

Recogito

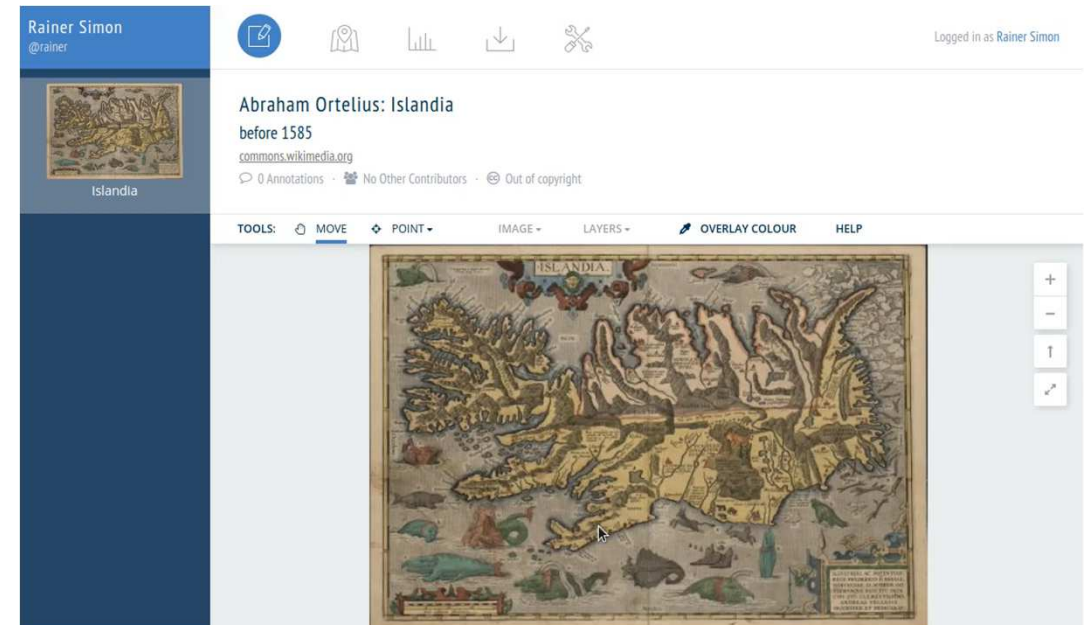
A Platform for Semantic Document Annotation

Rainer Simon, AIT @aboutgeo

What is Recogito?

- Web-based annotation tool for **text** and **images**
- “Semantic annotation” (more anon)
- Open source software
- Developed by **Pelagios Commons**, an international Humanities Linked Open Data initiative

<https://recogito.pelagios.org>



Some Background & History (1)



- Started life as internal tool for Mellon-funded project in 2014
 - Geographic mapping of texts & maps from classical period up to 1492
 - 171 registered accounts, ~10 “power users”
 - 317 documents in 8 languages
 - ~151,000 acquired toponyms + ~50% mapped to gazetteers
- Further funding from Andrew W. Mellon Foundation in 2016, 2018
- Now a free public online service since ~2.5 years
- Two major version releases (but actually constant development)

Some Background & History (2)



2.06M Annotations	535 Edits last 24hrs	3,180 Users
----------------------	-------------------------	----------------

- 4.000+ documents
- 22% image documents
- 1.5% IIF presentation manifests



Semantic Annotation?



- Digitize **historical sources** (and bring them online)
- Identify **concepts** and **entities** (places, people, events)
 - Disambiguate and resolve against digital **authority lists**
 - Identify concepts, entities and authority records through **URIs**
- Publish data as **Linked Open Data (LOD)**
 1. URIs as names for things (source, annotations, concepts, relations)
 2. HTTP URIs so that people can look them up online
 3. Information provided in Open Standards, e.g. RDF, JSON-LD
 4. Link to other data on the Web

Live Demo



<http://recogito.pelagios.org>

DH Awards 2018 Winner: Best DH tool or suite of tools

Why would you use Recogito?



- **Map transcription & markup**
 - Single scholar or crowd-/community transcription
- **As an online collaboration space**
 - Collaboration for users with different digital expertise
- **Teaching**
 - Critical engagement with sources – compare, discuss, visualize
 - Digital skills
- **Contribute to the LOD ecosystem**
 - For others to query, harvest, remix, hack, use in their own code,...

Why is this useful?

- More information!
- Information is **connected**
- Information is more **varied**
- Information to **complement** existing metadata
 - Multiple readings
 - Multiple scholarly interpretations
- **Search and discovery**
 - Create new entry-points into your data
 - Establish link to other online sources & gain new metadata

Recent Additions & Outlook



- IIF
- TEI “roundtrip” & Named Entity Recognition
- Relationship annotation

Outlook & desiderata

- SKOS tagging
- Drawing tools
- Interop scenarios (Omeka, FromThePage)
- IIF/geo

And Minerva said, "Father, son of Saturn, King of kings, if, then, the gods now mean that Ulysses should get home, we should first send Mercury to the Ogygian island to tell Calypso that we have made up our minds and that he is to return. In the meantime I will go to Ithaca, to put heart into Ulysses' son Telemachus; I will embolden him to call the Achaeans in assembly, and speak out to the suitors of his mother Penelope, who persist in eating up any number of his sheep and oxen; I will also conduct him to Sparta and to Pylos, to see if he can hear anything about the return of his dear father- for this will make people speak well of him."

A diagram illustrating semantic relationships between entities in the text. Dashed arrows with labels connect entities: "Ulysses" is connected to "Telemachus" via a "father of" relationship; "Ulysses" is connected to "home" via an "is" relationship; "Ulysses" is connected to "Ithaca" via a "goes to" relationship; "Telemachus" is connected to "Penelope" via a "goes to" relationship; "Telemachus" is connected to "Sparta" via a "goes to" relationship; and "Telemachus" is connected to "Pylos" via a "goes to" relationship.

Making Recogito Sustainable



- Partnerships, partnerships, partnerships...
 - Member at the **DARIAH EU GeoHumanities WG**
 - Close collaboration with **University of Pittsburgh World History Center**
 - Working relations within our community
NYU, Harvard, Stanford, Lancaster U, Tufts U, German Archaeological Institute, ÖAW, American Numismatic Society, British Library, Lund U, Europeana, Mainz, Digital Mappa project, Lab1100, Performant Software, Eldarion, Fitzwilliam Museum,...
- Ties with infrastructure initiatives (e.g. DARIAH)

- **Grow our Community**
 - Training & workshops
 - Embed Recogito in teaching
- **Seed a tech community**
 - Self-host your own copy
 - Configure and tweak Recogito
 - Find your way around the code & contribute
- **Learn from you!**
 - Find new uses we haven't thought of
 - Engage with new communities
 - Work with them to enhance, extend and disaggregate Recogito



Learn more...



Recognito |  AN INITIATIVE OF
Pelagios
commons

The best way to learn about Recognito is to try it!

- 10 minute tutorial: <http://recognito.pelagios.org/help/tutorial>
- Source code & self-installation: <http://github.com/pelagios/recognito2>
- Project website: <http://commons.pelagios.org>

Grateful Acknowledgement to the Andrew W. Mellon Foundation
and all [@Pelagiosproject](#) Partners