

Project Newsletter

For the members of the Advisory Board

Dear Members of the Advisory Board,

We hope you are well and have enjoyed a pleasant summer break. We realised it would have been difficult to organise a meeting while a few members of both the team and the advisory board were on leave, so we have decided to prepare a newsletter instead, to update you about the progress of our project. As always, we are very grateful for your feedback and input, and happy to answer any question you may have.

We will have a more traditional meeting in November, but, in the meantime, here are some highlights from the first months of Machines Reading Maps.

Prototype and Workshops



About the Project

A prototype version of our annotation platform, that combines the functionalities of [Recogito](#)—the semantic annotation platform developed by Rainer for Pelagios— and [mapKurator](#)—the tool developed by Yao-Yi and Zekun for automatically detecting, disambiguating, and linking text on maps—is ready

to be tested! Thanks to this integrated platform we will be able to use Recogito to create manual annotations of historical maps that will also become the gold standard for the automatic annotation performed by mapKurator. We are very keen on receiving feedback from different kinds of users, and we have organised two small workshops, one for Europe-based participants (on October 12th) and one for North America-based ones (on October 18th). We will use initial feedback gathered after these events to improve the present workflow, and we will then organise larger-scale workshops and training events.

In planning this grant, we have scoped a workshop with each of our cultural heritage partners (British Library, National Library of Scotland, Library of Congress, USC Libraries). We are ready to begin planning these events with you and will be contacting you soon to discuss ideas and details. We also welcome requests for workshops from other institutions/projects, so if you want to organise something around our tools and approaches, please get in touch.

A screenshot showing our annotation interface, integrating Recogito and mapKurator



Data Transfer



We are finalising data acquisition with all our library partners. First, we have accessed National Library of Scotland maps and map sheet metadata. We are finalizing the data transfer method for Library of Congress Sanborn fire insurance maps and this should be complete in December. The Library of Congress will use this opportunity of working with MRM to make their Sanborn collection available not only to us, but to the public as a data download. We have begun to explore the Sanborn collection thanks to an export of the collection metadata. Finally, we are discussing the best approach for acquiring the Goad fire insurance maps held at the BL as well as other small₂ map collections.

GeoReferencing



we have just started an exciting conversation with the LC about georeferencing the Sanborn Map Collection. The scale of the collection and the limited resources available present a number of challenges, but we are very enthusiastic about the potential outcomes, and we have started exploring some practical approaches. We plan to use mapKurator to produce approximate locations for the whole collection, which will enable LC to achieve an acceptable balance between quantity and quality/accuracy. The information can be then possibly verified and/or improved by human users, but we will discuss that more in detail at a later stage.

Communications and Outputs



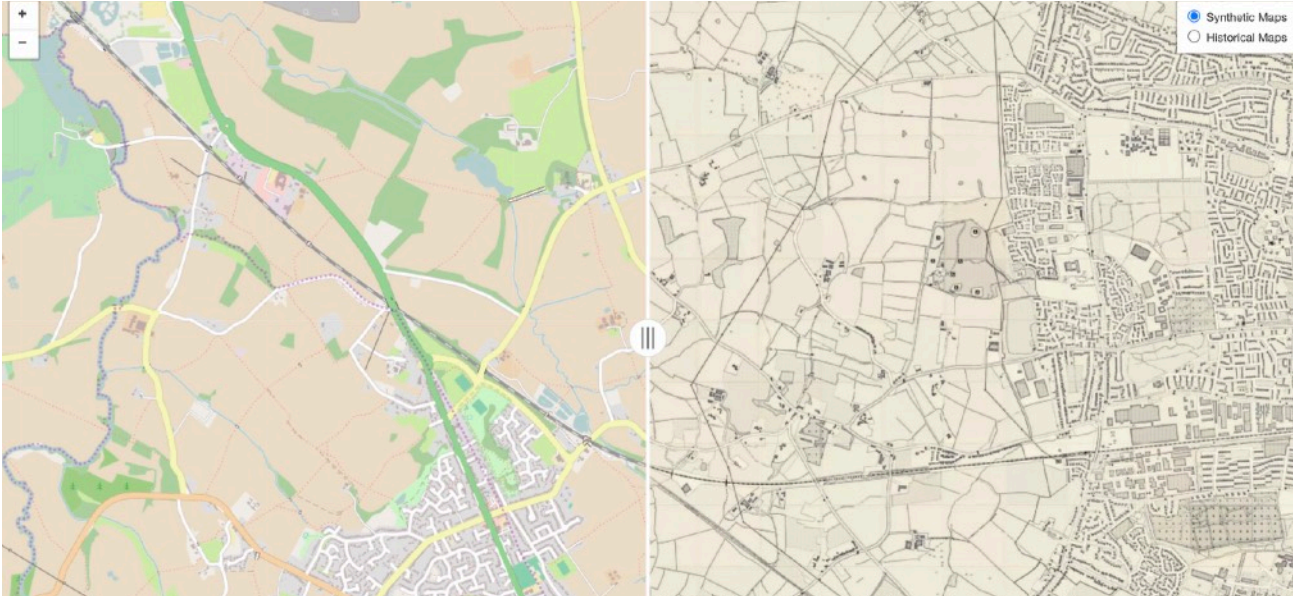
A [blogpost](#) discussing the use of synthetic maps in machine-learning algorithms to detect and read text on historical maps has just been published in the National Library of Scotland's blog. We would like to thank Chris Fleet for his help in the writing and editorial process.

If you want us to write a piece for a digital venue in your institution, we would be delighted to do so. We can guarantee amazing images too!

A poster we submitted to [Datafication in the Historical Humanities: Reconsidering Traditional Understandings of Sources and Data](#) has been accepted. We are also going to propose a collaborative activity for the [Linked Pasts 7 Symposium](#), that will take place in Ghent this December. We have also submitted a paper introducing the synthetic historical map generation pipeline to the [GeoAI 2021](#) workshop at [SIGSPATIAL2021](#). We are working on linking text labels on maps to the WikiData knowledge base, and inferring semantic types (e.g. road, building) for labels that can not be linked to WikiData (i.e. no corresponding entity in WikiData KB), and we plan to submit this work to [ESWC 2022](#). We also have a paper about stitching together complex map scans in the pipeline, and we are about to submit an abstract to the Programming Historian for their call for papers, in the form of tutorials, around [computational analysis skills for large-scale humanities data](#).

In the longer term, we are also working on a paper detailing our annotation methodology, and how it can aid the study of historical maps, and a paper looking at the representation of

Side by side view of Open Street Maps and the corresponding synthetic map recreating OS maps style.



User Profiles and Collaborations



We have been thinking of who the people are, outside of our project partners, who could make use of the workflow and tools we are developing, and how their needs may differ. Through meetings and conversations, we are identifying different typologies of users, ranging from researchers in the humanities to professionals in cultural heritage to machine learning experts, and trying to create user profiles that capture their expectations and the kind of applications they may find interesting. We are also exploring different types of relationships with potential partners, with varying degrees of involvement on our side, but always based on the principles of reciprocity and mutual benefit. Considering our current interactions, we are keen to offer collaborations around:

1. The use of our manual annotation methodology (users will be able to apply a free research tool within their project/collection, and will provide us with useful feedback on the flexibility of our semantic approach).
2. The use of our computer vision technology (users will be able to apply a computer vision technique for the detection of map labels within their project/collection, and they will provide us with feedback on the limits of our pipeline)

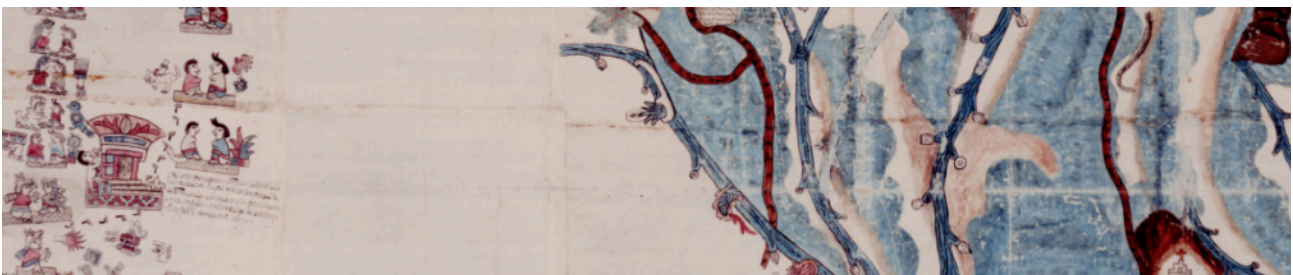
3. The support of humanities research focused on historical maps, in particular looking at ethnic urban enclaves (users with research questions around the distribution and evolution of ethnic urban enclaves in the US and the UK will be able to use our technology and methods, but also receive some support in the process of gathering and analysing the data. This will be an important test for the validity of our workflow as a research tool able to generate new, relevant knowledge from historical maps.)

We have had very useful and stimulating conversations with other projects that have a strong focus on historical maps, such as [Unlocking the Colonial Archive](#) and [OS2000](#). These collaborations may widen the impact of our tools, and help us test our approach on different datasets. These relationships may also lead to interesting data comparisons and, potentially, to more formal outputs such as collaborative academic articles or joint grant applications.

We have also explored possible collaborations at a different scale, getting in touch with individual experts, academics as well as curators, who work with smaller collections and have, sometimes, little to no resources for digital projects, like a striking collection of maps drawn by Holocaust survivors, or a collection of maps of Los Angeles through time held at USC. These unique collections represent an interesting challenge for our tools, and we are curious to discover how flexible is our method, and how able it is to cope with more idiosyncratic corpora. Collaborations with smaller projects also help us to understand how accessible our outputs really are, especially to users less familiar with computational technologies, and how easy they are to set up and maintain, even when our potential users can't access institutional resources.

All these collaborations, of course, proceed hand in hand with our own research questions, that revolve around maps as complex semiotic systems, and the representation of antiquities and historical sites.

Detail of an historical Mexican map as featured on the [Unlocking the Colonial Archive](#) website



Say hello (and goodbye) to the new team members



About the Team

We want to introduce to the advisory board the two brilliant new PhDs who have worked/are working on our project.

Daniil Feldman worked from May to August on the integration between Recogito and mapKurator, and helped to set up the dedicated virtual machine at the Turing. He was with us only for a few months but he did a great job, and we wish him all the best for his PhD. Our latest addition is Jina Kim, who is doing her PhD at the University of Minnesota, under the supervision of Yao-Yi, and is currently working for MRM on automatic toponym disambiguation.



New Affiliations

Yao-Yi and Zekun have traded their sunglasses for fleece hoodies and moved from University of Southern California to University of Minnesota, where they were joined by Jina.

Official Welcome

After a long legal limbo, we were delighted to officially welcome Rainer to the team.

The New Recogito (feat mapKurator)



About the Tools

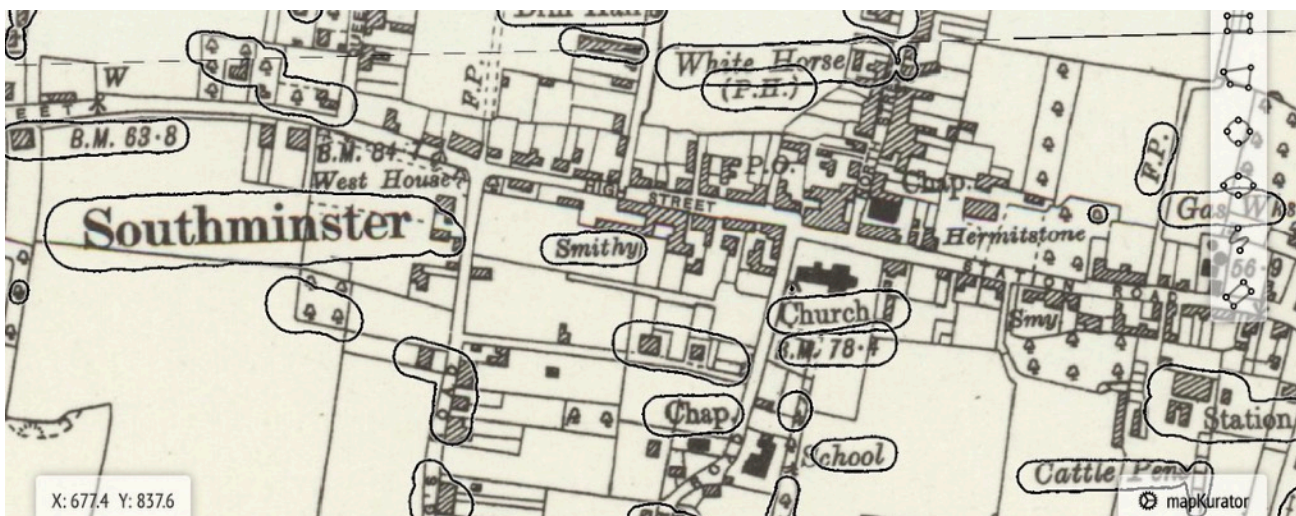
Recogito has incorporated a number of functions that Rainer had developed for his image-annotation tool Annotorious. Recogito now features a wide range of drawing tools that should ensure a high degree of flexibility in the annotation, even with the most unusual maps. Although our main goal is, ultimately, to gather gold standard data for the automatic recognition of text labels on maps, we think that many users will be interested in exploring the manual annotation per se, and we are very keen on making this process self standing and as rewarding as possible. Thinking of users interested in manual

annotations of maps, we have redesigned the Recogito annotation pop-up, and substituted the old categories, “place”, “person” and “event”, with something focused on place information, but also able to capture the complex relationships between text and other signs on the maps. The new categories are: “entity”, “label” and “symbol”. If you want to know more about the reasoning behind this decision, feel free to ask us to elaborate. It will be good practice for the article we are preparing on this topic!

Like in the previous instance of Recogito, there will be room for comments and tags. It will also be possible to constrain the tags within a controlled vocabulary, as well as to use free tagging.

But the most exciting new development in Recogito is its integration with mapKurator, which enables users to automatically generate a large number of text annotations on a map image. These annotations can then be refined and enriched manually in the Recogito’s annotation interface. We are quite proud of our platform, and we believe that it combines two complementary and powerful approaches to cater for a range of different research and curatorial needs.

Detail of an OS map annotated automatically with mapKurator



Gazetteers

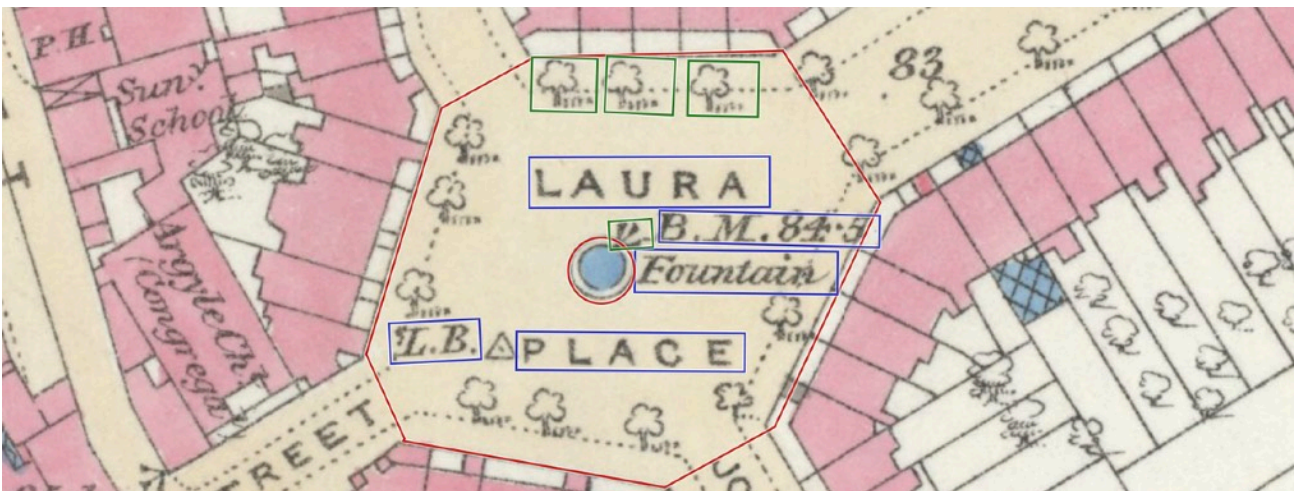


A key feature in Recogito and mapKurator is the opportunity to link any annotation to a gazetteer’s entry. However, there isn’t, at the moment, a suitable gazetteer of places in the UK in the 19th and 20th century. We have decided to join forces with the colleagues working on the [Living with Machines](#) project, and reuse a resource they created collecting information from Wikidata and storing it into a CSV. We have converted the data

in the CSV into [Linked Places](#) format (a type of geoJSON that is becoming a de facto standard in Linked Open GeoData), and transformed it into a (pseudo)gazetteer. We are going to use this resource as a library in our platform, so that annotations can be easily disambiguated via, and linked to, Wikidata IDs. In collaboration with LwM, we plan to make both the gazetteer and a guide to the data conversion workflow available to all in the near future.

We think that this pipeline is simple and effective, and could help many small projects customise their own local instance of Recogito, and deal with the current lack of historical gazetteers that still makes semantic annotations of places challenging.

If you are now curious to try our annotation platform, we are always happy to organise demos and workshops for small groups or even one-to-one sessions. You can also test the tool on your own, just let us know so that we can set up your access credentials. We are preparing detailed tutorials that we hope you'll find useful.



Detail of an OS map annotated manually with Recogito

That's all from us this time. We look forward to any feedback, questions, and comments. We hope to see many of you at our next meeting, we will soon circulate a poll for selecting a date.

The Machines Reading Maps Team