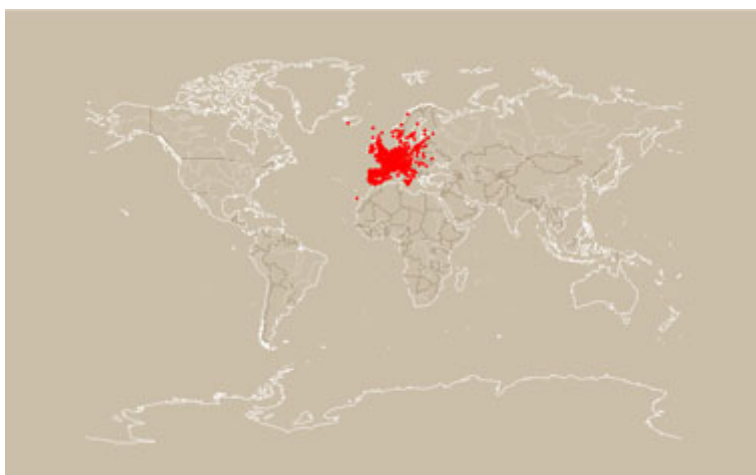




YERSINIA

Georeferences for places of plague occurrence in Europe 1347-1600

Vlad ATANASIU¹, Claire PRIOL², Anne TOURNIEROUX², Ezio ORNATO²



(1) Commission for Scientific Visualization, Austrian Academy of Sciences, Vienna, Austria,
atanasiu@alum.mit.edu, <http://www.viskom.oeaw.ac.at>

(2) Laboratory for Medieval Studies of Paris, National Center for Scientific Research / University Paris 1, Paris, France,
erialch@gmail.com, anne.tournieroux@wanadoo.fr, ezornato@vjf.cnrs.fr, <http://lamop.univ-paris1.fr/W3/>

Purpose

Yersinia provides georeferences for the places of plague occurrences in Europe 1347-1600. These include for each year of occurrence, placenames, coordinates and region names. There are approximately 850 georeferenced unique locations in the dataset. The file is distributed in three formats: Microsoft Excel 2003 file, Unicode UTF-16 comma-delimited text file, KML file. Yersinia is part of the Bernstein Paper Atlas (<http://www.bernstein.oeaw.ac.at>).

The source used for generating Yersinia is Jean-Noël Biraben, *Les Hommes et la peste en France et dans les pays européens et méditerranéens*, Paris/La Haye, Mouton/École des hautes études en sciences sociales, 1975-6.

Description

The following describes the columns of the Yersinia file.

A. *Serial number* –

B. *Year* – year of plague occurrence

C. *Status* – information about the nature of the placenames.

Georeferenced placenames – placenames for which coordinates are provided.

location – locations considered punctiform (e.g.: city, abbey, castle).

region – locations considered surfaces: administrative units, geographical and historical regions (e.g.: county of York, Land of Bayern, Transylvania). Regions were normalized to the principal locations in the region (e.g.: county of York > city of York, Land of Bayern > city of München, Bohemia > Prague).

Placenames not georeferenced – placenames for which coordinates are not provided

ambiguous – placenames referring to more than one location that could not be disambiguated (e.g.: Gazzo appears once in the region Emilia Romagna, once in Lombardia and twice in Veneto).

not identified – placenames that could not be identified (e.g.: Two Waters, Dorf “Zahaeim”).

D-E. *Placenames*

D. *Original placename* – placenames as given in the sources. Accents are preserved.

E. *Normalized placename* – placenames providing a unique name homonymous placenames in the sources (e.g.: Bécés, Vienne, Vienna, Wien > Wien). The local or most common name is the preferred one.

F-G. *Coordinates* – Decimal coordinates, rounded to two decimals. When more than one set of coordinates were found for a given location, only one was retained. Empty fields mean that the location is of a type that didn't allow georeferencing.

F. *Latitude*

G. *Longitude*

H-M. *Regions* – Names and codes of the placenames' regions (e.g.: Oxford > Oxfordshire).

H. *Code* – Code of the region. For EU-27 and EFTA countries this is the code level 3 of the Nomenclature of Territorial Units for Statistics 2003 (<http://ec.europa.eu/eurostat/ramon/nuts/> and http://www.bfs.admin.ch/bfs/portal/fr/index/international/11/geo/analyse_regionen/11.html). For other countries the country code according to ISO 3166-1 alpha 2 was used (http://www.iso.org/iso/country_codes.htm).

I. *Country* – Name of the country as given by the two standards named above.

J. *NUTS 1 (2003)* – name of NUTS region level 1, state codes for Canada and United States.

K. *NUTS 2 (2003)* – name of NUTS region level 2.

L. *NUTS 3 (2003)* – name of NUTS region level 3.

M. *NUTS 4 (2003)* – name of NUTS region level 4.

N. *Remarks* – Various remarks, mostly about how ambiguities were solved or why some placenames are considered as ambiguous.

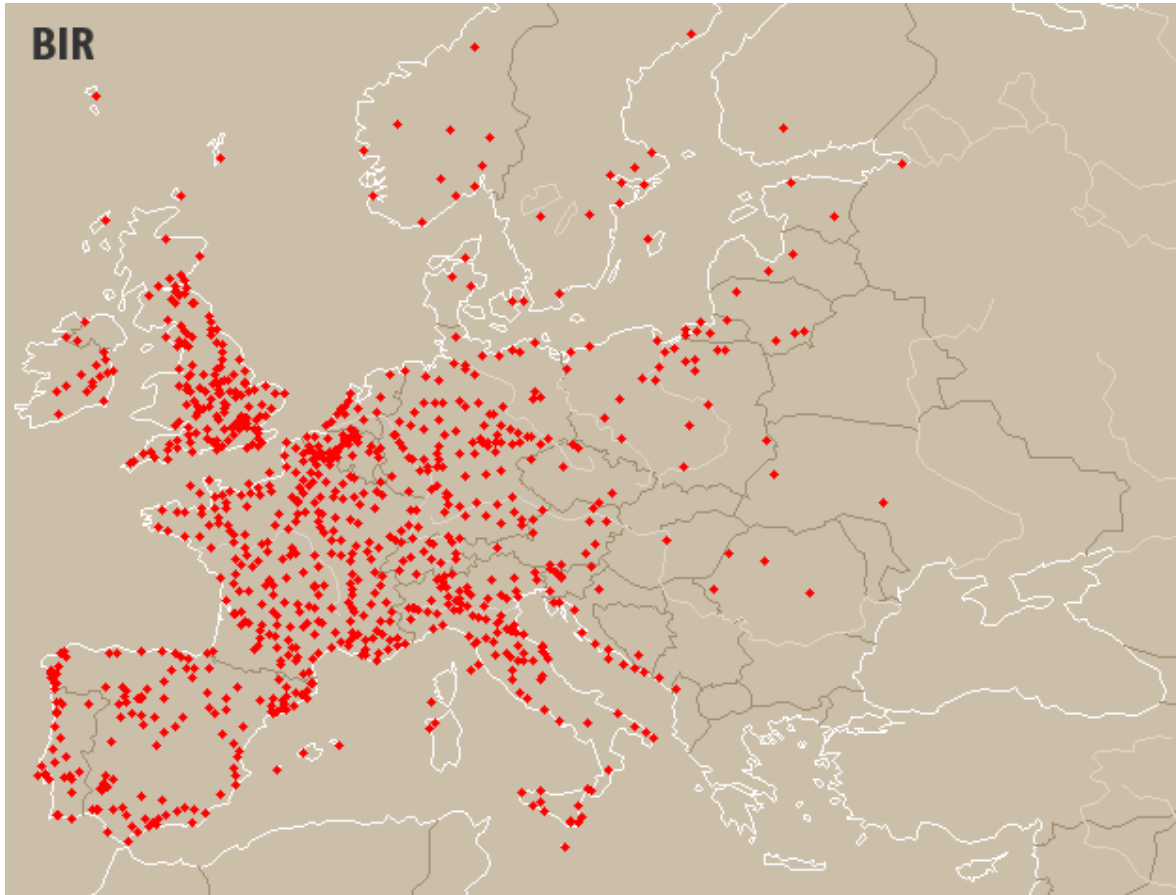
Quantities

The following tables provide a quantitative insight in Yersinia. Updates to the files might modify the quantities.

	Records		Unique values		
Locations	839	74 %	821	75	98 %
Regions	192	17	169	15	88
Ambiguous	36	3	36	3	100
Not identified	68	6	68	6	100

Visualizations

1. *Animation* – A timeframe animation is supplied in the Yersinia kit (file yersinia.mpg). For every location having a 'region' status, all locations in the region as given in the Bernstein Dalek file were made visible in the animation.
2. *Spatio-temporal distribution* – A static visualization that still allows to visually analyze plague occurrences in time without the need for an animation is also part of the Yersinia kit (file the_score_of_death.pdf).
3. *Spatial distribution* – The map below shows the spatial distribution of locations in Yersinia.



Method

Notes on the method of georeferencing.

1. *Reattribution* – Although some placenames are identified, they are either too narrowly defined or too wide, so that reattribution has been operated.

a. *Inclusion* – Streets and neighborhoods are attributed the name of the respective cities (e.g.: ‘Bread Street’ and ‘Lambeth’ are normalized to ‘London’), and abbeys and castles the names of the villages or cities close to which they are located.

b. *Focusing* – Names of regions are normalized to the principal settlement in the area (e.g.: ‘Alsace’ > ‘Strasbourg’). For cross-border regions the attribution might be arbitrary (‘Flanders’ is partly in present day Belgium, France and Netherlands).

2. *Find coordinates* – Once it was established which location the source means, its geographical coordinates had to be found. All coordinates were converted to decimal degrees and rounded to two decimals, giving a precision of about 1 km. The main sources of information are Geonames (<http://www.geonames.org>) and Wikipedia (http://en.wikipedia.org/wiki/Main_Page), various other online services being also used, such as Getty Thesaurus of Geographic Names (http://www.getty.edu/research/conducting_research/vocabularies/tgn/) and Falling Rain Genomics (<http://www.fallingrain.com/world/>).

3. *Attribute administrative units* – Providing the administrative unit to which a place belongs allows performing geographical statistics by clustering point-locations into areas. The overview thus gained helps the historian in his investigative work of the past. This is the reason why along with the coordinates’ information on the administrative units was provided. Because administrative units are dynamical historical entities – created, abolished, their boundaries changed – it was decided to refer to the present state of administrative units. Statistical units (NUTS) were chosen over administrative units (AU) because there were available to the creators of this dataset both as nomenclature and vector boundaries for almost the entire area covered by the dataset (Europe).

Limitations

While great care has been taken to georeference the placenames, errors might subsist, some which are even not stemming from the authors of the file. Here is a list of error sources identified during the development of Yersinia. Most of the errors could be identified using the Dibuk software (search on <http://www.bernstein.oew.ac.at>).

— *Historical sources issues*

1. error on the placename in the original document
2. error in reading the document by the modern-time author of the source used by Dalek

— *Geographical sources issues*

3. error in the gazetteers on coordinates, names, accents, regions of belonging (e.g.: non-ASCII characters are stripped of their accents in GeoNames and the NUTS)
4. incorrect boundaries due to digitization errors or faulty generalizations (e.g.: in Matlab there are two discontinuities in the boundary of the Faroe Islands)
5. reorganization of regional boundaries (e.g.: breakup of Yougoslavia, reorganization of NUTS every few years)
6. several closely situated coordinates for the same placename (e.g.: when one source geolocates Paris with the coordinates of the Eiffel Tower and another with those of the Notre Dame – which one to follow?)
7. coordinates with various levels of precision (e.g.: 45.1 and 45.11 considered as two distinct locations)
8. system transformation errors (e.g.: coordinates and datum)

— *Development issues*

9. typing errors
10. mixing of systems (e.g.: decimal/sexagesimal coordinates, geographical datum)
11. correct coordinates but attribution to wrong region (e.g.: human error or map scale with insufficient detail)
12. variants not identified as such (e.g.: ‘Dubrovnik’ and ‘Ragusa’ are the Serbo-Croatian and Italian names of the same place, which is not self-evident without the necessary background knowledge)

License

Copyright (c) 2008 Vlad ATANASIU, Claire PRIOL, Anne TOURNIEROUX, Ezio ORNATO

Permission is hereby granted, free of charge, to any person obtaining a copy of this dataset and associated documentation files (the "Dataset"), to deal in the Dataset without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Dataset, and to permit persons to whom the Dataset is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Dataset.

THE DATASET IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE DATASET OR THE USE OR OTHER DEALINGS IN THE DATASET.

Credit

The concept of this dataset originates with Ezio ORNATO, with input from Vlad ATANASIU; georeferencing was done by Claire PRIOL and Anne TOURNIEROUX; development software for visualization and debugging written by Vlad ATANASIU. The work has been undertaken in 2007-8 at the Laboratory for Medieval Studies of Paris, National Center for Scientific Research / University Paris 1, Paris, France, within the project “Bernstein – The Memory of Papers” (<http://www.bernstein.oeaw.ac.at>), co-funded by the European Commission, under the programme eContentPlus (ECP-2005-CULT-038097/Bernstein).