Exploring Old Maps
EOM 2017

Proceedings

Thomas C. van Dijk, Christoph Schommer (Eds.)
Universität Würzburg, 6–7 April 2017

International Workshop on Exploring Old Maps

University of Würzburg, Campus Hubland South
Thursday April 6 and Friday April 7, 2017

Preface

Many libraries own an extensive collection of historical maps. Beside their value as historical objects, these maps are an important source of information for researchers in various scientific disciplines. This ranges from the actual history of cartography and general history to the geographic and social sciences. With the progressing digitisation of libraries and archives, these maps become more easily available to a larger public. A basic level of digitisation consists of scanned bitmap images, tagged with some basic bibliographic information such as title, author and year of production. In order to make the maps more accessible, further metadata describing the contained information is desirable. This would enable more user-friendly interfaces, relevant queries of a database, and automatic analyses.

These are the proceedings of the Second International Workshop on Exploring Old Maps (EOM 2017). The workshop provides a forum for the communication of results that may be useful to the community. Researchers and practitioners of many areas working on unlocking the content of old maps have contributed to this year's program — humanities scholars, developers, computer and information scientists and map enthusiasts.

Program Committee
Christoph Schommer, *Dept. of Computer Science, University of Luxembourg*
Thomas van Dijk, *Dept. of Computer Science, University of Würzburg*
Geoffrey Caruso, *Institute of Geography and Spatial Planning, University of Luxembourg*
Alexander Wolff, *Dept. of Computer Science, University of Würzburg*

Organising Committee
Benedikt Budig, *Dept. of Computer Science, University of Würzburg*
Thomas van Dijk, *Dept. of Computer Science, University of Würzburg*
Sigrid Keller, *Dept. of Computer Science, University of Würzburg*
Alexander Wolff, *Dept. of Computer Science, University of Würzburg*

Programme: Thursday, April 6

Zentrales Hörsaal- und Seminargebäude (Z6), Room 2.012 (second floor).

13:15 **Opening & Ice breaker with maps**

13:30 **Invited Talk I** — *Yao-Yi Chiang*
Spatial Sciences Institute, University of Southern California

Querying Historical Maps as a Unified, Structured, and Linked Spatiotemporal Source

14:30 Break

14:50 **Georeferencing Old Large-Scale Maps from a Research on Historical Routes Perspective**

Daniel Blank and Andreas Henrich

15:10 **Metadata-Aware Map Processing – An Automated Metadata Retrieval and Management Workflow for Analysing Old Maps**

Hendrik Herold, H. Kollai, M. Neubert, G. Meinel, R. Grunktzke and P. Winkler

15:30 **Open Historical Data Map – a Prototype**

Thomas Schwotzer

15:50 End of contributed talks on day 1

16:00 **Tour of the Würzburg University Library map collection**

18:30 Dinner at Würzburger Ratskeller

Programme: Friday, April 7

Zentrales Hörsaal- und Seminargebäude (Z6), Room 2.012 (second floor).

9:00 **Invited Talk II** — *Martijn Storms*
Leiden University Libraries

Georeferencing by Crowdsourcing: The Maps in the Crowd-project of Leiden University Libraries

10:00 **Deep Learning for Place Name OCR in Early Maps**

Winfried Höhn

10:20 **Exploring Local Geography for Toponym Matching**

Benedikt Budig

10:40 Coffee break

11:00 **Digitising Schematic Maps: Recreating or Reinventing History**

Maxwell Roberts

11:20 **From Many User-Contributed Polygons to One Building Footprint**

Fabian Feitsch

11:40 **Annotating Old Maps with Recogito 2**

Rainer Simon, Valeria Vitale, Leif Isaksen, Elton Barker and Rebecca Kahn

12:00 **Tunnels of Knowledge: Mapping Today's "Secrets" from Yesterday's Public Maps (and Improving Public Safety)**

Paul Sladen

12:20 EOM 2018: When & Where?

Group Discussions: What is next for old maps?

13:00 Closing / Lunch

Keynote Speaker

Yao-Yi Chiang
University of Southern California (USC)

Querying Historical Maps as a Unified, Structured, and Linked Spatiotemporal Source

Abstract: Historical spatiotemporal datasets are important for a variety of studies such as cancer and environmental epidemiology, urbanization, and landscape ecology. However, existing data sources typically contain only contemporary datasets. Historical maps hold a great deal of detailed geographic information at various times in the past. Yet, finding relevant maps is difficult and the map content are not machine readable. I envision a map processing, modeling, linking, and publishing framework that allows querying historical map collections as a unified and structured spatiotemporal source in which individual geographic phenomena (extracted from maps) are modeled with semantic descriptions and linked to other data sources (e.g., DBpedia). This framework will make it possible to efficiently study historical spatiotemporal datasets on a large scale. Realizing such a framework poses significant research challenges in multiple fields in computer science including digital map processing, data integration, and the Semantic Web technologies, and other disciplines such as spatial, earth, social, and health sciences. Tackling these challenges will not only advance research in computer science but also present a unique opportunity for interdisciplinary research.

About the author: Yao-Yi Chiang is Assistant Professor of Spatial Sciences in the Spatial Sciences Institute, University of Southern California (USC). His general area of research is artificial intelligence and data science, with a focus on information integration and spatial data analytics. He develops computer algorithms and applications that discover, collect, fuse, and analyze data from heterogeneous sources to solve real world problems. He teaches data mining, spatial databases, and mobile GIS. Prior to USC, Chiang worked as a research scientist for Geosemble Technologies (now TerraGo Technologies), which was founded based on a patent of which he was a co-inventor on geospatial-data fusion techniques.

Keynote Speaker

Martijn Storms
Leiden University Libraries

Georeferencing by Crowdsourcing: The Maps in the Crowd-project of Leiden University Libraries

Abstract: Leiden University Libraries holds a vast collection of c. 100,000 maps and 3,500 atlases. To make these maps better accessible and searchable, the georeferencing of the map collection has started. Moreover, georeferenced maps can be used and analysed further in Geographical Information Systems. With the well-known Georeferencer application of Klokan Technologies, everyone who is interested could contribute to this project online. After a pilot project in 2015, consisting of c. 400 18th century manuscript charts of the Van Keulen collection, almost 7,000 Dutch colonial maps of the KITLV collection (Royal Netherlands Institute for Southeast Asian and Caribbean Studies) were made available for georeferencing. In eight months' time all maps were georeferenced by the crowd. At this moment a third phase of Maps in the Crowd is in preparation, with c. 1,000 maps of the Caribbean in general and the Netherlands Antilles in particular.

This talk will focus on the crowdsourcing aspects of the project. How did we reach people who were interested to contribute? What kind of PR activities did we develop to bring the project to the attention? Who are the people who participated? And how did we keep them satisfied? The project cannot be dissociated from the broader digital revolution in the academic and librarian world. Leiden University Libraries is developing a Center for Digital Scholarship to assist researchers and students in all aspects of digital scholarship. At the same time the library is creating a new repository infrastructure to manage and display its digital and digitised collections. Maps in the Crowd perfectly fits in this larger development.

About the author: Martijn Storms MA (Arnhem, 1978), is curator of maps and atlases at Leiden University Libraries. Besides, he is project coordinator at Brill publishers and member of the editing board of Caert-Thresoor, the Dutch journal on the history of cartography. Martijn Storms studied GIS and cartography at Utrecht University.

Georeferencing Old Large-Scale Maps from a Research on Historical Routes Perspective

Daniel Blank and Andreas Henrich
Media Informatics Group, University of Bamberg, Germany
Email: {daniel.blank|andreas.henrich}@uni-bamberg.de

I. INTRODUCTION AND RELATED WORK

Old large-scale maps are important information sources for research on georeferencing historical routes. This research is largely supported by local researchers working in the terrain. They georeference map entities such as cadastral units, boundary stones, and wayside shrines by hand. Such groundtruth data is valuable for the development of new digital map processing algorithms. By characterizing those georeferencing and disambiguation tasks we outline potential requirements for algorithmic solutions.

Understanding text and georeferencing entities of large-scale historical maps is still in its infancy. More focus lies on the segmentation and the matching of labels and markers (see e.g. [1]). Research on historical route networks and travel itineraries is often concerned with the georeferencing of route stops only. As a particular algorithmic approach, we address the georeferencing of route stops denoted by historical place names in [2]. The algorithm input is a sequence of place names and distance estimates between them as e.g. explicitly contained in mile disks and maps with mile dots (see Fig. 1a and 1c). Weinman [3] targets the georeferencing by aligning toponyms on maps and those from a gazetteer through a probabilistic model. Georeferencing maps using image processing based on line segment features is e.g. addressed in [4]. Tools for georeferencing historical maps in their entirety[1] and for the display of map overlays[2] exist. However, there is a need for more automatic approaches and an in-depth georeferencing.

II. GEOREFERENCING HISTORICAL LARGE-SCALE MAPS

The georeferencing of historical routes relies on an in-depth georeferencing of old large-scale maps. We distinguish four properties useful for the disambiguation of map entities:

Textual: Place names on old maps (e.g. names of lakes, rivers, roads, settlements) might match with gazetteer entries and ambiguities might have to be resolved (see e.g. Fig. 2 and the ambiguous *Steinbach* present in GeoNames). Historical spelling must also be addressed such as for *Eyersheim* (see e.g. [2] for algorithms). A special challenge arises from place names of locations that no longer exist (see e.g. Fig. 2: the abandoned village *Wettenburg*), and from unpopulated places (see e.g. Fig. 2: *Wart* as a reference to a historical tower). Attached markers might help in the disambiguation.

Field names and route names are full of semantics (see Fig. 3 where *Wein Straße*, *Post-Straße*, *alte Straße* all refer to the same route). It can also be observed that *Bircken Schlag* and *Wüstung* (Fig. 3) refer to an area also denoted as *Pechofenschlag* (Fig. 4b). Besides textual clues, image processing based on line and shape matching is useful as well.

Figures 1a and 1c show that the track between Tauberbischofsheim and Miltenberg is part of a larger route towards Frankfurt. Such information is important when georeferencing old maps (e.g. Fig. 1b can be identified as part of it). In addition, *Franckfürter Stras* can be detected as a fourth name addressing the old route (Fig. 1d).

Numbered: Boundary stones, wayside shrines, etc. provide important reference points and they might still be present in the terrain. References to boundary stones are often numbered. The same numbers may occur multiple times on a map segment. Stones may have been (re)moved, or the numbering may have changed over time (see Fig. 4). Disambiguation including geospatial and textual clues is necessary here, too.

Marked: Template matches of markers across multiple maps cannot always be assumed. Yet, symbolic representations can be useful (Fig. 5). Singular stones are difficult to georeference. However, shape properties from their connections and relative locations reveal useful information (see e.g. Fig. 3).

Image-content-based: Map entities such as lakes, roads, lots can be georeferenced using matching techniques from image processing relying on color, texture, and shape (see [4], [5]); sometimes in addition to text, numbers, and markers. However, the availability of such gazetteer information is limited. Fig. 3 shows a lot where shape similarity is obvious.

It is our goal to encourage local history researchers who georeference map entities for their research to publicly share this data on OSM and Wikimedia projects such as Commons (see e.g. [6]). OSM can be used to georeference historical routes and related items. Commons is helpful to display photographs with epigraphs and additional metadata. Georeferencing could involve textual resources, too. By linking such data and old maps, this can lead to groundtruth data for new digital map processing algorithms. Finally, it is crucial to validate the results, even if correct from an algorithmic perspective (see Fig. 2). OSM, Commons, etc. allow for a crowdsourcing approach here. In the future, we plan to build a public test collection.

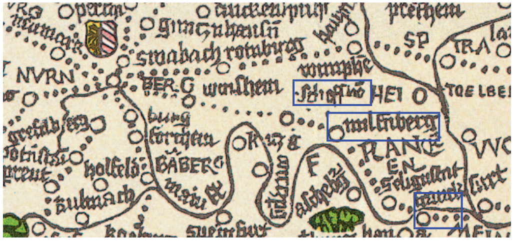
---

[1] e.g.: http://www.georeferencer.com/ and http://linkeddata.uni-muenster.de/georeferencer/georef.html (last visit: 15.3.17)

[2] e.g.: http://www.leo-bw.de/kartenbasierte-suche and http://kartenforum.slub-dresden.de/ (last visit: 15.3.17)

(a) Extract of Rogel's mile disk [7].

(b) Krettler's map [8] with the route between Tauberbischofsheim and Miltenberg near the village Steinbach.



(c) Map extract of Etzlaub's Landstraßen-Karte [9].

(d) A map showing the synonymy of *Wein Stras* and *Frankfürter Stras* on a route segment between Tauberbischofsheim and Miltenberg [10].

Fig. 1: The route from Tauberbischofsheim to Miltenberg and Frankfurt. Markers in magenta added for highlighting.

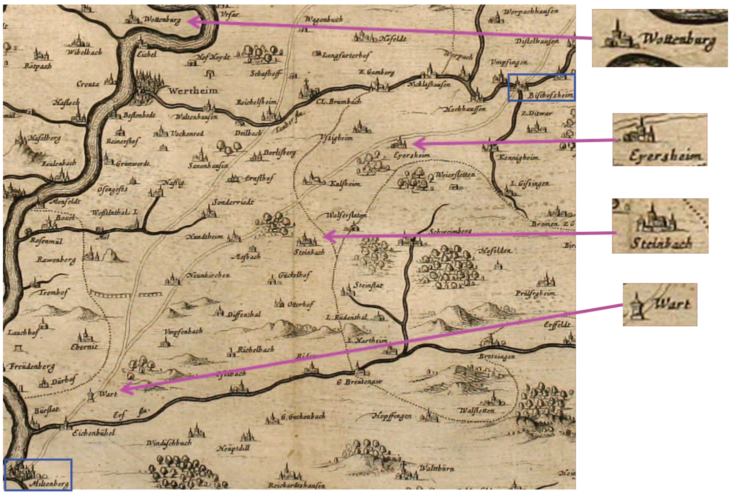
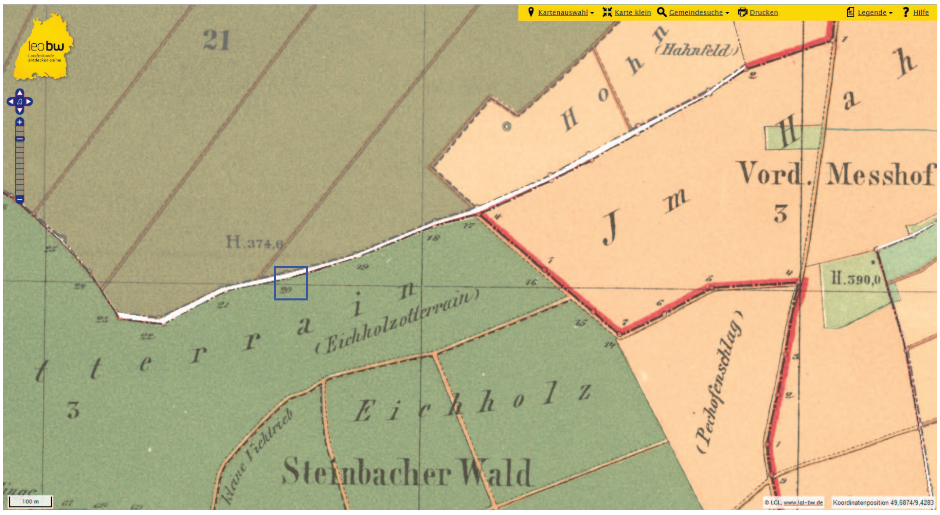


Fig. 2: The map [11] shows a route from top right to lower left: *Bischofsheim - Miltenberg*. It is only a schematic view: the actual route went south of, not north of *Eyersheim* and it went through *Diffenthal*.

Fig. 3: Four map segments showing the same part of a track but with a different naming. Left to right (LTR) [12], [13], [14], [15]: *die Weinstraße*, *die alte Straße*, *die alte oder Wein Straße*, *die Wein- oder alte Post-Straße*. The same cadastral unit is shown as can be supposed from the shape of the parcel and the position of the boundary stones surrounding it.

(a) Extract of Münzmeister Eberhard's map with the numbering of boundary stones [16].



(b) Ambiguous numbering of stones shown by the Historische Gemarkungspläne (Baden) [17].

Fig. 4: Stones where the numbering has changed. N. 17 (b) corresponds to N. 27 (a). It can thus be deduced that stone N. 20 (b) corresponds to N. 30 (a). Text labels such as *Pechofenschlag* (an alternative name for the parcel shown in Fig. 3) or *Eichholz* can help georeferencing the stones. Stone N. 1 occurs three times on the map (b).

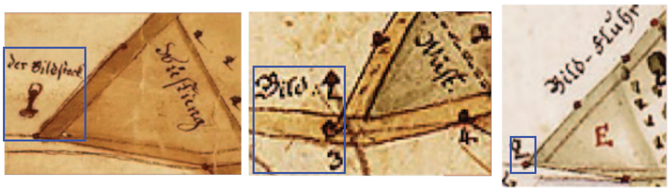


Fig. 5: Wayside shrine: *der Bildstock, Bild*, icon on the 3rd map (LTR: [13], [18], [19]). The name of the lot *Bild-Fluhr* addresses the prominence of the stone.

REFERENCES

[1] B. Budig, "Towards a Pipeline for Metadata Extraction from Historical Maps," in *Proceedings of EOM*, Luxembourg, 2016.

[2] D. Blank and A. Henrich, "A depth-first branch-and-bound algorithm for geocoding historic itinerary tables," in *GIR'16*. Burlingame, CA, USA: ACM, 2016, pp. 3:1–3:10.

[3] J. Weinman, "Toponym Recognition in Historical Maps by Gazetteer Alignment," in *ICDAR*. Washington, DC: IEEE, 2013, pp. 1044–1048.

[4] G. Minster, G. Moreau, and H. Saito, "Geolocation for Printed Maps Using Line Segment-Based SIFT-like Feature Matching," in *ISMARW '15*. Washington, DC, USA: IEEE, 2015, pp. 88–93.

[5] T. W. Shaw and P. Bajcsy, "Automation of digital historical map analyses," *Proc. SPIE*, vol. 7869, 2011.

[6] K. Bach, "How to share cultural heritage on Wikimedia Commons and OpenStreetMap," *www.opennordics.org/how-to-share-cultural-heritage-on-wikimedia-commons-and-openstreetmap/*, 2015.

[7] H. Rogel, https://commons.wikimedia.org/wiki/File:Augsburger_Meilenscheibe_Rogel_Hans.jpeg, public domain.

[8] Krettler, StAWt-R K Nr. 5956: http://www.landesarchiv-bw.de/plink/?f=7-319401.

[9] E. Etzlaub, https://commons.wikimedia.org/wiki/File:Rompilgerkarte_von_Erhard_Etzlaub_1501.jpg, public domain.

[10] StAWt-F K Nr. 16: http://www.landesarchiv-bw.de/plink/?f=7-307649.

[11] J. Janssonius, https://commons.wikimedia.org/wiki/File:Grafschaft_Wertheim_Janssonius_2.png; public domain.

[12] Staatsarchiv Wertheim StAWt-F K Nr. 217: http://www.landesarchiv-bw.de/plink/?f=7-307660.

[13] J. Kahl, StAWt R-K Nr. 712: http://www.landesarchiv-bw.de/plink/?f=7-320677.

[14] ——, StAWt-F K Nr. 232: http://www.landesarchiv-bw.de/plink/?f=7-307678.

[15] J. Goez, StAWt-R K in StAWt-R S 2 Nr. 599, Plan 1: http://www.landesarchiv-bw.de/plink/?f=7-313418.

[16] Münzmeister Eberhard, StAWt-F K Nr. 328: http://www.landesarchiv-bw.de/plink/?f=7-307963.

[17] Historische Gemarkungspläne (Baden), http://www.leo-bw.de/kartenbasierte-suche.

[18] Altenfelder, StAWt F-K Nr. 329: http://www.landesarchiv-bw.de/plink/?f=7-307964.

[19] J. Kahl, StAWt F-K Nr. 37: http://www.landesarchiv-bw.de/plink/?f=7-307680.

Metadata-Aware Map Processing

An automated metadata retrieval and management workflow for analysing old maps

H. Herold, H. Kollai, M. Neubert, G. Meinel
Monitoring of Settlement and Open Space Development
Institute of Ecological and Regional Development
Dresden, Germany
h.herold@ioer.de

R. Gruntzke, P. Winkler
ZIH - Center for Information Services and High
Performance Computing, TU Dresden
Dresden, Germany
r.grunzke@tu-dresden.de

Abstract—**Historical topographic maps are not only part of the cultural heritage but also a unique data source for land change research. Topographic maps preserve landscape and settlement patterns at certain points in time. Hence, old maps are a rich data source for tracking land use changes over long periods of time. However, metadata that is associated with the original maps and their processing is often not appropriately modeled and stored. This may complicate or even restrict the interpretation and usage of the results. Therefore, we develop and present a workflow for the automated enrichment and the ISO-conform storage and management of metadata for historical map processing and their usage for land use change monitoring.**

Keywords—Historical topographic maps; metadata; generic metadata management; image analysis; HPC; land use monitoring

I. INTRODUCTION

Historical topographic maps are often the only data source for tracking land use changes in a spatially explicit manner over long periods of time [1]. In the same way as other geospatial sources such as archival satellite images, topographic maps preserve states of landscapes and settlements at a certain point in time. The geoinformation that can be retrieved from these old maps could hence be used to “historize” existing land use and land cover databases [e.g., 2, 3]. In the last decades, libraries and national mapping agencies have begun to make large amounts of old maps digitally available over the internet.

To make the implicit map information available for large scale spatial analyses and change detection, advanced image analysis and pattern recognition algorithms have to be applied to the scanned map documents. The information extraction process comprises three major components: firstly, the scanning of the paper maps or − if available for the region of interest − the acquisition of digital images from a map repository [e.g., 4, 5], respectively; secondly, the georeferencing of the scanned maps (the reference is still only provided for the minority of available digital maps); and thirdly, the information or object extraction from the geo-referenced digital map images.

However, metadata that is associated with the original maps and their processing (see Fig. 1) is often not appropriately modeled and stored. This complicates or may even restrict the usage of the object extraction results. In this paper, we therefore develop and present a workflow for the automated retrieval and enrichment as well as the ISO-conform storage and management of metadata for historical map processing and for the usage of the historical geo-content in land use change monitoring.

II. METADATA RETRIEVAL AND MANAGEMENT

The automatic information acquisition from historical maps generates and necessitates a variety of metadata. Each of the three components generates a set of obligatory metadata entries. So far, there is no specific metadata standard of old maps. Therefore, we use the metadata standards for geodata of the Open Geospatial Consortium (OGC) and the Infrastructure for Spatial Information in the European Community (INSPIRE) [6]. To automatically enrich the bibliographic information (e.g., map sheet title and number, map series, date of publication), a web crawler was written to extract the descriptive metadata form the digital map archive [5] for each map file. Information on the geographic reference system and the corner coordinates of the georeferenced files are obtained by the interpretation of the GeoTIFF files using the Geospatial Data Abstraction Library (GDAL) [7]. The retrieved bounding box information is used to query OpenStreetMap (OSM) for a list of place names within the data frame through the Overpass API [8, 9]. All derived information is encoded in a simple Extensible Markup Language (XML) string using Python 3.5.2. To provide metadata as a base for interoperable geoprocessing workflows we chose to generate all metadata compliant to international standards for geospatial metadata, including ISO 19115:2014 - Geographic Information-Metadata [10]. The collected metadata is transformed into an ISO-conform XML structure (see Fig 2), using Extensible Stylesheet Language Transformation (XSLT) [11].

III. MAP CONTENT EXTRACTION AND PROVENANCE

The applied metadata standard also provides a suitable model to store provenance (or lineage [12, 13]) information, i.e. the description of processing steps, algorithms, and input/output data. This metadata is essential for both the interpretation of the retrieved map content and the granting of reproducibility and transparency. Therefore, we describe our current work on the integration of the provenance metadata retrieval into our map processing algorithms [2, 3, 14, 15, 16, 17] as well as into a High Performance Computing (HPC) environment for the extraction of settlement areas from up to 150 year old maps [18]. In conclusion, we discuss some synergies that arise from the combination of the retrieved metadata with data form other disciplines.

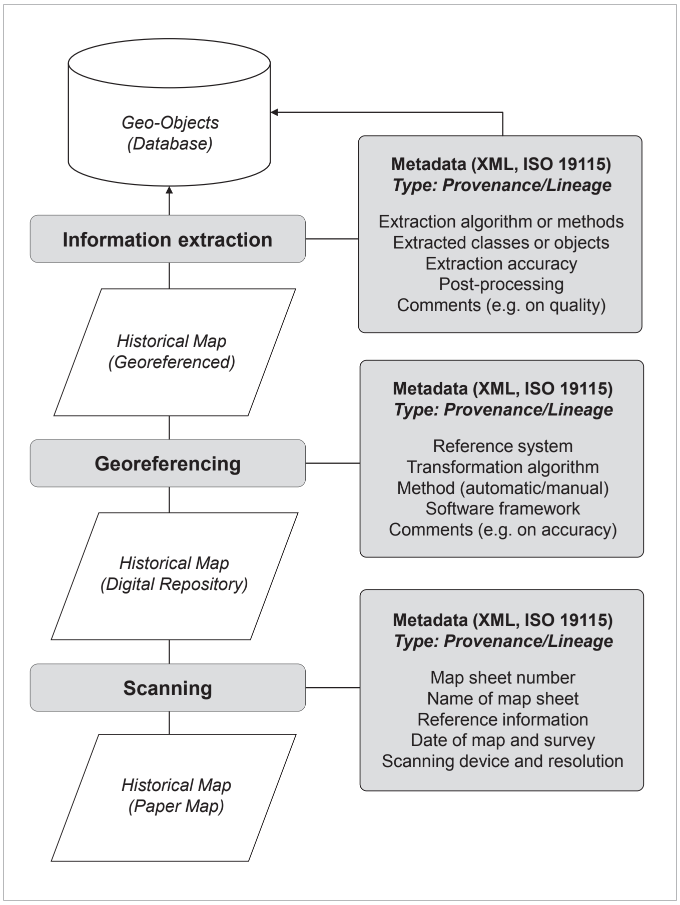


Fig. 1. Our metadata-aware map processing workflow and the extracted metadata, adapted and extended after [19].


```
    <gmd:role>
      <gmd:CI_RoleCode codeList="
      http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml#CI_RoleCode"
      codeListValue="pointOfContact">pointOfContact</gmd:CI_RoleCode>
    </gmd:role>
  </gmd:CI_ResponsibleParty>
</gmd:contact>
<gmd:dateStamp>
  <gco:DateTime>2017-02-07T14:18:08</gco:DateTime>
</gmd:dateStamp>
<gmd:metadataStandardName>
  <gco:CharacterString>ISO 19115:2003/19139</gco:CharacterString>
</gmd:metadataStandardName>
<gmd:metadataStandardVersion>
  <gco:CharacterString>1.0</gco:CharacterString>
</gmd:metadataStandardVersion>
<gmd:referenceSystemInfo>
  <gmd:MD_ReferenceSystem>
    <gmd:referenceSystemIdentifier>
      <gmd:RS_Identifier>
        <gmd:code>
          <gco:CharacterString>4314</gco:CharacterString>
        </gmd:code>
        <gmd:codeSpace>
          <gco:CharacterString>EPSG</gco:CharacterString>
        </gmd:codeSpace>
      </gmd:RS_Identifier>
    </gmd:referenceSystemIdentifier>
  </gmd:MD_ReferenceSystem>
</gmd:referenceSystemInfo>
<gmd:identificationInfo>
  <gmd:MD_DataIdentification>
    <gmd:citation>
      <gmd:CI_Citation>
        <gmd:title>
          <gco:CharacterString>Historische Topographische Karte 1:25000, Messtischblatt
          7113, Sufflenheim </gco:CharacterString>
        </gmd:title>
```

Fig. 2. Excerpt of the automatically generated metadata using an ISO-conform XML structure.

REFERENCES

[1] Haase, D., Walz, U., Neubert, M., and Rosenberg, M. (2007): Changes to Central European landscapes – Analysing historical maps to approach current environmental issues, examples from Saxony, Central Germany. Land Use Policy 24 (2007), p. 248-263.

[2] Meinel, G., Hecht, R., Herold, H., (2009). Analyzing building stock using topographic maps and GIS. Building Research & Information 37(5-6), pp. 468-482.

[3] Herold, H., Meinel, G., Hecht, R., and Csaplovics, E. (2012). A GEOBIA Approach to Map Interpretation - Multitemporal Building Footprint Retrieval for High Resolution Monitoring of Spatial Urban Dynamics. In Proceedings of the 4th GEOBIA, Rio de Janeiro, Brazil pp. 252–256.

[4] OldMapsOnline – www.oldmapsonline.org, The search engine for historical maps.

[5] Deutsche Fotothek Kartenforum, http://www.deutschefotothek.de/kartenforum; Virtual Map Forum 2.0, http://kartenforum.slub-dresden.de.

[6] Commission of the European Communities, EC (2007). Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). *Official Journal of the European Union*, L108/1–14.

[7] Warmerdam, F. (2008). The geospatial data abstraction library. Open source approaches in spatial data handling. pp. 87-104.

[8] Overpass API public instance, http://overpass-api.de/

[9] Olbricht, R. M. (2015). Data Retrieval for Small Spatial Regions in OpenStreetMap. OpenStreetMap in GIScience. pp. 101-122.

[10] ISO/TC211 (2014). ISO 19115-1:2014 Geographic information – Metadata – Part 1: Fundamentals. Geneva: International Standards Organisation.

[11] Mize, J., & Robertson, C. F. (2009). A solution to metadata: Using XML transformations to automate metadata. CEANS 2009, MTS/IEEE Biloxi-Marine Technology for Our Future: Global and Local Challenge, pp. 1-7.

[12] Di, L., Shao, Y., & Kang, L. (2013). Implementation of Geospatial Data Provenance in a Web Service Workflow Environment With ISO 19115 and ISO 19115-2 Lineage Model. IEEE Transactions on Geoscience and Remote Sensing, 51(11), pp. 5082–5089.

[13] Henzen, C., Mäs, S., & Bernard, L. (2013). Provenance information in geodata infrastructures. Geographic Information Science at the Heart of Europe. pp. 133-151.

[14] Herold, H., Roehm, P., Hecht, R., and Meinel, G. (2011). Automatically Georeferenced Maps as a Source for High Resolution Urban Growth Analyses. In Proceedings of the ICA 25th International Cartographic Conference, Paris, France, pp. 1–5.

[15] Hecht, Robert; Herold, Hendrik; Meinel, Gotthard; Buchroithner, Manfred F. (2013). Automatic Derivation of Urban Structure Types from Topographic Maps by Means of Image Analysis and Machine Learning. In: Proceedings of the 26th International Cartographic Conference.

[16] Muhs, S.; Herold, H.; Meinel, G.; Burkhardt, D.; Kretschmer, O. (2016). Automatic delineation of built-up area at urban block level from topographic maps In: Computers, Environment and Urban Systems, 58, pp. 71-84

[17] Krüger, Tobias; Meinel, Gotthard; Schumacher, Ulrich (2013). Land-use monitoring by topographic data analysis. In: Cartography and Geographic Information Science 40 (2013) 3, pp.220-228.

[18] Schemala, D., Schlesinger, D., Winkler, P., Herold, H., & Meinel, G. (2016). Semantic segmentation of settlement patterns in gray-scale map images using RF and CRF within an HPC environment. *GEOBIA 2016: Solutions and Synergies.*, September 2016, University of Twente Faculty of Geo-Information and Earth Observation (ITC).

[19] R. Grunzke, V. Hartmann, T. Jejkal, A. Prabhune, S. Herres-Pawlis, A. Hoffmann, A. Deicke, T. Schrade, R. Stotzka and W. E. Nagel (2016). Towards a Metadata-driven Multi-community Research Data Management Service. 8th International Workshop on Science Gateways (IWSG), pp. 1-8.

Open Historical Data Map - a prototype

Thomas Schwotzer
HTW Berlin
University of Applied Sciences
Germany, Berlin +49 30 5019-2604
Email: thomas.schwotzer@htw-berlin.de

***Abstract*—Open Street Map (OSM) is one example of free and open map service. Volunteers chart geographic objects and enter those information into the systems. OHDM is going to adopt that idea but for historic maps and data. OHDM will offer historic data as Web Map Service and interfaces to down- and upload historic spatial information. A technical prototype can be presented during the workshop. That paper describes the used components and some major obstacles which had to be overcome.**

I. OHDM DATA

The OHDM data model is already published [OHDM16] and will not be discussed here in detail. In general, OHDM stores geographic objects, geometries and their temporal relations. Quite often, same geometry describes different objects when time passes: Streets are renamed for instance and usage of buildings can change. Same objects can also change geometry for instance if a building is expanded.

PostGIS was our choice for the OHDM data base. Postgres is free and open software. The PostGIS plugin makes it a fast spatial database. We use a Debian server with 40 Terrabyte hard drive, 32 kernels and 30 GByte Memory. Planet wide OSM data from Jan, 19th 2017 (3.7 billion nodes, 250 million ways) require less than 1 Terrabyte. That hardware is sufficient for that amount of data.

II. WEB MAP SERVICE

Web Map Service is the OGC standard for offering digital maps [WMS]. We chose GeoServer[1] as WMS implementation. It is well documented any easy to administrate. These are significant criteria in a university environment in which students often enter and leave the project.

We produce rendering tables out of the OHDM core database, see figure below. Each line of those tables contains object description, its geometry and validity. The renamed street would appear in two lines which differ in description but not geometry. That data model is apparently redundant but optimized for rendering. With a spatial index on those tables, Geoserver is able to render tables with about 10 million lines under 10 seconds. Geoserver is an appropriate technology for the project.

There is a pleasant side effect: We can offer about 100 WMS layers each containing an OSM feature. For instance, there is a layer called `railway_lines` containing all railways. Other applications can use those layers for their own purpose.

It took our server about 5 hours to produce rendering tables covering whole Germany. We calculate that production of worldwide rendering tables are done within a day on our hardware.

III. OSM DATA IMPORT / UPDATE

OHDM requires data. OSM with its huge amount of data is a perfect choice to test our hardware and software components.

The current `planet.osm` file has a size of about 850 GByte. Those data are imported into an *intermediate* database what we call the *OSM import*, see figure below.

In a second step (*OSM extraction*), semantic information (like names), geometries and validity are extracted from OSM data and imported into the OHDM database. An annual update is planned. Changes in OSM are saved with the OHDM database which remember that history. We expect an annual growth rate of less than a Gigabyte each year.

That initial OSM importer was written in Java which is the usual language in our teaching programs. It communicates with the OHDM PostGIS database via JDBC and issues `INSERT` statements for each OSM entity. The whole import takes about 10 days on our hardware. It is sufficient for a prototype when having in mind that this process will only be performed once a year. Nevertheless, is worth considering alternatives.

`PSQL` is a PostGIS tool which can process these statement a fifth faster than JDBC. There is another alternative, though.

PostGIS offers another structure and API supporting fast import and export of tables (`pgdump`) for backup and recovery. That API is documented and could be used for our project. Some first measurements indicate that using that API would require a hundredth (!) of time. We are optimistic to import worldwide OSM data within 15 hours.

IV. CONCLUSION AND OUTLOOK

PostGIS is a stable and fast spatial database server that fits perfectly to our requirements.

Creating optimized rendering tables and creating spatial indexes are basis of a fast rendering process. Geoserver is a reliable and well-documented WMS server that is fast enough for our needs.

Importing other sources than OSM into OHDM is most relevant now. We are in close contact to libraries especially in Berlin. Our next task is to incorporate historic data into OHDM and to offer it to interested parties.

---

[1] www.geoserver.org

REFERENCES

[WMS] Jeff de la Beaujardiere (editor): OpenGIS Web Map Server Implementation Specification 1.3.0, 2006 http://www.opengeospatial.org/standards/wms

[OHDM16] Thomas Schwotzer: Open Historic Data Map (work in progress) First International Workshop on Exploring Old Maps, June 2016, Luxembourg https://www.researchgate.net/publication/303818952_Open_Historical_Data_Map_OHDM_-_work_in_progress

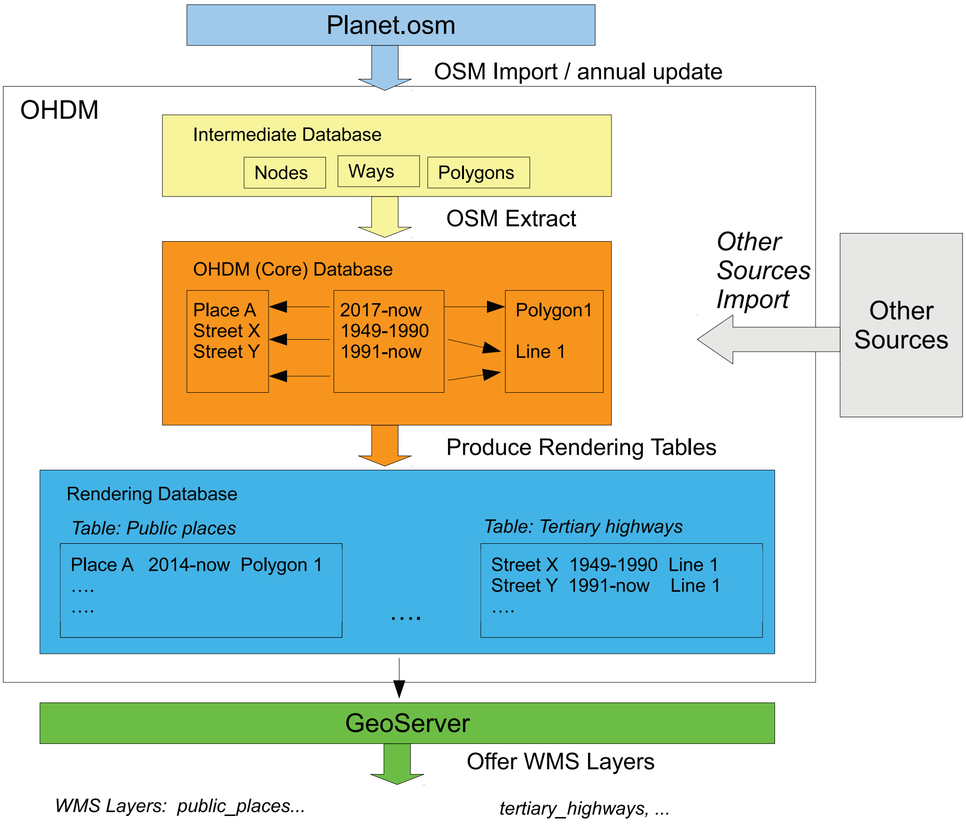


Fig. 1. OHDM Components and processes

Deep Learning for Place Name OCR in Early Maps

Winfried Höhn
Faculty of Science, Technology and Communication
University of Luxembourg
Email: winfried.hoehn@uni.lu

I. INTRODUCTION

Original early maps are usually only accessible for a small group of researchers and librarians because they are old and fragile. However, they are a valuable knowledge source for historical research. With the availability of early maps in digital libraries (e.g. http://www.oldmapsonline.org/) people from all over the world can now explore these artefacts of our cultural heritage. However, the digitization solely generates images of the artefacts without any access to the semantics of the documents. For most digital libraries the available metadata include only information about the map, e.g. author, title, size, creation date. Unfortunately, there is only little information about the data contained in the map. Thus this information stays unsearchable even when it is available online. Therefore, tools that can automate the extraction of information from early maps are of interest.

In the recent years deep learning had a huge impact in many research areas, especially in the computer vision area. Some successes, which motivated this work, are the use for handwritten text recognition [1] and OCR [2]. Deep networks have also been successfully used in scene text recognition [3], [4], which consists of the two tasks of localizing text in natural scenes and creating a transcript of the text. We will here focus on the text recognition part for already localized place labels in early maps using deep learning.

II. PLACE LABEL OCR FOR EARLY MAPS

The character recognition for place labels we are proposing is constructed similar to Ocropy[1] using a bidirectional recurrent neural network (RNN) [5], which makes a prediction for each frame of the feature sequence (column of the image) and a Connectionist Temporal Classification (CTC) loss layer [6] that is used to learn the prediction of a character sequence. In contrast to Ocropy we are not feeding the input image directly to the RNN but using additional alternating convolutional and max pooling layers, which extract a feature sequence from the input image. The output of the convolutional layers is then processed by the RNN. Also data augmentation as described in [7] is used on the input images to reduce overfitting. The whole system was built in Python with TensorFlow [8] as Deep Learning Framework.

We evaluated our approach on a early map with clear writing and diplomatic transcripts of the labels (see example labels in Fig. 2). The map contained 929 labels of which we used 260 as test set. On this set we were able to reach CERs of 10% while only using 200 labels for training. We were also able to further reduce the CER to 6% by first training the network on artificially generated labels (see Fig. 3), which used randomly selected place names[2] rendered with blackletter fonts on a background image extracted from the map. The network trained with only artificial data had a CER of ca. 90%, but when it was fine-tuned with the data from the training set, it converged in only a few steps to a CER of 6%.

As a second test case we used data available in Recogito [9], [10], a tool built specifically for place name annotations in texts and maps (see Fig. 1 and 4). We tested our approach on a large set of image annotations, consisting of 2322 annotations from 6 portolan maps. Since the labels in the Recogito set were not created for OCR training but to annotate the places on the map, they have some shortcomings for this task. For example, the bounding boxes do not always including the ascenders and descenders and in some of the label transcriptions, abbreviations are expanded or the case is adjusted to modern spelling. The rhumb lines in these maps, often also crossing the place names (see Fig. 1) make this an even harder OCR problem than for other early maps. We split this label set in 2000 labels for training and 322 labels for testing. The proposed architecture then reaches on the test set a character error rate (CER) of 19.6%. This results in a correct recognition of over a fourth and only one error in a third of the place labels (see Fig. 5). When trained on only 1000 of the labels, the character error rate goes up to 27.5%.

III. DISCUSSION

The presented results show that even for early maps, where it is often doubted [11], OCR with a reasonable quality is within reach, although the results are still inferior to OCR on modern texts. For scanning books, a CER of 10% would not be considered sufficient, but here we have shown that, even with a CER of 20%, over a quarter of the place labels are recognized without a single error. To further improve the quality of the recognized labels they could be matched against a gazetteer, which would also allow localizing the map and narrowing down the possibilities for yet unmatched labels.

REFERENCES

[1] A. Graves, M. Liwicki, S. Fernndez, R. Bertolami, H. Bunke, and J. Schmidhuber, "A novel connectionist system for unconstrained handwriting recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 5, pp. 855–868, 2009.

---

[1] https://github.com/tmbdev/ocropy

[2] http://www.ling.uni-potsdam.de/ kolb/DE-Ortsnamen.txt

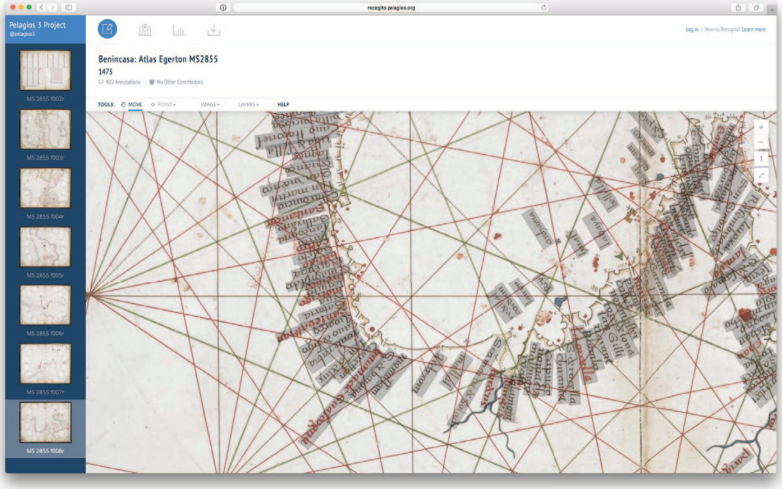


Fig. 1. Sample portolan map with annotated place markers

Fig. 2. Examples of labels extracted from the map

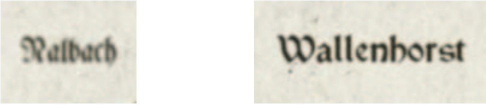


Fig. 3. Examples of generated labels

[2] T. M. Breuel, A. Ul-Hasan, M. A. Al-Azawi, and F. Shafait, "High-performance ocr for printed english and fraktur using lstm networks," in *Proceedings of the 2013 12th International Conference on Document Analysis and Recognition*, ser. ICDAR '13. IEEE Computer Society, 2013, pp. 683–687.

[3] B. Su and S. Lu, "Accurate scene text recognition based on recurrent neural network," in *Asian Conference on Computer Vision*. Springer, 2014, pp. 35–48.

[4] B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, no. 99, 2016.

[5] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[6] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks," in *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006, pp. 369–376.

[7] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Proceedings of the Seventh International Conference on Document Analysis and Recognition - Volume 2*, ser. ICDAR '03. IEEE Computer Society, 2003, pp. 958–962.

[8] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning," in *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI).*, 2016.

[9] R. Simon, E. Barker, L. Isaksen, and P. de Soto Cañamares, "Linking Early Geospatial Documents, One Place at a Time: Annotation of Geographic Documents with Recogito," *e-Perimetron*, vol. 10, no. 2, pp. 49–59, 2015.

[10] Recogito. Recogito. [Online]. Available: http://commons.pelagios.org/about/

[11] R. Simon, P. Pilgerstorfer, L. Isaksen, and E. Barker, "Towards semi-automatic annotation of toponyms on old maps," *e-Perimetron*, vol. 9, no. 3, pp. 105–112, 2014.

Transcription:
Venetia
Comment:
Not 100% certain about spelling though.

Fig. 4. Recogito Image Annotation Editor. Source http://pelagios.org/recogito/static/documentation/index.html

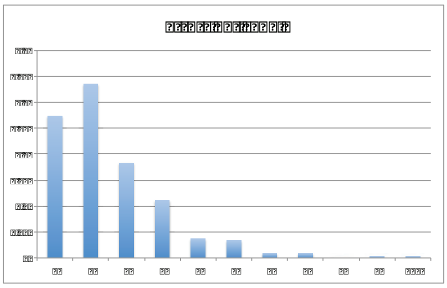


Fig. 5. Error distribution per label. The test label lengths range from 3 to 26 characters with an average length of 8 and a median length of 7 characters.

Exploring Local Geography for Toponym Matching

Benedikt Budig
Chair for Computer Science I, University of Würzburg
`benedikt.budig@uni-wuerzburg.de`

I. INTRODUCTION

Historical maps are an important source of information for scholars of various disciplines. In this report, we identify a problem related to historical maps, consider it from different perspectives and formulate it as an optimization problem. We advocate this approach of mathematic modeling and optimization and showcase it on the problem of identifying toponyms.

In order to make the most use of historical maps that have already been scanned, we need to be able to query them as a *structured spatiotemporal source* [1] and *deeply georeference* [2] their contents. One particular task towards this goal is to match settlements depicted in a historical map to the corresponding modern places.

II. STATE OF THE ART

A straightforward approach to this problem is to "read" the place labels in the historical map (using OCR or crowdsourcing) and then compare them to modern place names from a gazetteer. The applicability of OCR techniques to historical maps is being actively studied [3]–[6]. Furthermore, Yu et al. [7] match OCR results from large-scale maps of different time periods to a large (general) dictionary. Due to possibly weak OCR results and the natural change of toponyms over the centuries, it can however be necessary to combine the textual with geographic information.

Weinman [8] presents such an approach based on probabilistic modeling. He combines geometric considerations with OCR results to find a maximum likelihood explanation for matching toponyms in a Bayesian framework. This model involves finding a most likely affine transformation between the historical and a modern map. Weinman successfully applies this technique to a set of maps of the 19th century.

In contrast, we aim to solve this problem for considerably older (i.e. 16th- to 18th-century) maps, which are often heavily distorted and not described by a reasonable projection. For such maps, a meaningful affine transformation to a modern map does not exist. Still, local geography—that is, the relative position of map elements to other elements nearby—is often correct (for an example, see Figure 1). Based on this observation, we propose a problem formulation that avoids global transformations, but rather combines local geographic information with toponym similarities.

III. MODELING THE PROBLEM

We want to exploit both textual and geometric similarities. Our target set of historical maps for experimentation will be the *Franconica* collection, a set of roughly 700 maps created between the 16th and early 19th century. The maps cover the area of Lower Franconia, thus containing German toponyms.

String Metrics for Toponyms: In many cases, toponyms change only slightly over the centuries: particularly spelling details differ, while the names remain phonetically similar. There is a large body of research on distance measures between strings. A comparison specifically for (modern) toponym matching has been presented by Recchia and Louwerse [9], who suggest using skip-grams for matching German toponyms. However, when reproducing their results in our setting with historical toponyms, we found distance measures based on the Needleman-Wunsch algorithm [10] to be more suitable. Using specific weights for characters that are likely to have changed, this approach can also be adapted using phonetic considerations (for example those from the Kölner Phonetik [11] or Soundex [12]).

Geometric Considerations: Since historical maps can be heavily distorted, we only want to consider local geometry. Our basic assumption is that when two places are located near each other in the historical map, their modern equivalents should be near each other in a modern map. Thus, we want our problem formulation to favor matching pairs of places that are close both on historical and modern map. This effectively results in a web of local neighborhood relations, which can be used to distinguish between places that have similar names and to match places if textual information is weak.

Optimization Problem: We combine these textual and geometric considerations into a single optimization problem. We introduce a complete bipartite graph on the sets of historical and modern toponyms as vertices and ask for a *matching*, that is, each historical toponym is assigned to at most one modern toponym and vice versa. These assignments are weighted according to the string similarity of the place names. Further, we introduce a bonus for pairs of assignments if and only if the involved historical and modern places are close-by on their respective maps. The objective is to find a matching that maximizes the sum of these weights and bonuses. For an exact statement, see Figure 2. Unfortunately, the given problem is NP-hard (see Figure 3 for a sketch of the proof).

IV. CONCLUDING REMARKS

Our model may at this point be impractical for real-world instances, but the approach taken makes sense: Formal problem statements are necessary both for finding good heuristics and for solid experimentation. An important question we are currently investigating is how to obtain suitable weights and bonuses (e.g. how to learn [13] them from sample maps).

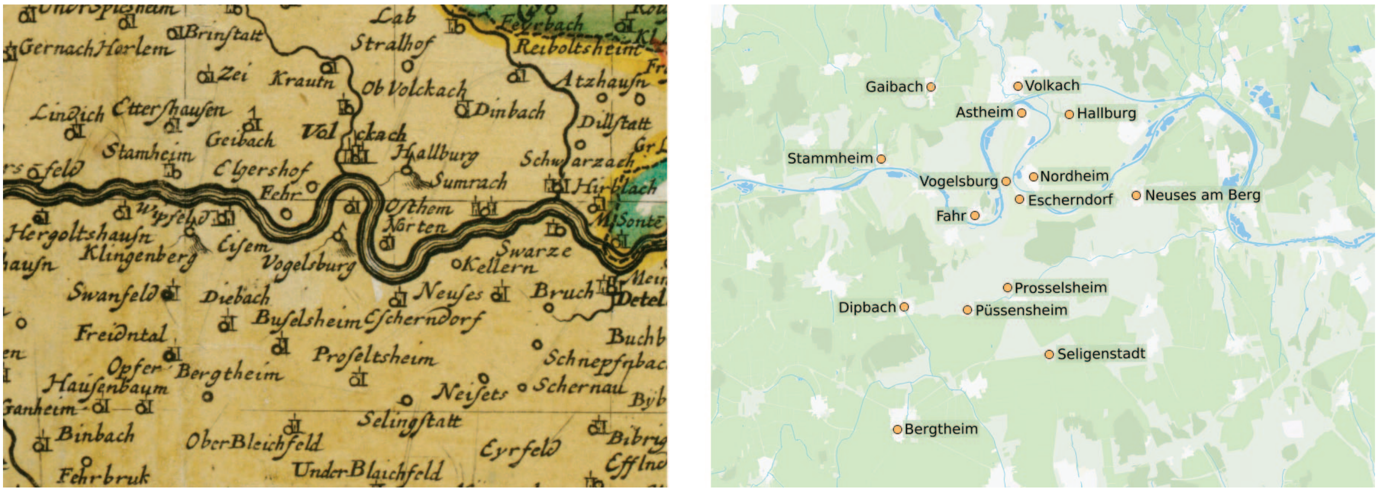
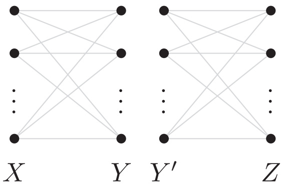


Fig. 1. Toponyms in a map from 1676 (left) and some modern toponyms in the same area (right, place names from GeoNames [14], base map from [15]). Note that corresponding toponyms are either identical or very similar in spelling, for example Selingftatt/Seligenstadt, Volckach/Volkach and Fehr/Fahr. In the historical map, the relative positions of the places to each other are also mostly correct, despite the considerable global distortion.

$$
\begin{aligned}
\text{Maximize} \quad & \sum_{\substack{g\in G\\ b\in B}} c_{g,b}\cdot x_{g,b} + \sum_{\substack{g,g'\in G\\ b,b'\in B}} k_{g,g',b,b'}\cdot y_{g,g',b,b'} & \\
\text{subject to} \quad & \sum_{g\in G} x_{g,b} \le 1 & \forall b\in B\\
& \sum_{b\in B} x_{g,b} \le 1 & \forall g\in G\\
& y_{g,g',b,b'} \le x_{g,b} & \forall g,g'\in G,\ \forall b,b'\in B\\
& y_{g,g',b,b'} \le x_{g',b'} & \forall g,g'\in G,\ \forall b,b'\in B\\
& x_{g,b}\in\{0,1\} & \forall g\in G,\ \forall b\in B\\
& y_{g,g',b,b'}\in\{0,1\} & \forall g,g'\in G,\ \forall b,b'\in B
\end{aligned}
$$

Fig. 2. ILP formulation for our toponym assignment problem. For each pair (g,b) of a historical toponym $g\in G$ and a modern toponym $b\in B$, the decision variable $x_{g,b}$ indicates whether g and b are matched. They are weighted by $c_{g,b}$, which corresponds to the string similarity between b and g. The variables $y_{g,g',b,b'}$ are nonzero if and only if g is matched to b and g' to b' as well. If both g and g' are geographically close and b and b' as well in the respective maps, the corresponding weight $k_{g,g',b,b'}$ is set to a positive value (and thus gives a bonus).



$$
\begin{aligned}
c_{x,y} &= 0 \qquad \forall x\in X, y\in Y\\
c_{y',z} &= 0 \qquad \forall y'\in Y', z\in Z\\
k_{x,y,y',z} &= \begin{cases} 1 & \text{if } (x,y,z)\in T\\ 0 & \text{otherwise}\end{cases}
\end{aligned}
$$

Fig. 3. Sketch of our NP-Hardness proof for the toponym assignment problem. We reduce from 3-dimensional matching [16], which is defined as follows: Let X, Y, and Z be disjoint sets and let $T\subseteq X\times Y\times Z$; is there a subset of T of size at least k that covers no element from $X\cup Y\cup Z$ more than once? We can model this problem as a toponym assignment problem by duplicating Y and using the construction and weights shown above. Here $X\cup Z$ and $Y\cup Y'$ are used as the historical and the modern toponyms. The size of the resulting toponym matching is equal to the size of the optimal 3-dimensional matching for the given instance.

REFERENCES

[1] Y.-Y. Chiang, "Querying historical maps as a unified, structured, and linked spatiotemporal source," in *Proc. 23rd ACM SIGSPATIAL Int. Conf. on Adv. in Geographic Information Systems*, 2015, pp. 270–273.

[2] B. Budig, T. C. van Dijk, F. Feitsch, and M. G. Arteaga, "Polygon consensus: Smart crowdsourcing for extracting building footprints from historical maps," in *Proc. of the 24th ACM SIGSPATIAL Int. Conference on Advances in Geographic Information Systems*, 2016, pp. 66:1–66:4.

[3] Y.-Y. Chiang and C. A. Knoblock, "Recognizing text in raster maps," *GeoInformatica*, vol. 19, no. 1, pp. 1–27, 2015.

[4] Y.-Y. Chiang, S. Leyk, N. Honarvar Nazari, S. Moghaddam, and T. X. Tan, "Assessing the impact of graphical quality on automatic text recognition in digital maps," *Computers & Geosciences*, vol. 93, pp. 21–35, 2016.

[5] J. Pouderoux, J.-C. Gonzato, A. Pereira, and P. Guitton, "Toponym recognition in scanned color topographic maps," in *Proceedings of the 9th International Conference on Document Analysis and Recognition (ICDAR '07)*, 2007, pp. 531–535.

[6] W. Höhn, "Deep learning for place name OCR in early maps," in *Proc. of the 2nd Int. Workshop on Exploring Old Maps (EOM '17)*, 2016.

[7] R. Yu, L. Zexuan, and Y.-Y. Chiang, "Recognizing text in historical maps using maps from multiple time periods," in *Proceedings of the 23rd International Conference on Pattern Recognition (ICPR '16)*, 2016.

[8] J. Weinman, "Toponym recognition in historical maps by gazetteer alignment," in *Proceedings of the 12th International Conference on Document Analysis and Recognition (ICDAR '13)*, 2013, pp. 1044–1048.

[9] G. Recchia and M. Louwerse, "A comparison of string similarity measures for toponym matching," in *Proceedings of the first ACM SIGSPATIAL International Workshop on Computational Models of Place (COMP '13)*, 2013, pp. 54:54–54:61.

[10] S. B. Needleman and C. D. Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins," *Journal of Molecular Biology*, vol. 48, no. 3, pp. 443–453, 1970.

[11] H. J. Postel, "Die Kölner Phonetik. Ein Verfahren zur Identifizierung von Personennamen auf der Grundlage der Gestaltanalyse," *IBM-Nachrichten*, vol. 19, pp. 925–931, 1969.

[12] M. Odell and R. Russell, "The soundex coding system," 1918, US Patent 1261167.

[13] E. S. Ristad and P. N. Yianilos, "Learning string-edit distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 5, pp. 522–532, 1998.

[14] The GeoNames geographical database, http://www.geonames.org/.

[15] OpenStreetMap, http://www.openstreetmap.org.

[16] R. M. Karp, "Reducibility among combinatorial problems," in *Proceedings of a symposium on the Complexity of Computer Computations*, R. E. Miller, J. W. Thatcher, and J. D. Bohlinger, Eds., 1972, pp. 85–103.

Digitising Schematic Maps

Recreating or reinventing history?

Maxwell J Roberts
Department of Psychology, University of Essex
Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK
mjr@essex.ac.uk

***Abstract*—Historic schematic maps are well-suited to digitisation using vector graphic software, and this can assist in the interpretation and understanding of unpublished or lost designs. Provided that the potential for anachronisms is fully understood, it is argued that such exercises can provide valuable insights into this domain.**

Keywords—schematic maps; vector graphics; design investigations

I. INTRODUCTION

A researcher commencing investigation into the history of schematic maps will soon discover that the available published versions comprise a very incomplete record of the development of the genre. Countless prototypes remain unpublished, and there are many promising releases that might have benefitted considerably from further development, but were withdrawn early before refinements could be applied. Worst of all, for some published designs, all printed copies of have been lost so that, for example, only tantalising appearances in photographs demonstrate their existence.

Mapping is an inherently visual medium. For example, the undoubted success of Henry Beck's 1933 London Underground schematic [2] must in part be due to its visual organisation and impact, providing a uniquely effective graphical device for comprehending and understanding the structure of the network [4]. In the absence of pictorial representations, only a partial evaluation is therefore possible of the viability and potential of the unimplemented or lost elements of the historical jigsaw.

Schematic maps are usually straightforward to create using vector graphics software. Hence it is possible to implement and evaluate various lost or unpublished designs from drawings and photographs, and also to reconstruct published versions with a view to investigating the consequences of alternative design decisions that might have been applied to them.

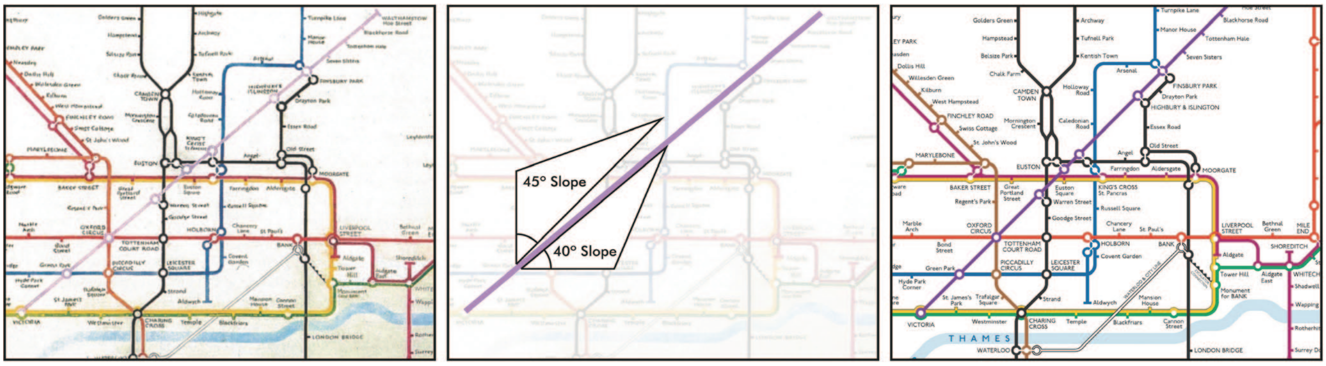
II. BENEFITS OF DIGITAL RECREATION

In general, digitising a map forces an attention to detail to its design and construction that would not usually be attained in the normal course of merely inspecting it visually. Prototype maps are sometimes rejected for political rather than practical reasons, and an attempt at implementing these in the form of their intended end-products can enable an evaluation of their viability that would not otherwise be possible. For example, the demise of Henry Beck's London Transport commission co-occurs with a deteriorating relationship, in the late 1950s, between himself and the London Transport Publicity Officer [2, 4]. Fig. 1 shows a section from one of Beck's rejected prototypes, and an attempt to implement this has revealed a number of problems that hitherto have not been reported.

Going beyond drawings, it is possible to identify written specifications and proposals for designs that, for various reasons, were never implemented. Fig. 2 shows a section from George Salomon's New York City Subway diagram of 1958, but his original suggestion was for a colour-coded design that might have had considerably more utility [5], and this can easily be investigated. Going further, it is possible to explore ideas and concepts that might have been rejected prematurely, and are difficult to evaluate because of the co-existence of less satisfactory features. One example of this is Henry Beck's experiment with 60º diagonals as a means to save space [2, 4]. The design only survived for one issue (Fig. 3) but its potential can be explored via speculative reworkings.

III. PITFALLS OF DIGITAL RECREATION

Digitising a historic map will result in many anachronisms and it is essential that these are understood by observers. Some can be minimised with close attention to the original. For example, hand-implemented curves are subtle and can be very difficult (but not impossible) to mimic via Bézier vectors. Much harder to address is the inevitable variability inherent in

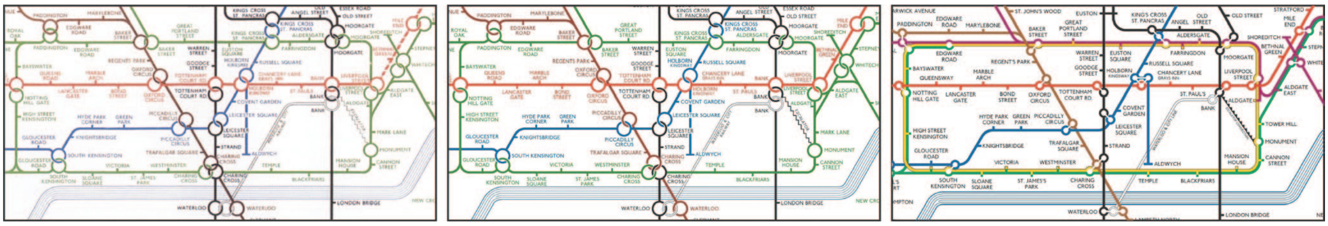


(a) *Scan from original drawing* (b) *Angle for the Victoria Line* (c) *Vector implementation by the author*

1. Henry Beck continued designing maps in the early 1960s even after London Transport ceased to employ him. Amongst his most ingenious proposals was a straight line trajectory for the under-construction Victoria Line (lilac) [2]. However, an attempt to digitise his drawing revealed that the diagonal angle of this was 40º rather than the standard one of 45º. Correcting this 'narrows' central London, with adverse consequences for station name placement. Beck's proposal was effectively unpublishable without comprehensive reworking, or else the breaking of standardised design rules [4].

(a) *Scan from published version* (b) *Vector implementation by the author* (c) *Adaptation by the author*

2. George Salomon designed a diagrammatic New York City Subway map that was first published in 1958. This was typeset using an impressive variety of faces [5] but digital versions of all of these are available, enabling a very good approximation to be created. This map is of interest because the published version used obsolete colour-coding based upon historic railway company ownership. Salomon himself proposed separately colour-coding the major north-south trunk lines in Manhattan [4, 5]. His specification was sufficiently detailed for the digitised map to be easily reconfigured to match this, thus enabling the utility of Salomon's proposal to be appreciated, compared with the version issued to New Yorkers instead.

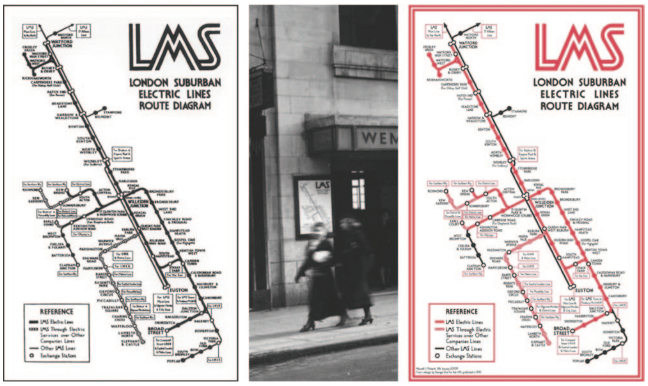


(a) *Scan from published version* (b) *Vector implementation by the author* (c) *Adaptation by the author*

3. In 1940, a completely new London Underground map was issued, designed by Henry Beck and using 60º diagonals rather than 45º. The motivation was an attempt to save horizontal space so that extensions to the east could be accommodated [2, 4]. Unfortunately, the utility of this innovation is difficult to evaluate because of other ones simultaneously applied, including large-ring interchange circles with multiply named stations in their own line colours. This increases visual congestion at already-complex regions of the map. All of these were quickly abandoned. The 60º diagonal innovation can be investigated further by combining this with the station naming and interchange circle conventions subsequently adopted in that decade. The result is purely speculative, but at the very least indicates that a steeper angle of diagonal is a viable solution to accommodate an extensive network.

the design and print process, such as imprecise stroke width and unreliable print registration. A digital reconstruction of a map, ink-jet printed, will be more regular and exact than could have been possible even thirty years ago – the first digital London Underground map was issued in 1987 [3].

The major challenge in digitising a map is implementing the lettering. Hand-drawn characters vary in compactness, form and spacing (Fig. 4), and even if this approximates a known typeface, substituting this for a computer font is not entirely satisfactory, with a far cleaner result than would have been possible historically. Even when a design has been typeset, and a computer version of the exact same typeface is available, obsolete print technology results in far more variability than digital reproduction today. For a published map, the extreme solution would be to trace every single letter individually, but it may be argued that this imprecision would look out of place in the context of the exactness of other aspects of the digitised design, such as perfect straight lines of constant stroke width. For implementing a prototype map, with no print version at all for guidance, introducing artificial speculative variability into the lettering would be even harder to justify.



(a) *Scan of reproduction* (b) *In situ* (c) *Vector implementation*

4. George Dow created a dramatic map for the London area electric services of the LMS railway company. It was released in 1935, but all printed copies have been lost. A monochrome reproduction in a contemporary railway magazine flags its existence [1] and a photograph at the entrance to Wembley station proves that, at the very least, prototypes were made available to the general public. The original was hand-lettered, with variable shapes and spacing presenting difficulties for recreation. Here, a font was devised based upon recurring features of the letterforms. The outcome shows the power of the original, but it is important to emphasise that the lettering on the map is representative rather than authentic.

Overall, digital recreations of historic maps can only be representative rather than authentic, but provided that people are fully aware of the necessary compromises and resulting anachronisms, then this should not be regarded as misleading *per se*. It is suggested here that the potential insights outweigh the costs. Furthermore, it could be argued that the vagaries of the historic production process mask the ingenuity of the original designs themselves, and digital recreation, even of published well-known easily-available versions, enables them to stand beside and withstand comparison to modern-day equivalents in a way that would otherwise not be possible.

REFERENCES

1. A. Dow, Telling the Passenger Where to Get Off. Harrow Weald, London: Capital Transport Publishing, 2005.
2. K. Garland, Mr Beck's Underground Map. Harrow Weald, London: Capital Transport Publishing, 1994.
3. M.J. Roberts, Underground Maps After Beck. Harrow Weald, London: Capital Transport Publishing, 2005.
4. M.J. Roberts, Underground Maps Unravelled: Explorations in Information Design. Wivenhoe, Essex: Published by the author, 2012.
5. P. Shaw, Helvetica and the New York Subway System. Cambridge, MA: MIT Press, 2011.

From Many User-Contributed Polygons to One Building Footprint

Fabian Feitsch*
Chair for Computer Science I, University of Wuerzburg
fabian.feitsch@stud-mail.uni-wuerzburg.de

I. INTRODUCTION

The *Building Inspector* is a crowd-sourced project by the New York Public Library aiming to create a digital map of New York City in the 19th century [1]. The project is based on scans of hand-drawn atlases of insurance companies of that time showing building footprints. These are our main interest.

The first step of the digitization is to run an image processing pipeline to recognize the buildings [3]. A building is a depicted area which is completely enclosed by dark lines without gaps. On a website suitable for mobile devices, New York citizens are asked to assess and fix those polygons, if necessary. Users can add, move or delete corners of the polygons as shown in Figure 1. The same polygon can be fixed by several users, resulting in similar, but not identical polygons. This is caused by different interpretations of the map and individual input inaccuracies. Therefore, using one single user-given polygon is not satisfying. We present two approaches to integrate those answers into a consensus polygon.

II. ALGORITHMS

In this section we describe two approaches to find the consensus of a set of similar polygons. An example of what a typical input looks like is given in Figure 2. It also shows a major difficulty as two different footprints are given in one instance. We denote the vertices of a polygon as *corners* to distinguish them from the vertices of a graph.

A. Geometric Approach

This approach depends on a parameter α and outputs the polygon defined by the area covering at least $\alpha \cdot n$ of the n input polygons. The results are not appropriate for the *Building Inspector* because the resulting polygon usually has many more corners than the average input polygon. Also, care must be taken with numerical precision.

B. Combinatorial Approach

Our combinatorial approach first clusters similar corners using DBSCAN [4]. Let $G = (V, E)$ be a directed graph where V is the set of clusters and $(u, v) \in E$ if there is a corner c in cluster u such that (c, c') is an edge of an input polygon and c' is a corner in cluster v. We assume that all of the input polygons are given in clockwise edge order. Let $w(u, v)$ be the number of such pairs (c, c') between u and v.

We evaluated an algorithm which was originally developed by Giraldo Arteaga [2]. We call this algorithm the Voting algorithm. The Voting algorithm first selects the largest cluster u in V, i.e. the cluster containing the most vertices. Ties are broken arbitrarily. In Figure 3, the blue cluster is chosen as start cluster. Then the algorithm follows the edge (u, v) given by the expression $\arg\min_{v \in \text{Adj}(u)} w(u, v)$. From cluster v it proceeds in the same manner until a cycle is found.

The cycle of clusters then needs to be translated back into a polygon. We calculate the centroid for each cluster in the cycle and use those for the consensus polygon.

The crucial part of the combinatorial algorithm is the clustering step. If the clusters are too large, details may be missed by the algorithm (see Figure 4). If the clusters are too small, not all corners of the input polygons are taken into account. Giraldo Arteaga provided parameters for clustering that work well on the *Building Inspector* dataset, but might not be applicable to differently scaled datasets. We developed a heuristic to estimate those parameters scale-independently by choosing parameters such that the number of clusters is equal to the median number of vertices in the input polygons [5].

III. EVALUATION

We focused on evaluating the combinatorial approach. By manual inspection, we found that only 85 % of the user-contributed polygons were describing a footprint only with true corners, i.e. are *semantically correct*. These polygons were all given to the Voting algorithm, which calculated a consensus polygon from up to 15 user-contributed polygons per building. The rate of semantically correct polygons from the algorithm was 96 % using the parameters provided by Giraldo Arteaga. We also examined the precision of the user-contributed and the consensus polygons regarding the black ink lines on the scans. In general, the algorithm gives polygons that are more precise than the average of the user-contributed polygons.

IV. FUTURE WORK

The Voting algorithm works well if every user contributed polygon in fact describes the same building. We found that this is not always the case in our data (see Figure 5). Therefore, a reliable method is necessary to remove outlying polygons. The main open problem for this method is to find a way to automatically learn the parameters needed for clustering.

---

[*] This abstract is based on a paper by F. Feitsch, B. Budig, T. C. van Dijk and M. Giraldo Arteaga with the title “Polygon Consensus: Smart Crowdsourcing for Extracting Building Footprints from Historical Maps” published at ACM-SIGSPATIAL 2016

REFERENCES

[1] New York Building Inspector. http://buildinginspector.nypl.org/.

[2] Mauricio Giraldo Arteaga. Finding shape consensus among multiple geo polygons. http://nbviewer.jupyter.org/gist/mgiraldo/a68b53175ce5892531bc.

[3] Mauricio Giraldo Arteaga. Historical map polygon and feature extractor. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on MapInteraction*, pages 66 – 71, 2013.

[4] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In Usama M. Fayyad Evangelos Simoudis, Jiawei Han, editor, *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, pages 226 – 231. 1996.

[5] Fabian Feitsch. From Many User-Contributed Polygons to One Polygon Consensus. Bachelor Thesis At *University Of Wuerzburg, Chair for Computer Science I*, 2016.

Fig. 1. *Building Inspector* presents a polygon that needs to be fixed. Users can add, move and delete corners such that it matches the building footprint. In this instance, users may delete three of the unnecessary corners, others may push them on the edge. This results in several different user-contributed polygons.

Fig. 2. The polygons in this group describe two different buildings. There is a single polygon representing the footprint of the Hog slaughtering & Packing house on top, and 50 polygons describing the L-shaped building below (of which one is noticeably imprecise).

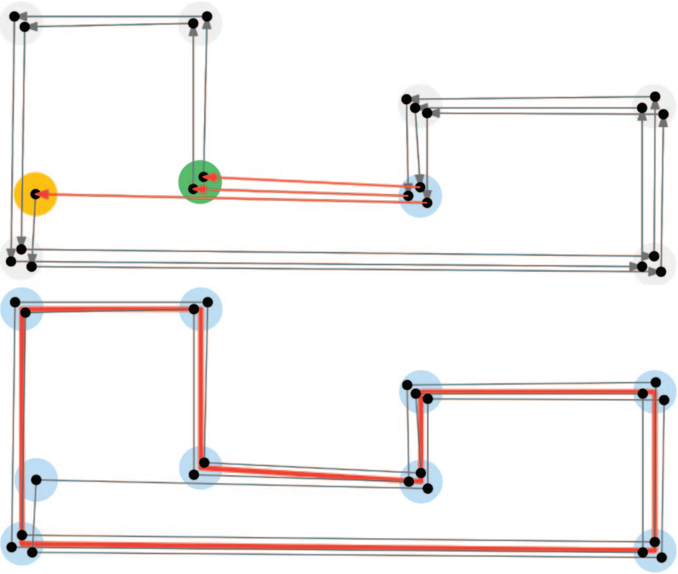


Fig. 3. Example of the Voting algorithm. It starts at one of the largest clusters (here blue). Then it proceeds to the next cluster. In this instance, this is the green cluster. The edge between the blue and the green cluster will be present in the consensus polygon (bottom).

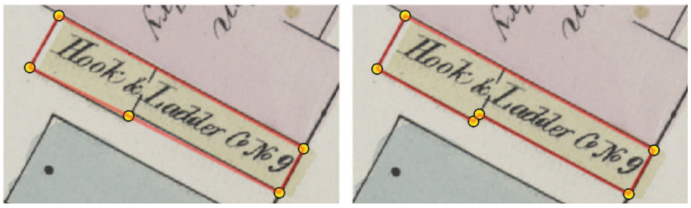


Fig. 4. DBSCAN has a parameter which quantifies the expected distance between the points in a cluster. In the left picture, the parameter was too large causing the consensus polygon given by the Voting algorithm to miss the detailed corner. The right picture shows the expected result.

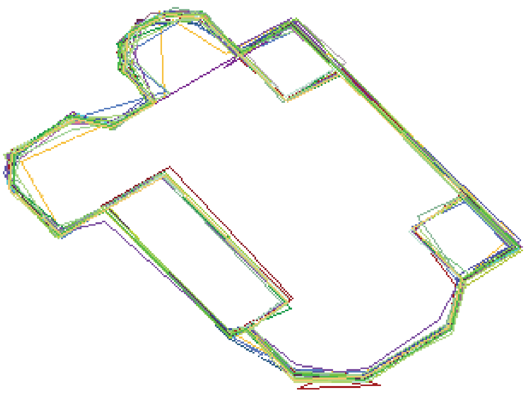


Fig. 5. A problematic input instance. Note that while most polygons look somewhat similar, the users did not at all agree on many of the corners of the building footprint.

Annotating Old Maps with Recogito 2

Rainer Simon
AIT Austrian Institute of Technology
Vienna, Austria
rainer.simon@ait.ac.at

Valeria Vitale
Institute of Classical Studies
London, United Kingdom
valeria.vitale@sas.ac.uk

Leif Isaksen
Lancaster University
Lancaster, United Kingdom
l.isaksen@lancaster.ac.uk

Elton Barker
The Open University
Milton Keynes, United Kingdom
elton.barker@open.ac.uk

Rebecca Kahn
Alexander von Humboldt Institute for Internet and Society
Berlin, Germany
rebecca.kahn@hiig.de

Abstract— **Recogito is a Web-based annotation environment for texts, images and data tables. It has a specific focus on geographical annotation, i.e. the transcription and geo-resolution of place references, and provides a number of features that are of particular relevance for digitized old maps. Among other things, Recogito provides: a zoomable interface for viewing high-resolution imagery, with free rotation; support for images served via the IIIF protocol; the integration of gazetteers to support quick and effortless geo-resolution; sharing options for collaborative annotation; versioning and provenance tracking; and a range of formats for exporting annotation data. Recogito is developed by Pelagios Commons, a Digital Humanities initiative currently funded through the Andrew W. Mellon Foundation, and provided as open source software under the terms of the Apache 2 license. Free access to a hosted version is available at http://recogito.pelagios.org.**

Keywords—annotation; maps; transcription; georeferencing;

I. INTRODUCTION

In recent years significant progress has been made towards enabling the "semantic digitization" of old maps—i.e. not just reproducing the imagery but capturing aspects of the content through geo-referencing [1], vectorization [2] and recognition of symbology and toponymy [3], [4]. Although it is becoming increasingly feasible to employ automatic approaches for implementing these tasks, technical challenges remain high, especially when dealing with earlier, hand-drawn maps [5]. Consequently, the majority of recent projects in this field rely on manual methods.

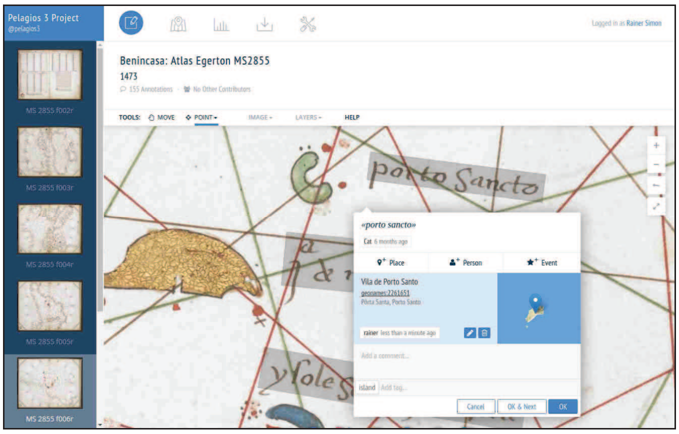
Recogito is a tool that provides, on the one hand, an efficient and intuitive platform for manually annotating maps, and, on the other hand, a framework for integrating different tools and processes. It does so through: a range of data import and export options; a JSON API; and a plugin mechanism (currently under development), with the aim of simplifying potential future extensions. Version 2 of Recogito (which builds on an earlier prototype [6]) was released in December 2016. A publicly hosted version is available for free use at [7]. The code for Recogito is released as open source (Apache 2 licensed). Therefore, it is possible to download Recogito and set up individual instances for personal or institutional use easily and free of charge [8].

II. KEY FEATURES

Recogito provides a personal workspace where documents can be uploaded and managed. With regard to images, upload of common formats (JPEG, TIFF, PNG) is supported, as well as import of images served via the IIIF protocol [9]. When opening a document, different "areas" are available, which address different aspects of the annotation workflow. The **image annotation area** provides a zoomable view of the image, along with selection tools and an editor for adding transcriptions, comments, or tags. For each part of an annotation, Recogito tracks provenance (i.e. who contributed what to an annotation). Unlike general-purpose annotation tools, Recogito has special support for *geo-resolution*. By making use of digital gazetteers (e.g. GeoNames [10] or Pleiades [11]), users can quickly add location information (Fig. 1). The **map area** plots these locations on a Web map, with clickable markers providing additional information about associated annotations as well as image snippets (Fig. 2). From the **download area** (Fig. 3), annotations can be exported in different formats (CSV, GeoJSON, RDF). The **settings area** (Fig. 4) provides a means to edit metadata, access version history, and configure sharing options. Annotated documents can be made visible (but not editable) to the public, while other Recogito users can be invited to collaborate. A tutorial that offers a guide through the steps involved with importing and annotating documents is available at [12].

III. TARGET AUDIENCE AND FUTURE POTENTIAL

The primary audiences for Recogito are researchers and students in the humanities. We are working closely with these communities in order to explore use cases, and set the future path for our roadmap. In addition to supporting scholarly research, we also see potential for using Recogito as a tool to build ground truth and training data for automated annotation methods. We hope that in these ways, Recogito will further empower the study and exploration of old maps in the future.



Benincasa: Atlas Egerton MS2855
1473
155 Annotations · No Other Contributors

Vila de Porto Santo
geonames:2261651
Pôrta Santa, Porto Santo

Fig. 1. The annotation area provides a zoom- and rotatable view of the map image (main area of the screen), and a navigation sidebar (left). Using a set of (currently point and box) selection tools, users can add transcriptions, comments, tags and *geo-references* (links to gazetteers) to regions of the image.

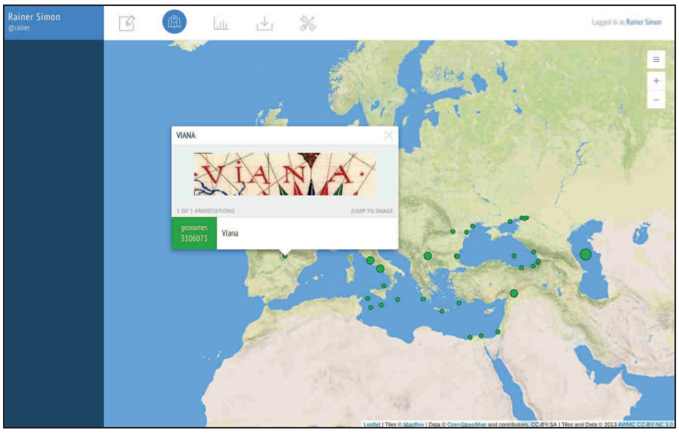
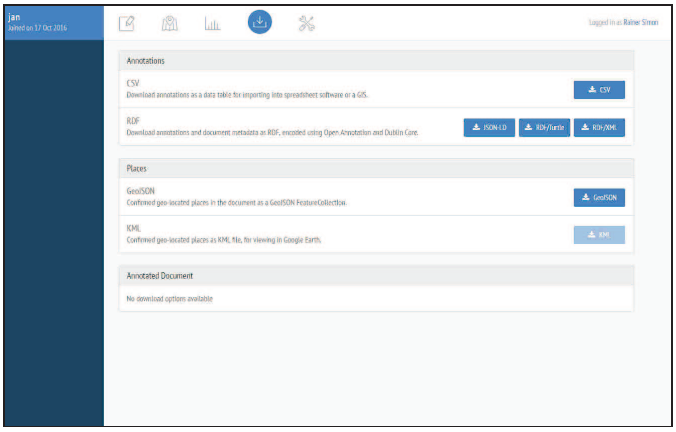


Fig. 2. The map area displays geo-references added earlier in the annotation area (Fig. 1). Clicking a marker brings up a popup shwoing i) information about the gazetteer record that was linked to, and ii) the annotated image region.



Annotations

CSV
Download annotations as a data table for importing into spreadsheet software or a GIS.

RDF
Download annotations and document metadata as RDF, encoded using Open Annotation and Dublin Core.

Places

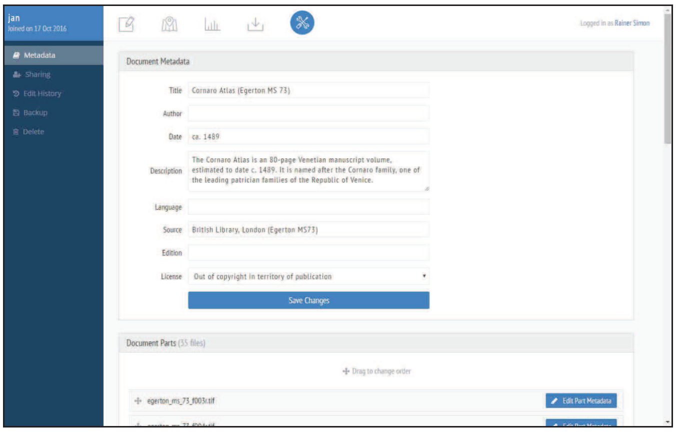
GeoJSON
Confirmed geo-located places in the document as a GeoJSON FeatureCollection.

KML
Confirmed geo-located places as KML file, for viewing in Google Earth.

Annotated Document

No download options available

Fig. 3. The download area provides different export options for annotation data. Currently: comma-separated values (CSV), RDF (using Dublin Core and Open Annotation vocabularies), and GeoJSON (geo-references only).



Document Metadata

Title: Cornaro Atlas (Egerton MS 73)
Author:
Date: ca. 1489
Description: The Cornaro Atlas is an 80-page Venetian manuscript volume, estimated to date c. 1489. It is named after the Cornaro family, one of the leading patrician families of the Republic of Venice.
Language:
Source: British Library, London (Egerton MS73)
Edition:
License: Out of copyright in territory of publication

Document Parts (35 files)

egerton_ms_73_f003r.tif

Fig. 4. The settings area provides means to edit document metadata, define sharing options, and access and revert the annotation version history.

ACKNOWLEDGMENTS

Pelagios Commons and the development of Recogito are funded by the Andrew W. Mellon Foundation.

REFERENCES

[1] C. Fleet, K. C. Kowal and P. Přidal, "Georeferencer: Crowdsourced georeferencing for map library collections", *D-Lib Magazine*, 18 (11), 5, 2012.

[2] I. Iosifescu, A. Tsorlini and L. Hurni, "Towards a comprehensive methodology for automatic vectorization of raster historical maps", *Proceedings of the 8th International Workshop Digital Approaches to Cartographic Heritage*, Rome, 19-20 September 2013.

[3] W. Höhn and C. Schommer, "RAT: A Referencing and Annotation Tool for Digitized Early Maps", *DH Benelux Conference 2016*, June 9-10, Luxembourg.

[4] Y. Chiang and C. Knoblock, "Recognizing Text in Raster Maps", *Geoinformatica* 19 (1), 2014, pp. 1-27.

[5] R. Simon, P. Pilgerstorfer, L. Isaksen and E. Barker, "Towards Semi-Automatic Annotation of Toponyms on Old Maps", *e-Perimetron*, 9 (3), 2014, pp. 105-128.

[6] R. Simon, E. Barker, L. Isaksen and P. de Soto Cañamares, "Linking Early Geospatial Documents, One Place at a Time: Annotation of Geographic Documents with Recogito", *e-Perimetron*, 10 (2), 2015, pp. 49-59.

[7] Recogito, http://recogito.pelagios.org (accessed March 17, 2017)

[8] Recogito source code repository, http://github.com/pelagios/recogito2 (accessed March 17, 2017)

[9] International Image Interoperability Framework (IIIF) http://iiif.io/ (accessed March 17, 2017)

[10] http://www.geonames.org/ (accessed March 17, 2017)

[11] http://pleiades.stoa.org (accessed March 17, 2017)

[12] Recogito 10 Minute Tutorial, http://recogito.pelagios.org/help/tutorial (accessed March 17, 2017)

Tunnels of Knowledge:

Mapping today's 'secrets' from Yesterday's public maps

* (and improving public safety)

Paul Sladen
Astronomical Calculation Institute (ARI–ZAH), University of Heidelberg
sladen@{ari.uni-heidelberg.de,gmail.com}

***Index Terms*—Channel Tunnel. Citizen science. Crowdsourcing. London Underground. OpenStreetMap. Public safety. Public transport. Tunnels. Underground infrastructure.**

I. INTRODUCTION

Everyday millions of people travel by U-Bahn, Metros, subways and the Tube. Most travel in a blinded fashion, steered through a brightly-lit linear maze, and unaware of where they are. Modern society often choses actively or passively to restrict maps of underground infrastructure, even when it is open for all. This is in contrast to the social dynamics of the early 1900s, when engineers were enthusiastic to share and the public eager to absorb.

OpenStreetMap [1] is a worldwide geo-database containing a contributed and crowdsourced vector dataset along with textual metadata, quickly making it one of the largest datasets available for research and analysis [2]. For citizens involved with mapping transport and underground metro systems, use of old out-of-copyright maps allows obtaining and rescuing information otherwise unavailable in the modern world. The presentation will show mapping of features from the complex London Underground system using old parliamentary maps as a basis; and efforts for precision mapping of the Channel Tunnel using custom software written to reproject numerical tables of engineering survey data—where existing libre software where existing software for projecting absolute azimuth/distance into WGS84 was not readily available.

II. TUNNELLING TOWARDS OPENNESS

Following incidents in both London and the Channel Tunnel, official reports from the public humanities investigating the handling of the incidents have highlighted lack of available information as a significant contributing factor [1], [3–5]. Yet, other players continue to advocate for non-publication in digital media form [5]. Using historical maps via digitisation, a [citizen] transport user can proactively enact official recommendations to improve public safety by dissemination of information. For myself, as an OpenStreetMap practitioner with a background in programming and an active interest in open knowledge, the possibility of citizen science being able to have direct agency through the use of old maps is intriguing.

A. London Underground

During the 2010s a member of London Underground's Infrastructure Protection team began research for their own MA and PhD theses. The published papers resulting contain many examples contrasting historic maps with output from London Underground's internal infrastructure maps [6], [7]. The allowance for including internal London Underground maps is perhaps an indication of a more open attitude by London Underground. Having such plans enables the opportunity for comparision against OpenStreetMap, but also the possibility of tainting OpenStreetMap. During one comparison a substantial difference was observed—when queried it was confirmed that the alignment on OpenStreetMap was correct, and it was the London Underground's internal maps that were out-of-date.

In 2015 Transport for London released additional quasi-3D (axonometric) diagrams of the internal structure of most of the underground stations [8]. This caused additional information to fall into the public *knowledge* domain—but not necessarily into the public domain of (legally usable) information and data available for freely-using. OpenStreetMap must therefore still turn to information that is truely out-of-copyright. [Fig. 1] Luckily, many of the metro and underground railways were constructed in the 1900s–1920s meaning original construction plans have fallen out of copyright and can be safely used as a research base for digital mapping.

B. Channel Tunnel

This international undersea tunnel was designed in the 1980s, privately constructed, and opened in the early-1990s. Plans and maps remain in copyright, split between Eurotunnel, and the various private construction companies. An entirely different approach has been applied, tabulating individually published numerical constraints (spacing, lengths, radii) to recreate a virtual survey [9], [10]. Publications cover survey traverses performed using gyrotheodolite, first along the tunnel walls, then secondly along the centre-line—to avoid a newly-observed problem of horizontal refraction caused by temperature gradients in the air at the tunnel walls [11], [12]. We confirm correct re-projection into WGS84 of the surveys by observing one zig-zagging around other traverse. [Fig. 2]

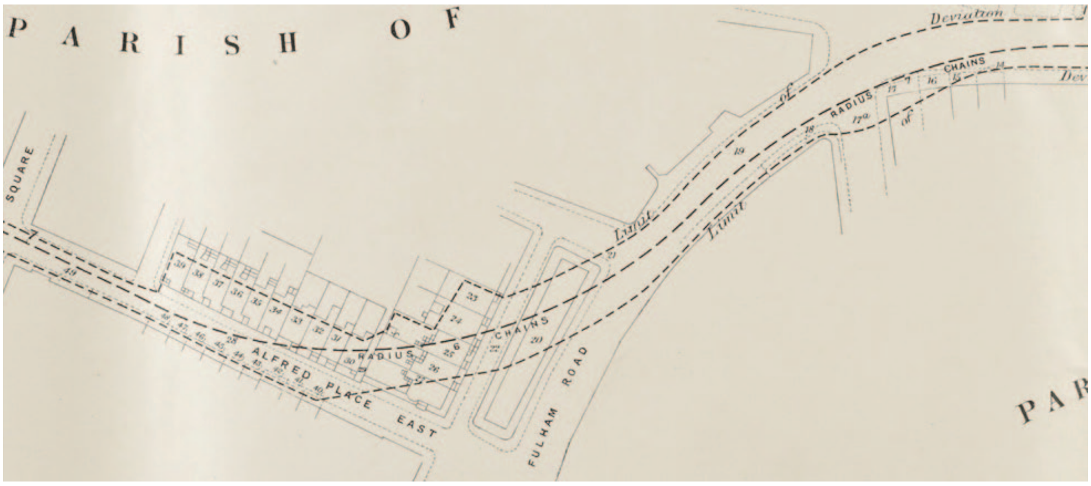


Fig. 1. London Underground (1897 submitted plans) later the Piccadilly line. Development of "South Kensingtons Bends" double S-curves under highway.

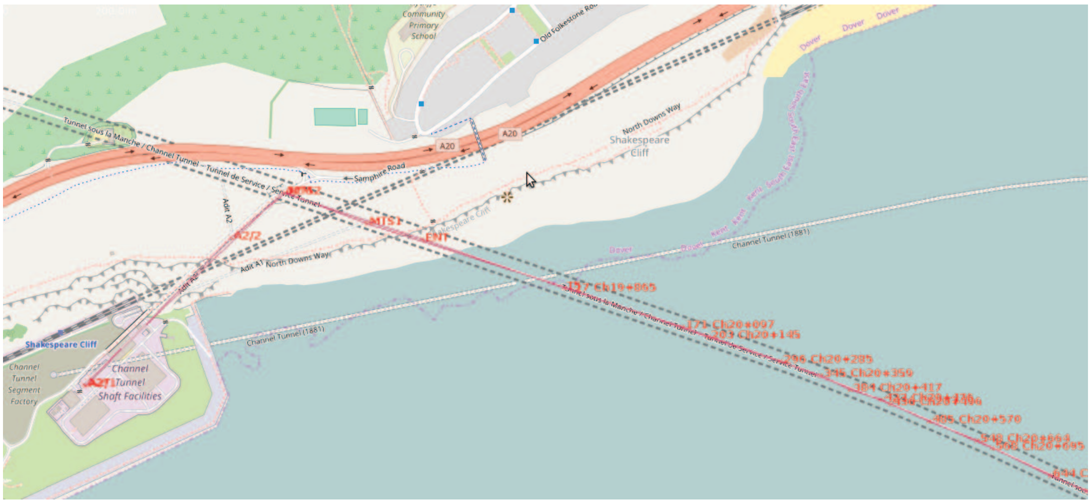


Fig. 2. Channel Tunnel: Shakespeare Cliff, tunnels close to Dover on the British coast (OpenStreetMap, CC-BY-SA). 1989/1990 survey traverses in red.

REFERENCES

[1] OpenStreetMap contributors, "Openstreetmap geo database," http://www.openstreetmap.org, 2004–2017.

[2] N. Street, "TimeContours: Using isochrone visualisation to describe transport network travel cost," Jun. 2006, final Report. [Online]. Available: http://www.imperial.ac.uk/pls/portallive/docs/1/18619712.PDF

[3] Bureau d'enquêtes sur les Accidents de transport terrestre and Rail Accident Investigation Branch, "Technical Investigation Report concerning the Fire on Eurotunnel Freight Shuttle 7412 on 11 September 2008," Nov. 2010, "Only knowing the PK at which mission 7412 has stopped, he looks for the corresponding cross-passage number in the tunnel pocket book. This takes about two minutes.". [Online]. Available: https://assets.publishing.service.gov.uk/media/547c900c40f0b60244000185/101122_ReportET2010_eurotunnel_eng.pdf

[4] Rail Accident Investigation Branch, "Penetration and obstruction of a tunnel between Old Street and Essex Road stations, London 8 March 2013," Mar. 2014, "Railway Infrastructure Managers with tunnels and associated subterranean structures which are under urban areas ... should then take all reasonable steps to publicise this information,". [Online]. Available: https://assets.publishing.service.gov.uk/media/547c8fb940f0b60241000157/R032014_140213_Old_Street.pdf

[5] R. Taylor and London Underground, "Location of Mid Tunnel Vent Shafts and Intervention Points," *WhatDoTheyKnow*, Oct. 2011, Freedom of Information request. [Online]. Available: https://www.whatdotheyknow.com/request/location_of_mid_tunnel_vent_shaf

[6] N. Darroch, "London's deep tube railways: Visibly invisible," University of York, Sep. 2012, MA by research. [Online]. Available: http://etheses.whiterose.ac.uk/3905/1/London%E2%80%99s_Deep_Tube_Railways-_Visibly_Invisible,_Nathan_Darroch.pdf

[7] N. Darroch, "A brief introduction to London's underground railways and land use," *Journal of Transport and Land Use*, vol. 7, no. 1, pp. 105–116, 2014. [Online]. Available: http://econpapers.repec.org/RePEc:ris:jtralu:0124

[8] Georg Vehres and Transport for London, "Maps of public corridors on larger Tube stations," 2015. [Online]. Available: https://www.whatdotheyknow.com/request/maps_of_public_corridors_on_larg

[9] P. Sladen, "Channel Tunnel," 2013–2017, Numerical constraints. [Online]. Available: https://wiki.openstreetmap.org/wiki/Channel_Tunnel

[10] P. Sladen, "Tube to Chunnel—Underground Mapping," *State of the Map*, Sep. 2013. [Online]. Available: http://lanyrd.com/2013/sotm/scpkgy/

[11] N. Korittke, "Zur Anwendung hochpräziser Kreiselmessungen im Bergbau und Tunnelbau," Technische Universität Braunschweig, 1997, doctoral thesis, ISBN: 3-926146-09-5.

[12] N. Korittke, "Influence of horizontal Refraction on the traverse Measurements in Tunnels with Small Diameters," pp. 13–17, 1989. [Online]. Available: http://www.slac.stanford.edu/econf/C9009106/papers/023.PDF