Opening-up historical maps

Hannes Kulovits
Austrian State Archives

Rainer Simon
Austrian Institute of Technology

November 6-7 2013 | Cartography and cadastral maps – Visions from the past, for a vision of our future, Pisa, Italy

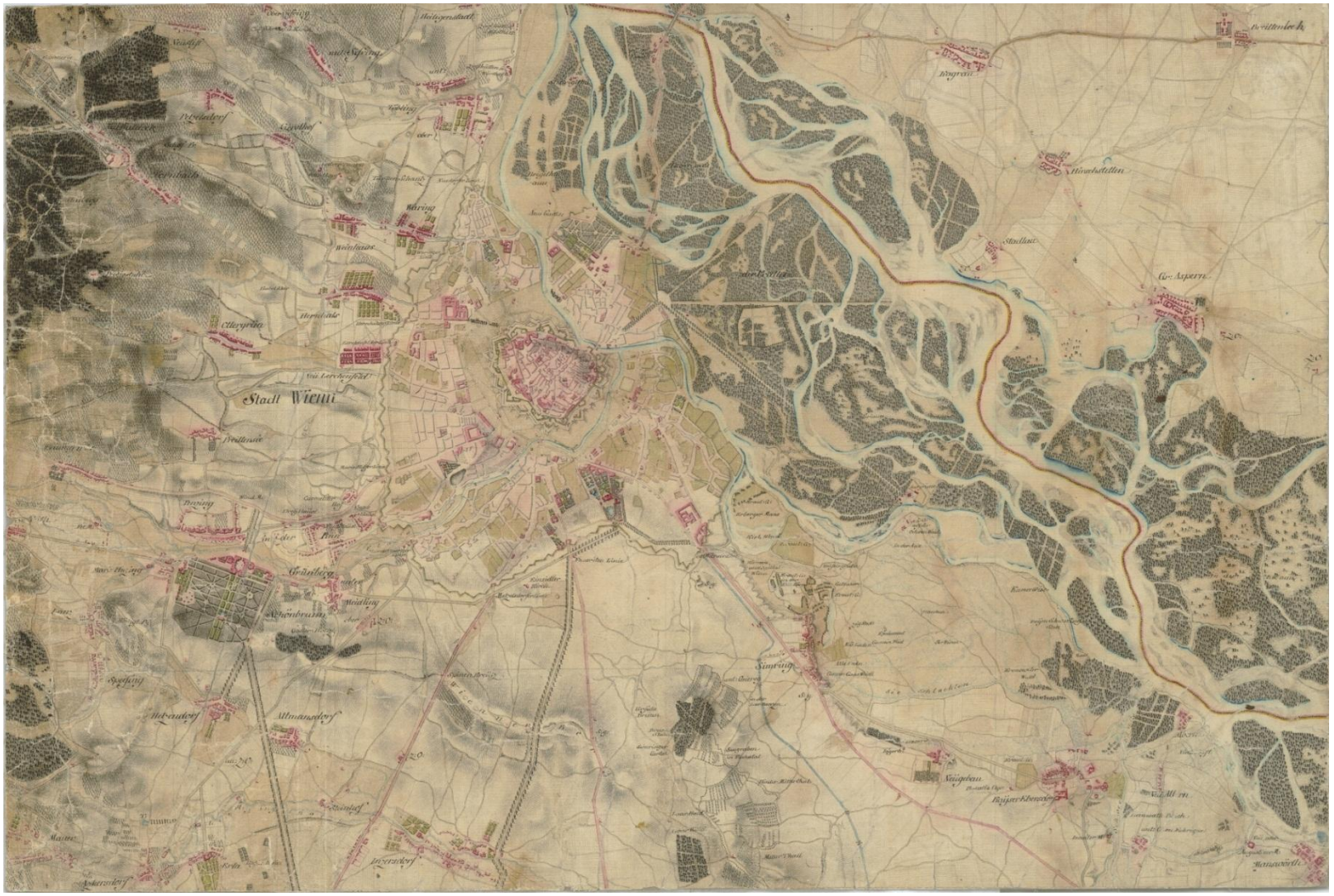
Austrian State Archives

Holding 600’000 maps from 16th to present.
No experience in making digital maps available online however. We don’t want to re-invent the wheel. Technical partner required.

Stakeholders

Historians/ Researchers of various fields
Military history, local history, changes in landscape, economic history

Map Digitization



But:

Search & discovery limited
by available catalog metadata!

Spatial Search?

Toponyms?

„Venta“

(Caistor St. Edmund)

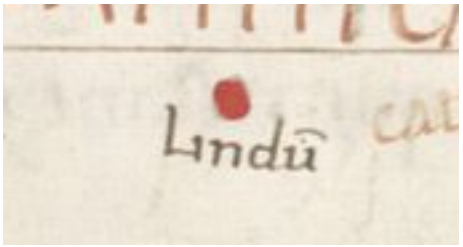
Toponyms?

„Mediolanum“

(Whitchurch)

Toponyms?

„Lindum“



(Lincoln)

Requirements

R1 – Enable semi-automatic process
R2 – Make available on the internet
R3 – Enable feedback/annotation by public

Transcription? OCR?

Remains a huge challenge.

Pelagios 3 envisions a *semi-automatic* process in which a human expert is assisted in "annotation grunt work".

Give recommendations.

Goal:

detect toponyms on old maps
with open source tools.

Approach:

determine *location, extent* and *orientation*
using image processing and
connected component analysis

Step 1: Background/Foreground

Note: our steps still include manual tuning/intervention.

Step 1: Background/Foreground

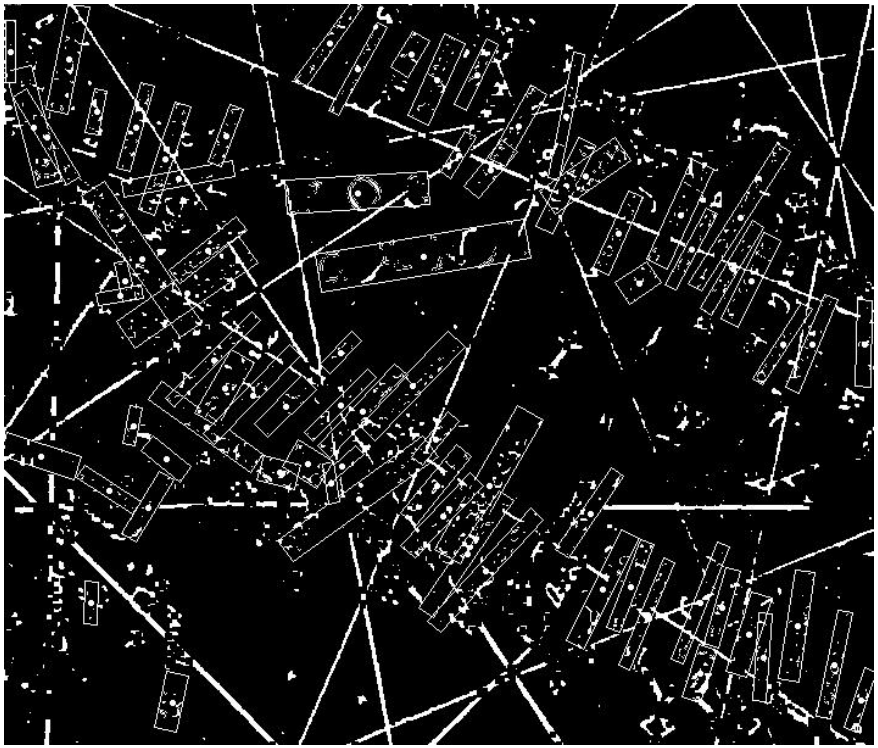
- Manual definition of color- and brightness ranges to treat as background layer
- Optional:
  - Median filtering (which has a „blurring effect“)
  - Substracting the filtered image from the original yields a foreground image
- Additional image processing to remove lines & elements with low color gradient
- Morphological operations to clean up „connected shapes“ (such as characters)

Step 2: Feature Detection

Step 2: Feature Detection

- Automatic identification of connected shapes (characters or groups of characters) using a contour detection algorithm
- Implemented using OpenCV open source image processing software
- Additional heuristics to filter for plausibility
  - Covered area of the shape
  - Shape width
  - Shape aspect ratio

Step 3: Feature Linking



Step 3: Feature Linking

- Aggregates features (=characters or character groups) into sets representing a single toponym
- Sets are „grown“ from a single feature by including one nearby feature at a time
- Empirical constraints applied in each step:
  - Distance between features
  - Bounding box covered by features combined vs. features individually
  - Baseline orientation of features combined vs. individual
  - Absolute bounds on width and height

TOS
CANA

Some preliminary results:

Experiment #1

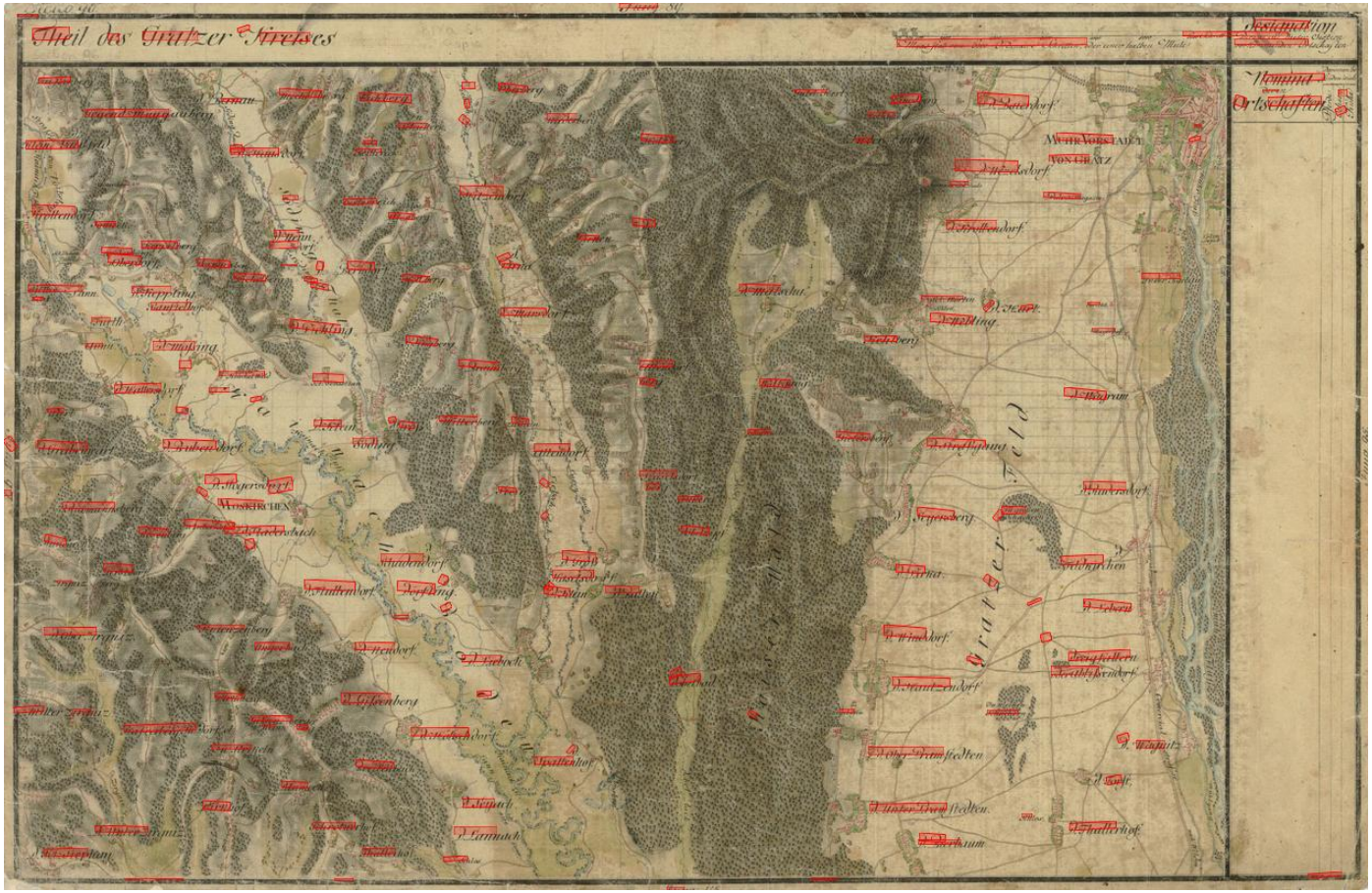
Ptolemaic Map of the British Isles, ca. 1480

Experiment #2

Salvator Oliva, Mediterranean. Portolan Atlas, Marseilles, 1619.

- 333 toponyms counted
- 532 detections
- ~50% recall
- ~31% precision

Experiment #3



Josephinische Landesaufnahme: Gebiet von Moos-kirchen bis Grazer Feld, Österreich, 1764-1787.

Full-resolution views at

http://rsimon.github.com/toponym_identification

Future work – *Pelagios 3*:

- Provide an index of toponyms in early geospatial documents
- Create assistive annotation tools to enhance the index
- Develop an "analysis workbench" to conduct visual & statistical comparison between different collections and documents

Conclusions

- Appraisal of historical maps requires semi-automatic processing
- Assistance in transcription process required
- Maps should be made available online
- Maps and toponyms “linkable”

Thank you for your attention

http://pelagios-project.blogspot.co.uk

hannes.kulovits@oesta.gv.at

rainer.simon@ait.ac.at
@aboutgeo

References

[1] http://pelagios-project.blogspot.co.uk
[2] http://gamera.informatik.hsnr.de/
[3] http://rsimon.github.io/toponym_identification/