



Recogito

Tiziana Mancinelli, Simone Rebora, Rachele Sprugnoli

ASSOCIAZIONE PER
L'INFORMATICA UMANISTICA
E LA CULTURA DIGITALE



What is Recogito?

- Free online **annotation tool** and **publishing** platform
- Semi-automatic annotation of place references in maps and texts. It was developed as part of the Pelagios 3 research project
- Semantic annotation Linked Open Data
- Allow to create collaborative project and simultaneous annotations of the same document
- First developed in 2014, the community-oriented philosophy behind **Pelagios Commons** has made users an active agent in shaping its functionality and interface.

What is Recogito?

- Recogito allows people to create semantic annotation without worrying too much about the formal language and structure of Linked Open Data.

Why? What is a semantic annotation?

Producing semantic annotations usually requires a certain amount of knowledge of digital technologies such as RDF, ontologies and/or text encoding. These techniques can sometimes act as a barrier for users that are not already familiar with Semantic Web theory.

Recogito motto:

“Linked Data annotation without the pointy brackets”

Linked (Geo) Data Annotation

Definitions

- LINKED DATA

Connecting structured data on the Web / Linked Data is simply about using the Web to create typed links between data from different sources. (Bizer et al., 2011)

- LINKED GEO-DATA:

The text annotation area to demarcate place names in digital text, an image annotation area to mark up and transcribe place names on map or manuscript scans, and a geo-resolution area, where the identified (and transcribed) place names are mapped to a gazetteer (and, thus, to geographical coordinates).

Gazeeters

What is a gazeteer? What are the gazeeters in Recogito?

A gazetteer is a geographical dictionary or directory used in conjunction with a map or atlas. It typically contains information concerning the geographical makeup, social statistics and physical features of a country, region, or continent (...) - Wikipedia

- ✓ **Pleiades** Pleiades Gazetteer of the Ancient World
- ✓ **CHGIS** China Historical GIS
- ✓ **DPP Places** Places from the Digitizing Patterns of Power project
- ✓ **DARE** Digital Atlas of the Roman Empire
- ✓ **GeoNames** A subset of GeoNames populated places, countries and first-level administrative divisions
- ✓ **MoEML** Map of Early Modern London
- ✓ **HGIS de las Indias** Historical-Geographic Information System for Spanish America (1701-1808)



Example: <https://sites.fas.harvard.edu/~chgis/>

Connecting Knowledge

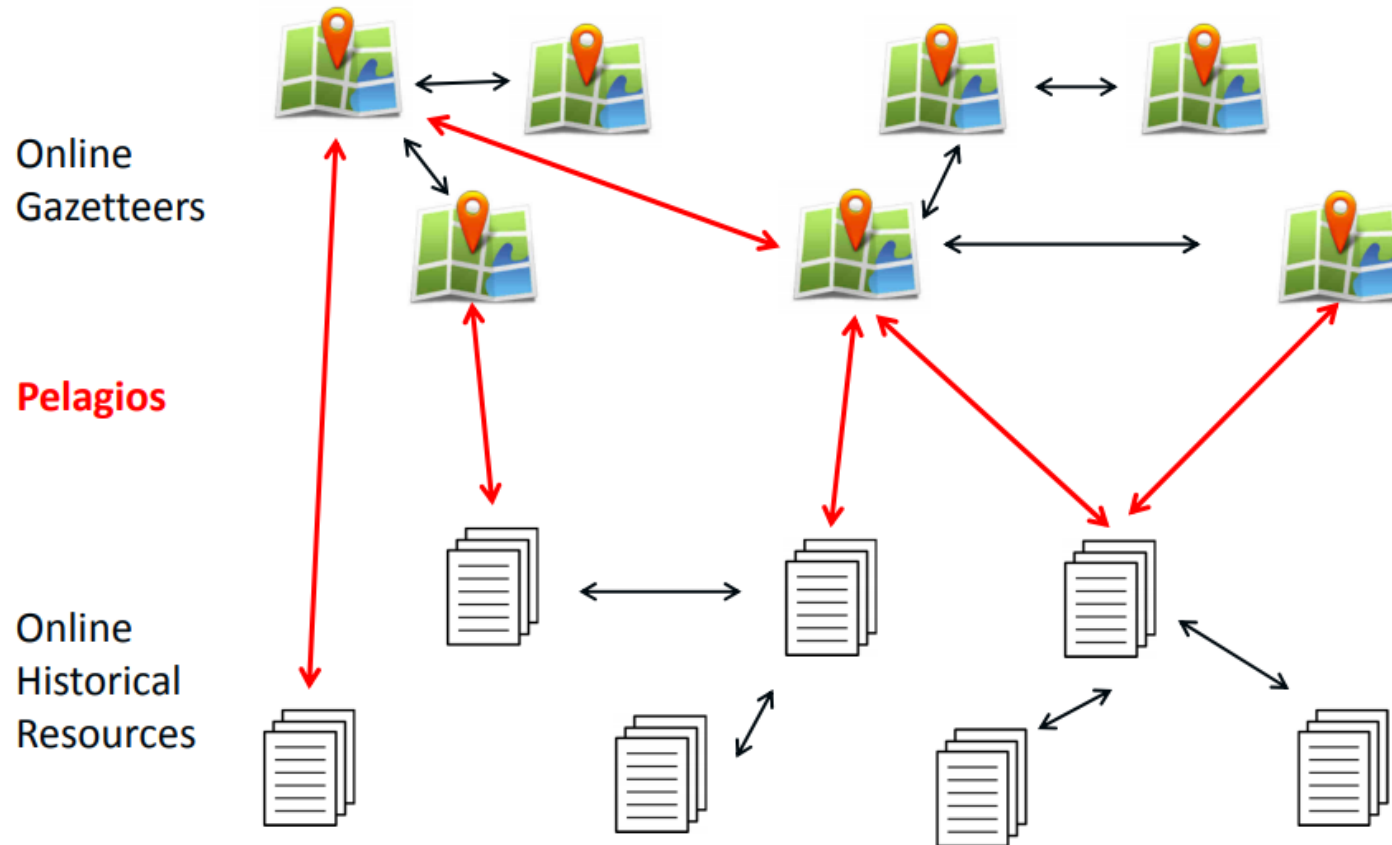
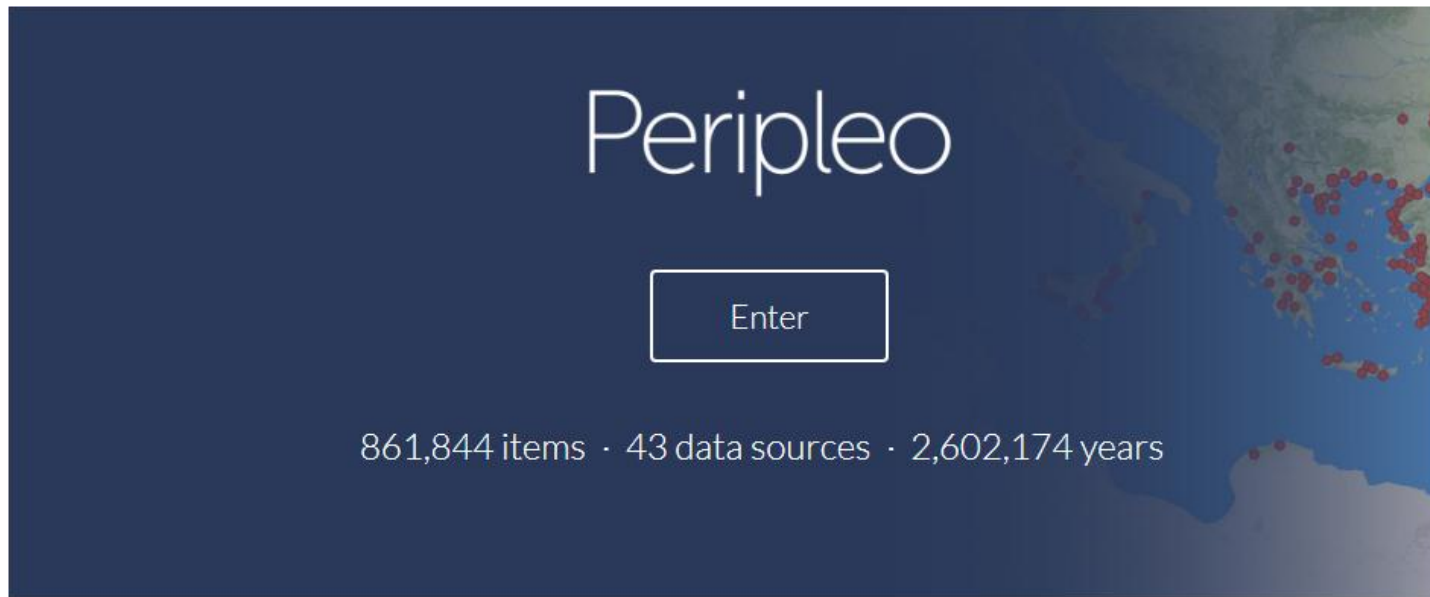


Image taken from: <https://f-origin.hypotheses.org/wp-content/blogs.dir/2899/files/2017/04/2017-04-Recogito-Paris.pdf>

From Annotation to Exploration

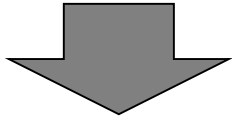
- From Recogito to Peripleo: <https://peripleo.pelagios.org/>
- Example “Herodotus: The Histories”:
<https://recogito.pelagios.org/document/tjrrsqn4dwmgep>



How Recogito Works

Step by step with Recogito

- Uploading a file - TXT (it is the only choice)



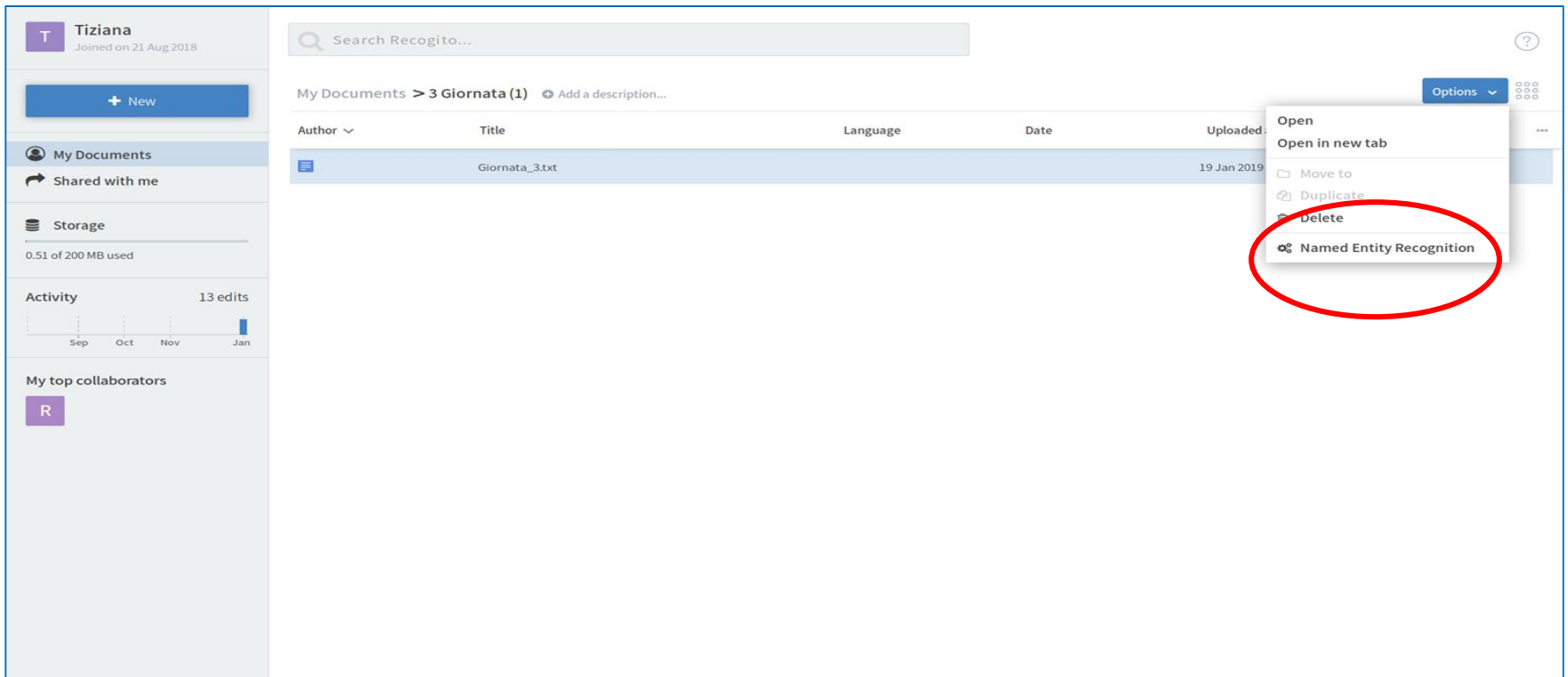
Search Recogito...

Shared with me (1)

Author	Title	Language	Date	Uploaded at	Last edit	...
Giovanni Boccaccio	Decameron Seconda Giornata	IT		31 Dec 2018	1 week ago	

Annotating Documents

- Automatic annotation - Named Entity Recognition (only EN)



The screenshot displays the Recogito web interface. On the left, a sidebar shows the user profile 'Tiziana' (joined 21 Aug 2018), a '+ New' button, and navigation links for 'My Documents', 'Shared with me', and 'Storage' (0.51 of 200 MB used). Below this is an 'Activity' section with a bar chart showing 13 edits from Sep to Jan, and a 'My top collaborators' section with a user 'R'.

The main area features a search bar 'Search Recogito...' and a breadcrumb 'My Documents > 3 Giornata (1)' with an 'Add a description...' link. A table lists documents with columns: Author, Title, Language, Date, and Uploaded. One document is listed: 'Giornata_3.txt' uploaded on '19 Jan 2019'. A context menu is open over this document, showing options: 'Open', 'Open in new tab', 'Move to', 'Duplicate', 'Delete', and 'Named Entity Recognition'. The 'Named Entity Recognition' option is highlighted with a red circle.

Author	Title	Language	Date	Uploaded
	Giornata_3.txt			19 Jan 2019

Step by step with Recogito

- Double-click on the text and you are in the working space for annotations for text

The screenshot displays the Recogito web interface. On the left, a dark blue sidebar shows the user profile 'rachelesprugnoli' (joined on 6 Mar 2018) and a list of document files, with 'seconda_giornata_1.txt' selected. The main workspace has a top navigation bar with icons for document, location, list, download, and share. The document title is 'Giovanni Boccaccio: Decameron Seconda Giornata', with 51 annotations and no other contributors. Below the title, the 'ANNOTATION MODE' is set to 'NORMAL', with other options like 'QUICK', 'RELATIONS', 'COLOUR', 'BY ENTITY TYPE', 'BY VERIFICATION STATUS', and 'BY FIRST TAG'. The text content is displayed with several entities highlighted in yellow: 'Martellino', 'Arrigo', 'Trivigi', 'Stecchi', and 'Marchese'. The text is in Italian, describing events from Boccaccio's Decameron.

rachelesprugnoli
Joined on 6 Mar 2018

seconda_giornata_1.txt
seconda_giornata_2.txt
seconda_giornata_4.txt
seconda_giornata_3.txt
seconda_giornata_5.txt
seconda_giornata_6.txt
seconda_giornata_7.txt
seconda_giornata_8.txt
seconda_giornata_10.txt
seconda_giornata_9.txt

Giovanni Boccaccio: Decameron Seconda Giornata
51 Annotations · No Other Contributors

ANNOTATION MODE: NORMAL QUICK RELATIONS COLOUR BY ENTITY TYPE BY VERIFICATION STATUS BY FIRST TAG

[1]
Martellino, infignendosi attratto,
sopra santo Arrigo fa vista di guerire e,
conosciuto il suo inganno, è battuto e poi preso;
e in pericol venuto d'essere impiccato per la gola,
ultimamente scampa.

Spesse volte, carissime donne, avvenne che chi altrui sé di beffare ingegnò, e massimamente quelle cose che sono da reverire, s'è con le beffe e talvolta col danno sé solo ritrovato. Il che, acciò che io al comandamento della reina ubidisca, e principio dea con una mia novella alla proposta, intendo di raccontarvi quello che prima sventuratamente e poi, fuori di tutto il suo pensiero, assai felicemente a un nostro cittadino adivenisse. Era, non è ancora lungo tempo passato, un tedesco a Trivigi chiamato Arrigo, il quale, povero uomo essendo, di portare pesi a prezzo serviva chi il richiedeva; e, con questo, uomo di santissima vita e di buona era tenuto da tutti. Per la qual cosa, o vero o non vero che si fosse, morendo egli adivenne, secondo che i trivigiani affermavano, che nell'ora della sua morte le campane della maggior chiesa di Trivigi tutte, senza essere da alcun tirate, cominciarono a sonare. Il che in luogo di miracolo avendo, questo Arrigo esser santo dicevano tutti; e concorso tutto il popolo della città alla casa nella quale il suo corpo giacea, quello a guisa d'un corpo santo nella chiesa maggior ne portarono, menando quivi zoppi, attratti e ciechi e altri di qualunque infermità o difetto impediti, quasi tutti dovessero dal toccamento di questo corpo divenir sani. In tanto tumulto e discorrimiento di popolo, avvenne che in Trivigi giunsero tre nostri cittadini, de' quali l'uno era chiamato Stecchi, l'altro Martellino e il terzo Marchese, uomini li quali, le corti de' signor visitando, di contraffarsi e con nuovi atti contraffacendo qualunque altro uomo li veditori

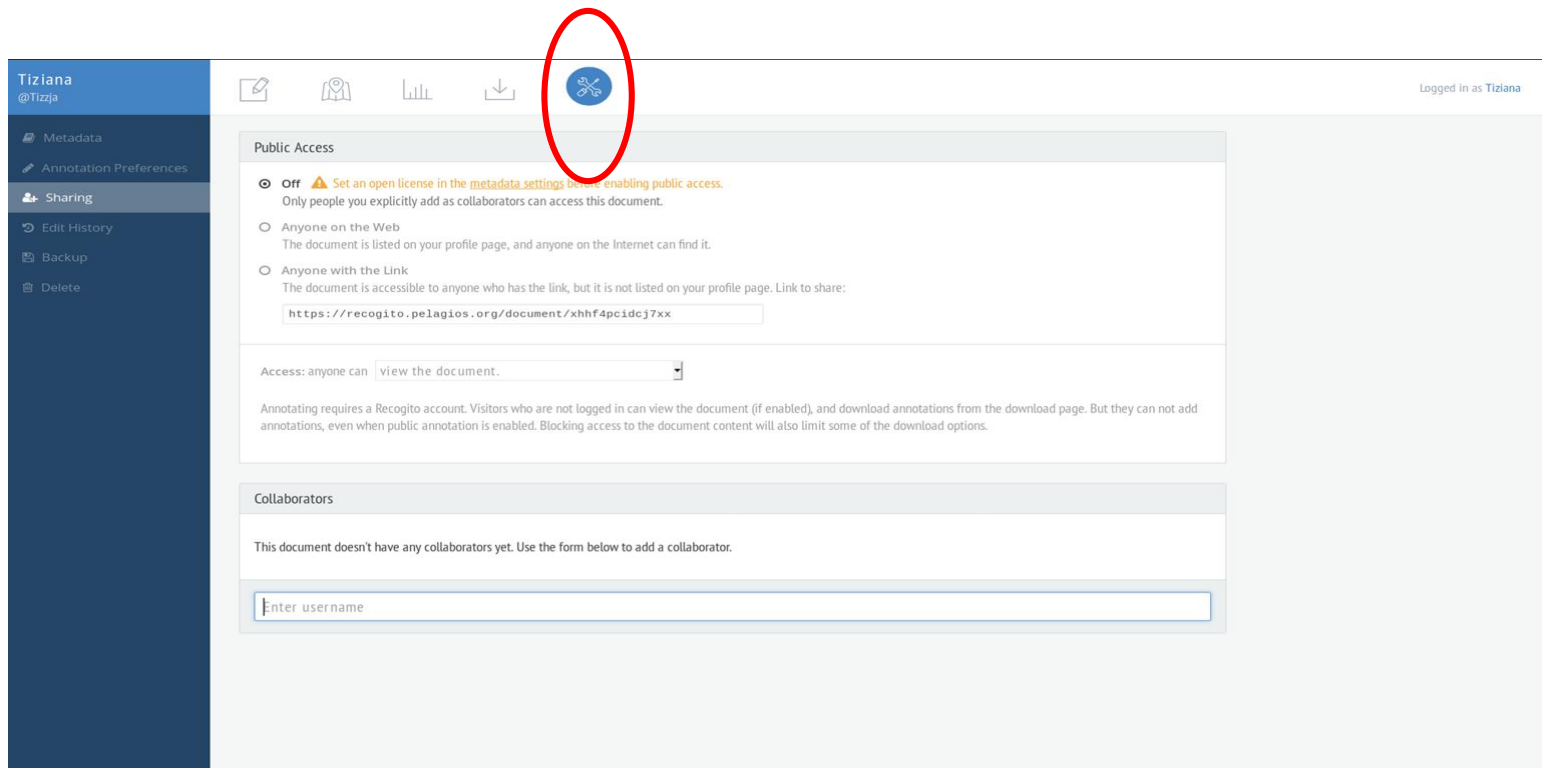
Adding metadata for your document

- The new interface does currently have the option to do it:
- Go to “Document Setting”

The screenshot shows the Tiziana user interface for editing document metadata. On the left is a dark blue sidebar with the user's name 'Tiziana @Tizja' and a menu containing 'Metadata', 'Annotation Preferences', 'Sharing', 'Edit History', 'Backup', and 'Delete'. The main content area is titled 'Document Metadata' and contains several input fields: 'Title' (filled with 'Crusades_1905.jpg'), 'Author', 'Date', 'Description', 'Language', 'Source', 'Edition', 'License' (a dropdown menu showing '-- Select a license --'), and 'Attribution' (with placeholder text 'Extra attribution statement or license information'). A blue 'Save Changes' button is at the bottom of the form. Below the metadata section is a 'Document Parts (1 file)' section with a drag handle and the text 'Drag to change order'. It lists the file 'Crusades_1905.jpg' with an 'Edit Part Metadata' button next to it. The top right of the interface shows 'Logged in as Tiziana'.

How to share a document with collaborators

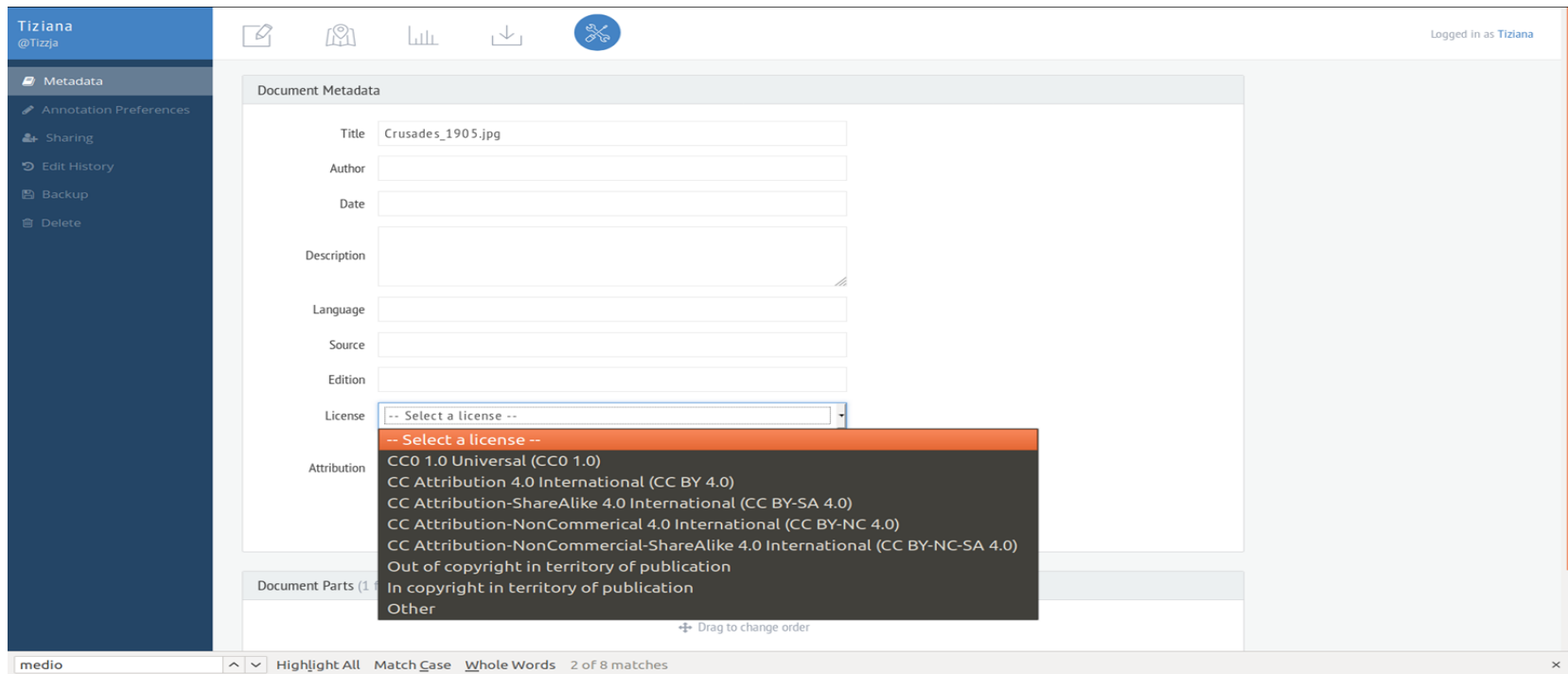
- In “setting document”, go to > sharing > you can also share a document with other collaborators.



The screenshot shows the Recogito document settings interface. On the left is a dark blue sidebar with the user's name 'Tiziana @Tizzja' and a list of menu items: 'Metadata', 'Annotation Preferences', 'Sharing' (highlighted), 'Edit History', 'Backup', and 'Delete'. The main content area has a top toolbar with icons for document, user, chart, download, and a red circle around a share icon. Below the toolbar, the 'Public Access' section is visible, with a warning: 'Set an open license in the metadata settings before enabling public access.' It offers three options: 'Off' (selected), 'Anyone on the Web', and 'Anyone with the Link'. A link to share is provided: 'https://recogito.pelagios.org/document/xhhf4pcidcj7xx'. Below this is an 'Access:' dropdown set to 'anyone can view the document.' and a note about annotation requirements. The 'Collaborators' section below states 'This document doesn't have any collaborators yet. Use the form below to add a collaborator.' and includes a text input field labeled 'Enter username'.

How to share a document with collaborators

- By default your document is private (only you can see your work)
To make your document public → go to the metadata



Annotating Documents

- Annotated manually in the text annotation area
- Recogito also includes a “transcription assistant”, which automatically suggests potential matches based on the proximity to other, already transcribed, toponyms
- Re-apply the same annotation function:

“There are Ntimes other annotated occurrences of Name in the text. Do you want to re-apply this annotation?”

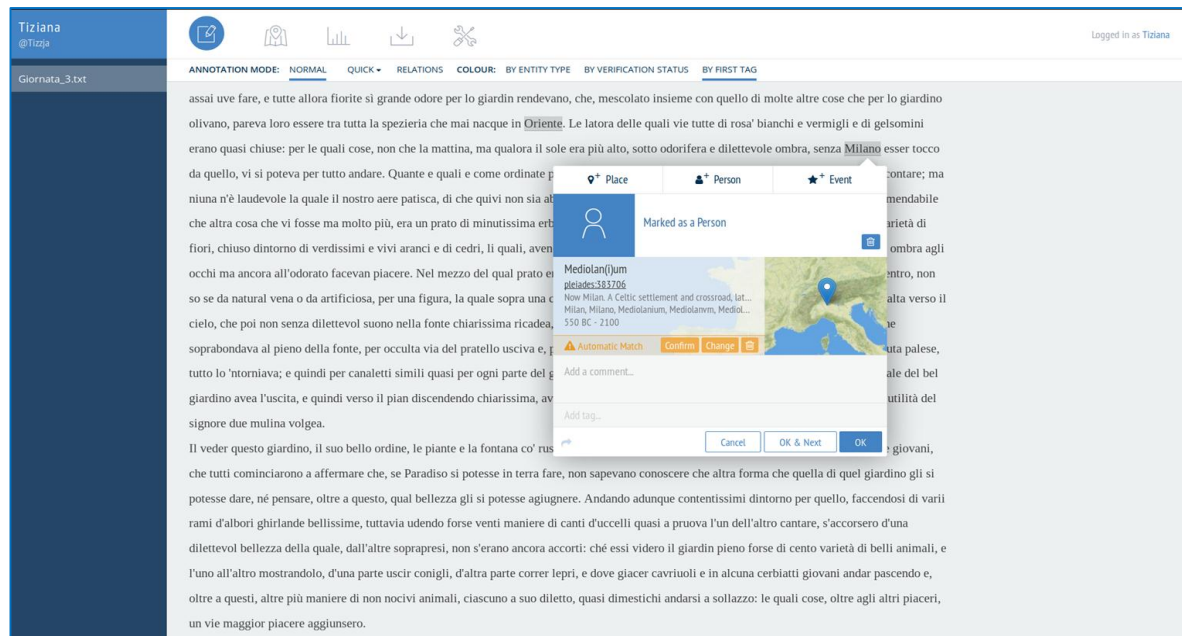
Annotating Documents

- Annotated manually in the text annotation area
- Sometime it doesn't really guess what you want!

The screenshot shows the Tiziana text annotation interface. On the left, the user's name 'Tiziana @Tizzja' and the document name 'Giornata_3.txt' are visible. The main area displays a text document with a paragraph about a garden. The word 'Oriente' is highlighted in the text. A modal window is open over the text, showing a map of San Rafael Oriente with a location pin. The modal includes tabs for 'Place', 'Person', and 'Event'. Below the map, there are buttons for 'Automatic Match', 'Confirm', 'Change', and 'Add a comment...'. At the bottom of the modal, there are buttons for 'Cancel', 'OK & Next', and 'OK'. The text in the background is partially obscured by the modal window.

Annotating Documents

- But when Recogito guesses that word is a place, it matches with the digital gazetteers that are integrated with Recogito.
- In this case, we can geo resolve an annotation, we have associated specific geo coordinates



Annotating Documents

- You can confirm the automatic guess or change it.
- If you don't confirm, the status of that geo-reference is set to “not verified”
- Some places are “unlocated”: Inghilterra
- Not all granularity levels are available (no streets or neighborhoods)

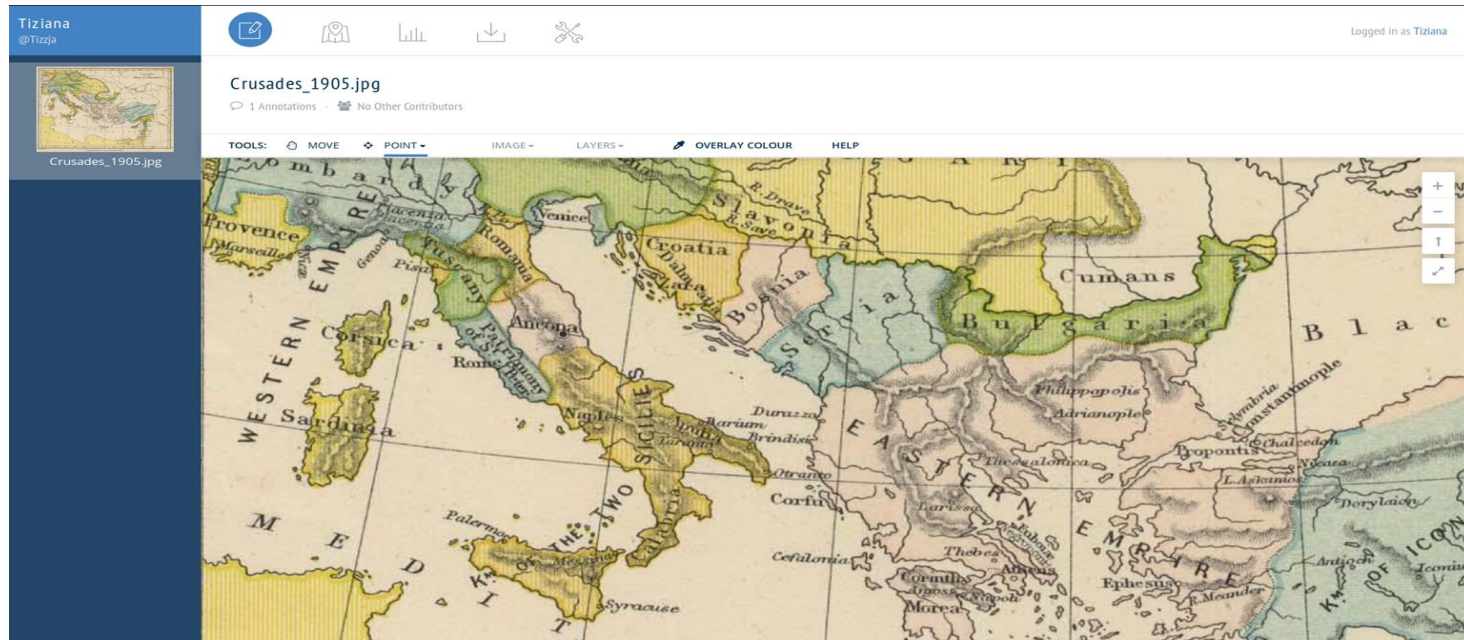


Not only places but people and events

- Very similar the way to annotate or even simpler. We can add **comments** and **tags**.
- What is important that information about “tags” - can be downloaded and used for our analysis.
- Different tags = different colors

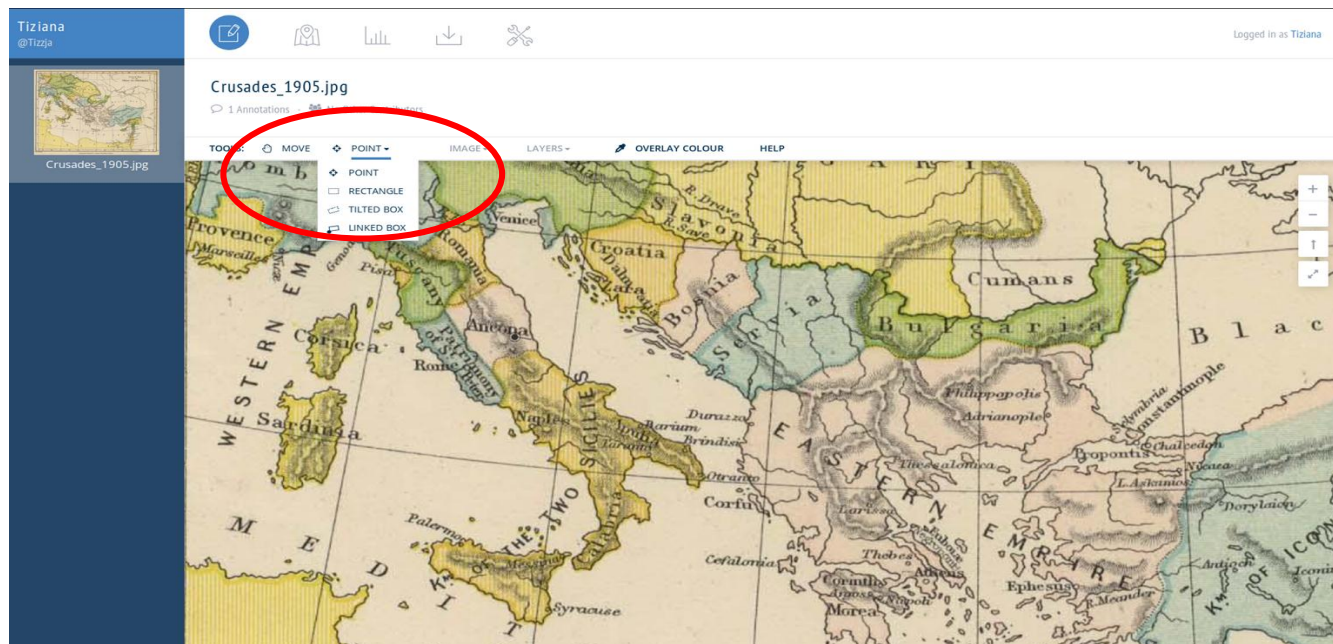
Images

- Uploading images is more or less the same as uploading text
- Point the place to the image



Images

- We can define the place in different ways: point, rectangle, titled box, linked box
- You can transcribe the name of the map



Images and gazeeters

Keyboard Shortcuts

SPACE toggle between MOVE and current annotation tool

ALT + V toggle annotation visibility

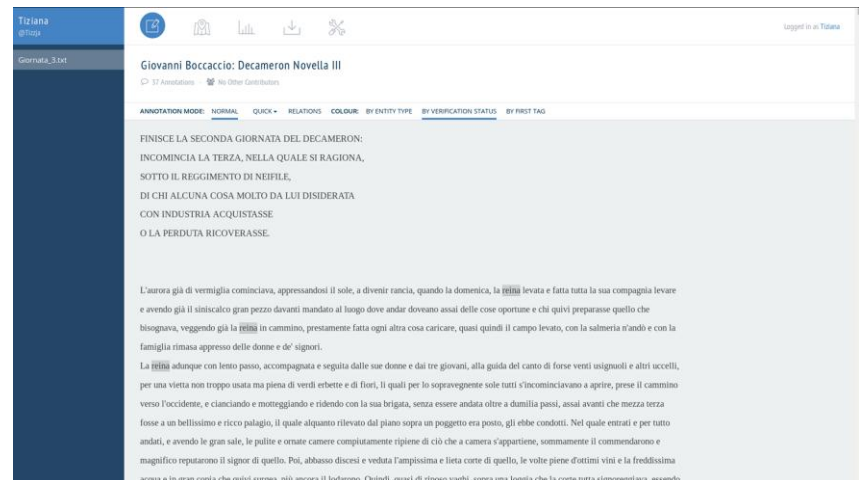
Map Navigation

Hold SHIFT to drag a box and zoom to the area

Hold ALT to rotate the map around the center of the screen

New Function: Relation Annotation

- Annotating relations among characters:



Tools, analysis and publishing

From Texts to Models to Visualizations



The screenshot displays the Brat NLP interface. On the left, a sidebar lists files: `rachelesprugnoli` (Joined on 6 Mar 2018) and several `seconda_giornata_*.txt` files. The top navigation bar contains icons for document, map, bar chart, download, and edit. Two red circles highlight the map and download icons. The main text area shows the title **Giovanni Boccaccio: Decameron Seconda Giornata** with 51 annotations and no other contributors. The annotation mode is set to **NORMAL**. The text being annotated is:

[1]
Martellino, infingendosi attratto,
sopra santo Arrigo fa vista di guerire e,
conosciuto il suo inganno, è battuto e poi preso;
e in pericol venuto d'essere impiccato per la gola,
ultimamente scampa.

Spesse volte, carissime donne, avvenne che chi altrui sé di beffare ingegnò, e massimamente quelle cose che sono da reverire, s'è con le beffe e talvolta col danno sé solo ritrovato. Il che, acciò che io al comandamento della reina ubidisca, e principio dea con una mia novella alla proposta, intendo di raccontarvi quello che prima sventuratamente e poi, fuori di tutto il suo pensiero, assai felicemente a un nostro cittadino adivenisse. Era, non è ancora lungo tempo passato, un tedesco a Trivigi chiamato Arrigo, il quale, povero uomo essendo, di portare pesi a prezzo serviva chi il richiedeva; e, con questo, uomo di santissima vita e di buona era tenuto da tutti. Per la qual cosa, o vero o non vero che si fosse, morendo egli adivenne, secondo che i trivigiani affermavano, che nell'ora della sua morte le campane della maggior chiesa di Trivigi tutte, senza essere da alcun tirate, cominciarono a sonare. Il che in luogo di miracolo avendo, questo Arrigo esser santo dicevano tutti; e concorso tutto il popolo della città alla casa nella quale il suo corpo giacea, quello a guisa d'un corpo santo nella chiesa maggior ne portarono, menando quivi zoppi, attratti e ciechi e altri di qualunque infermità o difetto impediti, quasi tutti dovessero dal toccamento di questo corpo divenir sani. In tanto tumulto e discorrimento di popolo, avvenne che in Trivigi giunsero tre nostri cittadini, de' quali l'uno era chiamato Stecchi, l'altro Martellino e il terzo Marchese, uomini li quali, le corti de' signor visitando, di contraffarsi e con nuovi atti contraffacendo qualunque altro uomo li veditori






Analyzing my Data

- How to export data from Recogito in a variety of formats suitable for visualizing and analysis in other tools, such as spreadsheets, databases or GIS
- Visualising and contextualizing geographical information within documents can be an important step in reaching a deeper understanding of their content, potentially highlighting phenomena that would have been otherwise difficult to identify

Creating different output

Tizzja

Joined on 21 Aug 2018




Logged in as

Annotations


CSV


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


 CSV

RDF

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

 JSON-LD


 RDF/Turtle

 RDF/XML

Places


GeoJSON

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

 GeoJSON

KML


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

 KML

Annotated Document

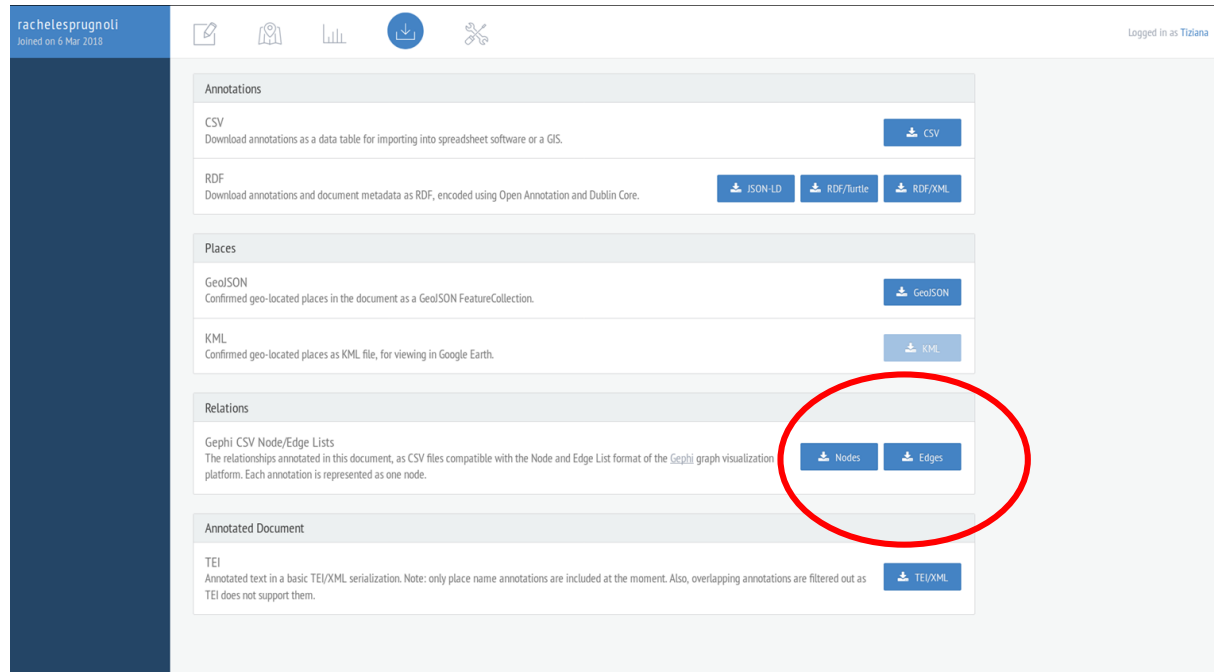
TEI

Annotated text in a basic TEI/XML serialization. Note: only place name annotations are included at the moment. Also, overlapping annotations are filtered out as TEI does not support them.

 TEI/XML

Gephi

- Download the two files “edges” and “Nodes”



Open Gephi and import:

- Edges
- Nodes

N.B. Post-processing needed to create correct networks...

Conclusion

- Allows to collaborate on the same project
- It is a very useful tool for teaching
- Allows to compare different sources
- People with different level of technical expertise to work with it

Methodological open issues

- How to deal with:
 - Ambiguity: e.g. places that are also people (and vice versa)
 - Imaginary places
 - Imaginary people or like the name Madonna
 - Places that you don't know if are existed or ...
- How to choose the right gazetteer?
- Do we annotate only named entities?
“Fu già nella nostra città” = Firenze

Hands-on Session

Boccaccio

Decameron collection of tales by Giovanni Boccaccio, probably composed between 1349 and 1353.

