Mining Ministers (1572-1815)
Using Semi-structured Data for Historical Research

Histoinformatics, 10 November, 2014

Serge ter Braake, Antske Fokkens, Fred van Lieburg
Main Issues

• Historical datasets from the past (eighties-noughties) usually not compatible with modern software

• Historical datasets from the past not open access

• Historical datasets from the past not able to communicate with the outside world (walled in knowledge)
Dutch Ministers 1572-1815

- At the center of any reformed community in the Netherlands
- Responsible for the religious well-being of their congregation
- Likely to be remembered in historical records
- Often well-documented
- 12405 men in the research period
The Data: a (nearly) complete set of ministers 1572-1815

Lists of Ministers, compiled since start of Reformation

Lists gathered by W.C. Regt, start 20th century, in a manuscript

CSV files, data available in github

Used by Fred van Lieburg (1995) as the basis for his dataset, originally stored in Dbase 3+
Working with the Data

• CSV: Comma-separated values, stored as plain text files. Ensures interoperability of different formats. 3 star Open Data
• Normalizing dates
• Disambiguating place names <= GeoNames (http://www.geonames.org), not a trivial task since many Dutch colonies were named after Dutch cities
Old and New Questions

• Old: Statistical Analyses: e.g. where were the ministers born? <= A good check to see if our conversion worked

• New: What can we say about the mobility of ministers through the centuries? Place of birth => Location first Job => Second Job, et cet. => Place of Death. <= Kilometres calculated with the help of GeoNames
Mobility of Dutch ministers 1575-1815

Fig. 3. Mobility in kilometres (distance moved between the known place of birth, work and death) per 50 years (based on dates of death of the ministers)
Conclusions

• Historical datasets from the past (eighties-noughties) usually not compatible with modern software <= CSV
• Historical datasets from the past not open Access <= Github; in the future: national institution responsible for hosting datasets
• Historical datasets from the past not able to communicate with the outside world (walled in knowledge) <= GeoNames; in the future: 5 Star Linked Open Data to enable more enhanced historical querying