Gansu Sheng, China)
According to name authority record nr97-19923, the area now established as Zhangye (China) was known as Ganzhou Fu until 1912 and as Zhangye Xian from 1912 until 1985. According to the Subject Cataloging Manual: Subject Headings, H708, it is subject cataloging policy to assign as a subject heading or as a geographic subdivision only the latest name of a political jurisdiction that has had one or more earlier names, as long as the territorial identity remains essentially the same. Hence the headings Ganzhou Fu (Gansu Sheng, China) and Zhangye Xian (China) should not be used as a subject heading for any item that deals with this part of China.

Nanyang Diqu (China); Nanyang (Henan Sheng, China); Nanyang Xian (China)
Nanyang Diqu, Nanyang Shi and Nanyang Xian were dissolved in 1994. At that time, a new Nanyang Shi was established, covering almost the same area as the original Nanyang Diqu (except for Teng-hsien -- Deng Xian in pinyin -- which was left out of the new city). The new Nanyang Shi is equivalent to the former Nanyang Diqu, while the old Nanyang Shi is an ADM2 -- a smaller unit. Therefore, no change will be needed.

Amoy (China) changed to Xiamen (Fujian Sheng, China)
Swatow (China) changed to Shantou (Guangdong Sheng, China)
Port Arthur (China) changed to Lushun (China)
- duality and co-location -

one name identifies both jurisdictional *areas* & administrative *seats*
two different instances share the same *footprint* at the same time

Shiping County (Yunnan)

duality < Shiping County area [polygon]
Shiping County seat [point]
Yilong Township [point] > co-location
- time series result in asynchronous changes -

and *areas* & *seats* often change independently from one another!
- true persistence -

geographic reference system that does not change over time
- historical gazetteer records as indirect geographic references -

MATCHES FOR CARNABY STREET CARNABY STREET
1. Carnaby Street, W1, HOLBORN
2. Carnaby Street, W1, LONDON
3. Carnaby Street, M8, HARPURHEY
4. Carnaby Street, M8, MANCHESTER

<table>
<thead>
<tr>
<th>id</th>
<th>name</th>
<th>grid-code</th>
<th>long</th>
<th>lat</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>Carnaby Street</td>
<td>7500:110:3</td>
<td>-0.13</td>
<td>51.5</td>
</tr>
</tbody>
</table>

Carnaby Street, London 1966

Carnaby Street, London 1881
- scale -

resolving global, regional, local datasets is problematic

need to solve with metadata