

## Linking Early Geospatial Documents, One Place at a Time: Geo-Tagging Texts and Maps with *Recogito*

Rainer Simon, AIT Austrian Institute of Technology, [rainer.simon@ait.ac.at](mailto:rainer.simon@ait.ac.at)

Elton Barker, The Open University, [elton.barker@open.ac.uk](mailto:elton.barker@open.ac.uk)

Leif Isaksen, University of Southampton, [L.isaksen@soton.ac.uk](mailto:L.isaksen@soton.ac.uk)

Pau de Soto Cañamares, University of Southampton, [p.desotocanamares@soton.ac.uk](mailto:p.desotocanamares@soton.ac.uk)

*Recogito is a Web-based tool for the structured annotation of place references in texts and images. As part of the Open Humanities Awards 2014, we held two “hackathon”-like workshops, where a mixed audience of students and academics of different backgrounds used Recogito to annotate literary texts from the Classical Latin and European Medieval period, as well as Medieval Mappae Mundi and Late Medieval maritime charts. At the end of the day, participants had added several thousand contributions, all of which are now openly available for download and further re-use. The resulting data can be used, for example, to “map” and compare the narrative of the texts, and the contents of the maps with modern day tools like Web maps and GIS; or to contrast documents’ geographic properties, toponymy and spatial relationships. Contributing to the wider ecosystem of the “Graph of Humanities Data” that is gathering pace in the Digital Humanities (linking data about people, places, events, canonical references, etc.), we argue that initiatives such as this have the potential to open up new avenues for computational and quantitative research in a variety of fields including History, Geography, Archaeology, Classics, Genealogy and Modern Languages.*

### 1. Background: the Pelagios Project and SEA CHANGE

Pelagios<sup>1</sup> is a community-driven initiative that facilitates better linkage between online resources documenting the past, based on the places that they refer to. Our member projects are connected by a shared vision of a world – most eloquently described in Tom Elliott’s article ‘*Digital Geography and Classics*’ [1] – in which the geography of the past is every bit as interconnected, interactive and interesting as the present. Each project represents a different perspective on our shared history, whether expressed through text, map or archaeological record. But as a group we believe passionately that the combination of all of our contributions is enormously more valuable than the sum of its parts.

The goal of Pelagios’ current project phase (“Pelagios 3”, funded by the Andrew W. Mellon Foundation) is to annotate, link and index place references in digitized *Early Geospatial Documents* – documents that use written or visual representation to describe geographic space prior to 1492. Through a series of six thematic work packages, Pelagios 3 will work with documents from the Latin, Greek, European medieval, maritime, as well as early Islamic and Chinese tradition. *Recogito* is a Web-based tool we developed specifically for use within the project team, to facilitate this work. However, the potentially unlimited number of documents to which our methodology would be suited means that establishing and honing community-based approaches will be essential in order to scale it beyond the pre-modern era.

The Open Humanities Awards have provided us with an impetus for trialing *Recogito* with a wider audience: under the title SEA CHANGE,<sup>2</sup> we held two public geo-annotation workshops with a mixed audience of students and academics of varying backgrounds (geography, history, engineering, and archaeology). Our primary goal was to explore the potential of *Recogito* as a tool for crowdsourcing and collaborative geo-annotation, but we were also interested in how and if a workshop format such as this is a suitable way to engage with a wider audience, and as a means to build community.

---

<sup>1</sup> <http://pelagios-project.blogspot.co.uk>

<sup>2</sup> Socially Enhanced Annotation for Cartographic History And Narrative Geography, <http://dm2e.eu/open-humanities-awards-round-2-winners-announced/>

## 2. Recogito

Recogito features several work areas (see Fig. 1), each dedicated to different stages of the geo-annotation workflow: an image annotation area to mark up and transcribe place names on map or manuscript scans, a text annotation area to demarcate place names in digital text, and a geo-resolution area, where the identified (and transcribed) place names are mapped to a gazetteer (and, thus, to geographical coordinates). *Recogito* also provides basic features for managing documents and their metadata, as well as functionality for viewing and downloading annotation data and usage statistics. Editing functionality is limited to registered users. However, data downloads and basic overview information is also available for access to the public. Our own production instance of *Recogito* is hosted at <http://pelagios.org/recogito>. The tool as such, however, is Open Source software (available from the Pelagios project's GitHub repository <http://github.com/pelagios/recogito>), which makes it possible to set up additional instances of *Recogito* for personal or institutional use.

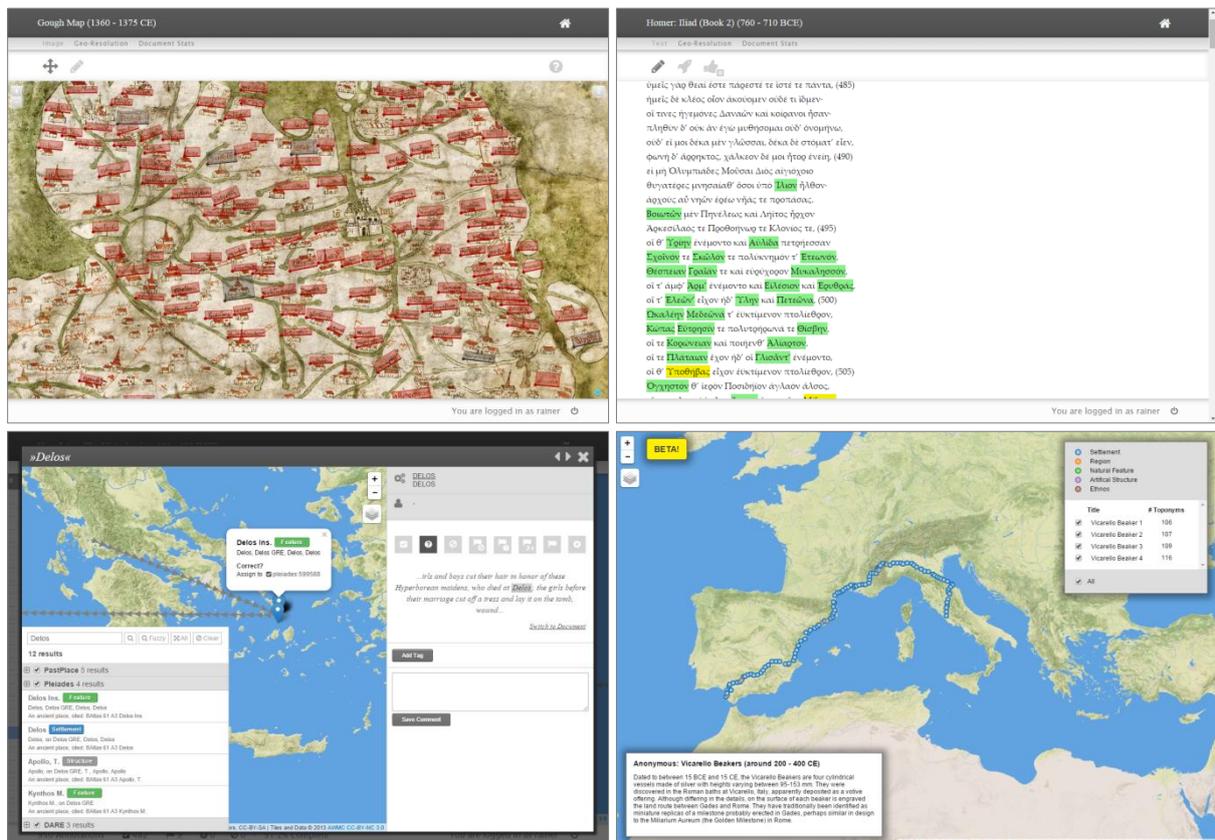


Fig.1 Recogito work areas: image annotation (top left), text annotation (top right), geo-resolution (bottom left), public map (bottom right).

## 3. Annotation Workshops

Our two workshops took place on October 31, 2014 at the Heidelberg University Institute of Geography, and on December 4, 2014 at the University of Applied Sciences Mainz. We started both days with a brief introduction to the goals and background of Pelagios, and a short tutorial of how to use Recogito's different work areas. (A written beginner's tutorial is also available online at <http://pelagios.org/recogito/docs>.) For each workshop, we defined a general thematic scope, and prepared material for annotating accordingly: Classical Latin texts and medieval maps for Heidelberg; Medieval travel writing and pilgrimage itineraries, and medieval nautical charts for Mainz. Beyond that, however, participants were free to choose which documents they wanted to work on, and which tasks they would focus on (tagging, transcribing, mapping toponyms to gazetteer records). Group sizes were roughly equal in both workshops, with 27 users in Heidelberg and 22 in Mainz.

After the introduction, we dedicated about 2 ½ hours to annotation work. The afternoon session, we used as a more open space for hands-on exploration. We wanted to get the audience thinking about the question: "now

*that we have annotated our documents, what can we do that we couldn't do before?"* As a concrete example, we prepared a tutorial which walked the audience through the steps necessary to download data from *Recogito* and analyze it further in QGIS (an open source Geographic Information System). This way, they could e.g. explore a medieval travel itinerary, and match the rate of stops and their different types against a 3D terrain model, pondering about the time taken – and the hardships endured – by travelers in the 4th century AD during their journeys. In the Mainz workshop, where part of the audience had an engineering background, we additionally prepared a short “hacking tutorial” consisting of small programming tasks that demonstrated how to re-use annotation data to create Web maps, timelines or network graphs, using JavaScript as a programming language.

### 3.1 Results Heidelberg

The quantity of contributions made by our participants greatly exceeded our expectations: on the first workshop day (Fig. 2), we recorded a total of 6,620 contributions, associated with 51 different documents (19 text documents, 8 of which were in Latin; and 32 map scans). Four participants even made it into our all-time top-10 list, which means that they managed to make more than 645 contributions in that morning session. The contributions consisted of approximately 2,650 place name identifications in text, 2,500 place name identifications on maps, 830 map transcriptions, 140 gazetteer resolutions and about 490 other actions, such as corrections, deletions or comments.



**Fig.2 Impressions from the SEA CHANGE Heidelberg workshop: participants working on medieval maps.**

Participants seemed to genuinely enjoy the process. Not only did we get positive feedback after the session, but several participants also followed our invitation to get permanent *Recogito* logins so that they can continue contributing after the workshop. (We recorded a further 1,648 contributions on Saturday, November 1st, the day after workshop.)

It was interesting for us to see such a clear division in terms of how the number of contributions was distributed over different task types. On the one hand, they reflect how different phases of the annotation workflow are more or less time consuming. Demarcating a place name in a text is usually a matter of a double click, for instance, whereas on a map it takes longer to navigate the image and select the area (selecting is a process that involves a mouse click, drag, and another click). Hence the roughly equal number of name identifications in texts and maps, despite the fact that more people were working on maps. Transcribing takes even more time, as we might expect; as does gazetteer resolution, i.e. searching through lists of potential gazetteer search results, and picking the one that most likely corresponds to the place name in question.

### 3.2 Results Mainz

For the workshop in Mainz, we followed the same procedure as in Heidelberg. In response to the low number of gazetteer resolutions (and feedback we had collected about it) we decided to re-design the user interface of this

particular *Recogito* work area beforehand, in particular with regard to where UI elements were placed, and the amount of screen real estate that was dedicated to them (e.g. giving more space to the map, while search results would be organized into groups and “folded” into collapsible lists to take up less screen space). The Mainz workshop was the first live trial run for this revised interface.

At the end of the day, we recorded a total of 7,511 contributions. These consisted of approx. 2,600 place name identifications in text (roughly an identical number to our first workshop); almost 3,200 place name identifications on images (significantly more than in the first workshop); about 620 map transcriptions (slightly less than the previous 830); 544 gazetteer resolutions; and 537 other activities such as corrections, comments, and deletions.

#### **4. Conclusion**

Overall, we were extremely happy with the amount of data our participants generated in the short time, and the continuity in terms of distribution of contributions over tasks. This seemed to show that *Recogito* is reaching a level of maturity that qualifies it for “non-expert use”, beyond the confines of our Pelagios project team.

It is also interesting to speculate about where some of the differences in the results may have come from: for example, it was interesting to see significantly more place name identifications on maps in the second workshop. We assume this was simply a result of the different material. The medieval nautical charts we prepared for the second workshop are very “dense” in place names, and the place names are typically arranged in sequence, in the same orientation. So there is less need for users to search and navigate the map. That may have allowed for slightly speedier tagging. On the other hand, though, the style of lettering in these maps was rather different from last time and much more challenging for the non-expert to decipher. This may well be the reason why the number of transcriptions was lower. Furthermore, we were particularly happy to see the almost 4 times increase in gazetteer resolutions, which is an indication of the positive impact our user interface redesign had.

The two workshops were our first significant attempt at reaching out to a broader community. The results have encouraged us to look more closely into “community-sourcing” as a future strategy for Pelagios and beyond, and to evolve our approach and toolset further into this direction. However, more work and experimentation will be needed to understand factors that influence crucial aspects such as ease of use, data quality issues, and what makes the annotation process motivating and fun (in particular to users that lack expert knowledge about ancient sources and historical background). In terms of the latter, light-hearted competition clearly played a part (which we helped foster with a live feed of statistics throughout the sessions). But motivation needs more than just point scoring: one specific feedback we took away from SEA CHANGE in this regard was that people seemed to enjoy the process most when they found meaning in it for themselves. One student, for example, commented on the experience of annotating an illustrated itinerary from a medieval manuscript – a document which, from a modern person’s point of view, wouldn’t be considered very “map-like” in appearance. She remarked that while she was annotating the document, the geographical nature of the document would progressively start to unfurl to her. As she identified places step by step, she would begin to “see it as a map”.

#### **5. Acknowledgements**

The authors wish to thank the Andrew W. Mellon Foundation, the DM2E project and the Open Humanities Awards for funding this work. Furthermore, we are indebted to our workshop hosts: Lukas Loos and Armin Volkmann from the University of Heidelberg, and Kai-Christian Bruhn and his team from the University of Applied Sciences Mainz. Last but not least we want to thank everyone who attended our workshops for their participation and enthusiasm.

#### **6. References**

- [1] Elliott, T. and Gillies, S. 2009. Digital Geography and Classics. In *Digital Humanities Quarterly*. Vol 3. Number 1. <http://www.digitalhumanities.org/dhq/vol/3/1/000031.html> (last visited January 22, 2015)