DISCOVERING AND ACCESSING SOCIAL SCIENCE DATA OF EAST ASIAN COUNTRIES: TRENDS AND OBSTACLES

JUNGWON YANG
UNIVERISTY OF MICHIGAN
IASSIST CONFERENCE 2014
Interdisciplinary Studies in Social Science

- The increasing use of geographic information system (GIS), combined with the wider availability of sub-national statistics

- For example,
  - Sinclair (2011) : voting behavior in New Orleans
  - Neckerman (2009) and Ghiradelli (2010) : public health & social policy
  - Apparcicio (2008) : urban policy
The Data Visualization Class Project, the University of Michigan School of Information (Winter 2014)
Information about the existence of micro-data and the availability for the re-use is often difficult to find. Moreover, the language barrier, legal, cultural, and technological obstacles often exacerbate the difficulties of reusing the discovered data.
Question

- What kinds of geospatial and sub-national statistics on China, Korea, and Japan have been recently developed?
- What obstacles researchers have encountered in using the data in their research?
Data Collection

- The National Bureau of Statistics of China (NBS) collects sub-national statistics
- No geospatial data from government agency.

Data Distribution

- Fee-based database
  - The China Knowledge Resource Integrated (CNKI)
  - China Data Center
China: Fee-based Databases

- **CNKI (Chinese only)**

- **China Data Center (ICPSR, U of Michigan)**
  - China Data Online: Statistics
  - China Geo-explorer: Web-based spatial data
    - Unlike ArcGIS software, users cannot combine the data which they create with the data of China Geo-explorer.
    - China Geo-Explorer II: Export geospatial data as a shapefile.
China: Historical Geospatial Data

- The China in Time and Space (CITAS), U of Washington
  - The vectorized count level base maps
  - Georeferenced Socioeconomic Data (1982-1993)
  - The Socioeconomic Data and Application center (SEDAC), NASA

- CHGIS V4, Harvard Yenching Institution
  - Time series data (221 BC- 1911 CE)
  - 1820 data: spatial data of territory of the Qing Dynasty
  - ChinaW data (UC Davis): 1820-1893
  - 1911 data: spatial data for the provinces of Anhui, Fujian, Gansu, Guangdong, Hebei, Henan, Hubei, Hunan, Jiangsu, Shaanxi, Shandong, Zhejiang, and Zhili for the year 1911
  - CITAS data (1990 & 1997)
Non-governmental organizations tend to collect sub-national data targeted toward a particular context or issue (Thompson, 2010).

The academic institutes which provide open geospatial data are more likely to archive their data into other bigger and stabilized academic institutions.

Good metadata, citation, and copyright information
Korea : sub-national statistics

- The Korean Statistical Information Service (KOSIS)
  - English version : provincial level of subnational statistics
  - Korean version : municipal level
Korea: Geospatial Data

- The National Geographic Information Institution of Korea (NGII)
- The National Image Exchange (NIX): JPEG image + metadata
- The information disclosure act of Korea
Korea: Geospatial Data

- The National Geographic Information Institution of Korea (NGII)
  - The National Image Exchange (NIX): JPEG image + metadata
  - The information disclosure act of Korea
  - Access to NGII’s data from a foreign IP address is strictly prohibited
  - Alternative option?
Korea: Geospatial Data

- The National Geographic Information Institution of Korea (NGII)
  - The National Image Exchange (NIX): JPEG image + metadata
  - The information disclosure act of Korea
  - Access to NGII’s data from a foreign IP address is strictly prohibited
  - Alternative option?
    - Use geospatial data service of local governments
Korea: Geospatial data of local governments

- Seoul metropolitan city map service
- English version map service for tourists vs. Korean version map service for academic research uses
- Article 21 of the Public Survey and Overseas Export Ban Act
The majority of provincial governments (군, gun) and metropolitan cities of Korea provide GIS information for their regions.

Overlay, download customized map

Technical problem

Internet Explorer package installation requirement
Good news!

The Korean central and local governments have a strong interest to develop and distribute geospatial data as well as the sub-national statistics to the public. However, the GIS information service is mainly developed in order to serve the Korean citizens. Some geospatial data are illegible for acquisition from overseas.
The Statistics Bureau and the Ministry of Internal Affairs and Communication: both statistics and geospatial data

- E-Stat for sub-national statistics
- Currently, the geocoded census data (2000, 2005, and 2010), the establishment and enterprise census (2001 and 2006), the economic census data (2009), and the census of agriculture and forestry data (2005 and 2010) can be downloaded in the shapefile format.
- Interactive map service (Japanese only) in e-Stat
The Global Map Japan database of the Geospatial Information Authority of Japan (GSI), the Ministry of Land, Infrastructure, and Transport of Tourism
- Land Cover, vegetation, administrative boundary, population, and transportation data

The collection and distribution procedures of government data in Japan is highly centralized. They are more willing to share their geospatial data with the overseas’ users compared to the Korean and Chinese governments.
Data Collection and Distribution

- Most of the Chinese sub-national statistics are not available from the government website, but all subnational statistics of Korea and Japan are freely available from their Statistics Bureau websites.

- Geospatial data
  - China: fee-based database
  - Korea: focused on Korean citizen
  - Japan: little restriction to overseas users
Trends and Obstacles

- **Reliability of Data**
  - China data, especially China Data Online

- **Language Barrier**
  - Korean and Japanese data, in particular interactive map service

- **Technological concern**
  - Korea: Internet Explorer package installation requirement
  - China Geo-explorer: no overlay option, out-of-dated data
Conclusion

- We can find quite a large amount of subnational statistics and geospatial data from the governments, as well as academic institutions in the United States. However, the reliability of data issue (China), the language barrier (Korea and Japan), the copyright and legal restriction (China and Korea), and technical issue (China and Korea) make it difficult for researchers to find and use the data of East Asian countries.

- In the academic library setting, collaboration and partnership among the data librarians, area specialists, and copyright librarians will be critical in order to overcome these legal, cultural, and technical issues and to support the social scientists’ interdisciplinary studies.
Thank you

- Any Question?

[Image source: http://thefamilycowlick.tumblr.com/]

THAT IS AN EXCELLENT QUESTION


Chua, H. (2012) Indiastat and China Data Center Online: an evaluation and comparison, Reference Reviews 26(2)


