# Recogito-in-a-Box: From Annotation to Digital Edition.

## del Rio Riande, Gimena y Vitale, Valeria.

Cita:
del Rio Riande, Gimena y Vitale, Valeria (2020). *Recogito-in-a-Box: From Annotation to Digital Edition. , 1.*

**ARTICLE – DIGITAL MODERN LANGUAGES**

# Recogito-in-a-Box: From Annotation to Digital Edition

Gimena del Rio Riande and Valeria Vitale

¹ Consejo Nacional de Investigaciones Científicas y Técnicas, AR

² University of London, GB

Corresponding author: Gimena del Rio Riande (gdelrio.riande@gmail.com)

Through the combination of two popular approaches in the Digital Humanities – digital editions and semantic annotation – this tutorial will present simple ways to create, analyse and export semantic annotations from texts and images, and publish them online. It will introduce intuitive, user-friendly, open-source tools interwoven in an integrated workflow from Recogito – a free online semantic annotation tool developed by the Pelagios Network – to documents encoded according to the Text Encoding Initiative (TEI) standard. With this tutorial, users interested in semantic annotation and digital editions will learn how to benefit from Recogito's automatic recognition of named entities, and how to refine them manually, checking the place references against historical gazetteers. They will learn how to create annotations *ex novo*, check or modify annotations identified by Recogito, and discover how the geo-annotations produced on the text can then be plotted on a digital map. Finally, users will learn how to use Recogito's export options and, in particular, the TEI format, which will become the starting point of a TEI-based simple minimal edition. As a case study, it will focus on the semantic and geographic annotation of an early Argentinian chronicle called *Historia de la Conquista del Río de la Plata*, better known as *La Argentina Manuscrita*, written by Ruy Diaz de Guzman in the early seventeenth century.

## 1. Context

This tutorial presents a free and easy-to-use digital workflow for annotating historical sources, with a particular focus on mapping the geography of texts. If you are interested in tracing itineraries, investigating migration or trading routes, or comparing accounts of the same places between different authors and media and/or across different times and cultures, Recogito-in-a-box could be for you.[1] In this tutorial, you will learn about the value of digital gazetteers,

---

[1] Further, a 10 minute tutorial for this tool can be found at: https://recogito.pelagios.org/help/tutorial, and a an FAQ at: https://recogito.pelagios.org/help/faq.

semantic annotation, map-based visualisations,[2] and digital editions using Recogito, an open-source, free and online semantic annotation tool developed by the Pelagios Network.[3]

To help you see how Recogito works in practice, we will use a specific case study – the *Historia de la Conquista del Río de la Plata*, best known as *La Argentina Manuscrita* – an early seventeenth-century chronicle written by a Spanish-Guarani officer, Ruy Díaz de Guzmán (del Rio Riande et al. 2019a). It is in *La Argentina Manuscrita* that, among others, we find the first description of the Rio de la Plata region in Spanish.[4]

The workflow demonstrated here includes technologies such as TEI-XML, the standard mark-up for text in the Humanities developed by the Text Encoding Initiative (TEI) Consortium[5] and Linked Open Data (LOD)[6] – but the good news is that you don't need to have any prior computing expertise related to these technologies to use Recogito.[7]

## 2. Procedure

### Step 1: Create an account

First of all, create an account on Recogito.[8] It's completely free, the user interface is available in several languages, including English, Spanish, German and Italian, and it works with most supported browsers (Firefox, Chrome, Safari). Once you have your account, you can start annotating straight away. This global version of Recogito doesn't require any installation. However, since Recogito is open-access software, you can also download a local version of it to customise, for example, by adding an additional gazetteer specific to your needs.

You can also annotate a document that another user has shared with you. We will discuss this possibility when talking about sharing options (see Step 8).



**Figure 1:** Create an account.

---

[2] A digital gazetteer is a list of places that often includes place names, coordinates and sometimes feature types for each place. More about digital gazetteers, semantic technologies, spatial search and web visualisations can be found in Berman et al. (2016).

[3] More about the newly born Pelagios Network association at: https://pelagios.org/.

[4] For details about the text and the digital edition of *La Argentina Manuscrita* see: http://hdlab.space/La-Argentina-Manuscrita/argentina-manuscrita-acerca-de/. The annotated text in *Recogito* is publicly available from: https://recogito.pelagios.org/document/wzqxhk0h3vpikm.

[5] You can learn more about this initiative and its guidelines and tools at: https://www.tei-c.org/. In the past few years, other technologies, such as GIS or LOD have interacted with TEI encoding, and proved to be a useful way of enriching texts. The title of the 2019 TEI conference – *What is text, really?: TEI and beyond* – and the contents of its Book of Abstracts give good account of this: https://graz-2019.tei-c.org/.

[6] LOD is structured open data expressed in machine-readable format and interlinked to other data, making semantic queries especially rich.

[7] More information about Recogito can be read at: Simon et al. (2017), (2019). A complete tutorial is Vitale & Simon (2019).

[8] Accessible at: https://recogito.pelagios.org/.

### Step 2: Upload a document

With Recogito you can annotate an array of digital documents (including image formats), but in this tutorial we just focus on text documents. To upload a text document to Recogito, we recommend using the .txt format. If your document is in another text format (e.g. .doc), you first need to convert it into the Unicode format UTF-8. You can do this in any of the most popular text editors, such as Word, Writer or Google Docs, simply by using the option *Save as*.

When you export the text file in .txt, please check that the text you are uploading is the final version: as Recogito is not a text editor, you won't be able to make changes on the text once it is uploaded. Although the optimum format for working on text documents in Recogito is currently .txt, Recogito now also has the capacity to enable the annotation of TEI-XML. We talk about this option later in the tutorial (see Step 7).

If you upload more than one text document at the same time, *Recogito* will collate the files to create a metadocument that brings them together. This function is particularly useful if you want to compare different chapters of the same book or if you wish to analyse accounts of the same trip or place in different authors.

### Step 3: Add your metadata

When you first upload a document, it is recommended that you fill in as much metadata about it as you can – information such as authorship, title and date of the text, and the provenance of its digital format, may all be important, particularly if you want to share the document. By default, all your documents will be visible to you only; if you want to share them with others (see Step 8), please be sure that you have the appropriate permissions and that you have supplied this in the metadata. You will only be able to share a document if you own the copyright to it or if it is under a Creative Commons license.[9]

### Step 4: Create annotations

Creating annotations in Recogito is simple and enjoyable. You simply select the word or words in the text that you wish to annotate. This action will bring up a small annotation pop-up window, which asks you to assign a category to your annotation. You can choose between three different categories: *Place*, *Person* and *Event*.

If you click on *Place*, Recogito will try to help you disambiguate your annotation by matching it to related entries from one or more global authority records for places through its gazetteers.[10] Recogito currently uses seven historical gazetteers – Pleiades (gazetteer of the
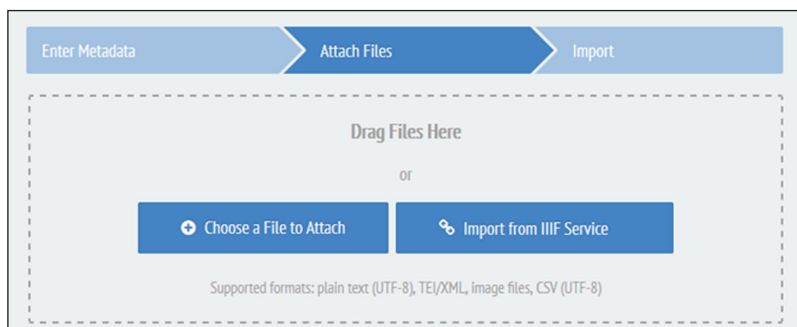


**Figure 2:** Upload a document.

Ancient World), CHGIS (China Historical GIS), DPP Places (Places from the Digitizing Patterns of Power Project), DARE (Digital Atlas of the Roman Empire), MoEML (Map of Early Modern London), HGIS de las Indias (Historical-Geographic Information System for Spanish America, 1701–1808), Kima (historical gazetteer with place names in Hebrew script) – as well as a contemporary one (a subset of Geonames). It remains up to you to choose the place record that you think best fits the place mentioned in the text that you are annotating. There is an added bonus of aligning your place annotation to a global place record: because gazetteers also provide other information (such as coordinates), Recogito will automatically visualise your place annotations on a map.

We will say a little more about visualisation options in Step 6. You are also able to mark your annotations as *People* or *Events*, though currently – given the lack of global authority records on these entities – you won't be able to disambiguate them using unique identifiers as you can do with places. You can even use tags and free text comments to further refine your annotations (see **Figure 4**), for example, by manually adding external identifiers to Wikidata or museum catalogues.
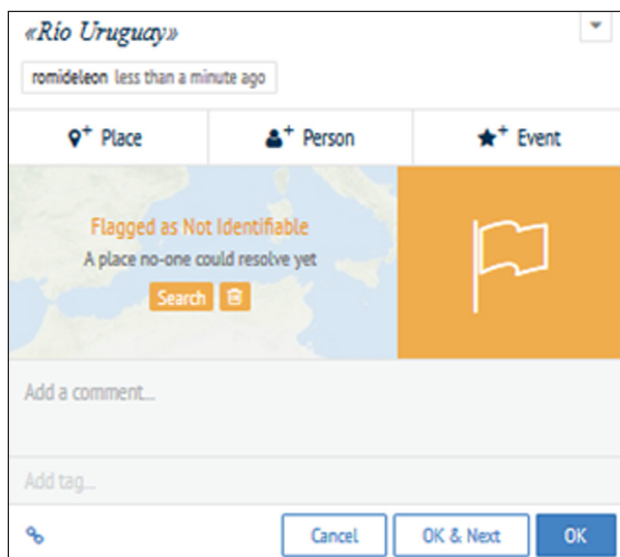


**Figure 3:** Place annotation.
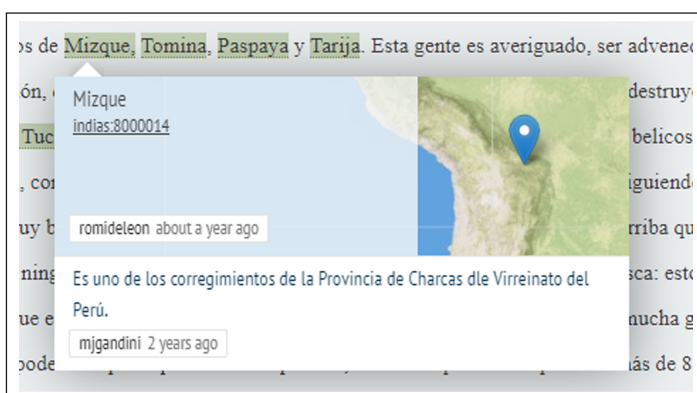


**Figure 4:** Automatic visualisation.

Recogito also offers the option of creating semi-automatic annotations using a Named Entity Recognition (NER) algorithm. NER algorithms are language specific, so you should select the most appropriate from those available in Recogito. At the moment, it offers NER in English, French, Spanish and German. More experimental NER algorithms are also available in Hebrew and Latin. If none of the available algorithms match the language of your text, you may try one that is linguistically close (for example, the Spanish algorithm for an Italian text). The algorithm parses the text and tries to identify all words that can be place names or person names. When the NER recognises a word as a possible place name, it will also try to match it automatically with an entry in one of Recogito's global gazetteers. These annotations will appear in grey highlight to reflect their automatic matching – to turn them green, a human user needs to confirm that: (i) the word is indeed a place; and (ii) it matches the particular place in the gazetteer.

Although, as noted above, there are several gazetteers in Recogito, the specific gazetteer most useful to the text we are using here as case study – *La Argentina Manuscrita* – is likely to be Indias, based on the HGIS de las Indias developed by Werner Stangl.[11] To narrow down the results from Recogito's automatic place matching, go to *Annotation Preferences* under *Document Settings*, and uncheck all those gazetteers that you don't think will be helpful.



**Figure 5:** Confirmed places.



**Figure 6:** NER options.

---

[11] Accessible at: https://www.hgis-indias.net/. More information about the Indias gazetteer can be found in Brumfield et al. (2018): https://medium.com/pelagios/interim-report-on-the-latam-gazetteer-of-colonial-latin-america-713613fb592c

Now it is time to work with our text, searching for places. In the example shown in **Figure 7**, we looked for Rio de la Plata (the River Plate). Recogito couldn't find the river or the region but it did find San Salvador, the first fort that Sebastián Gaboto founded in 1527.

Even though the user might not know at first hand that San Salvador is a fort located on the River Plate, if the (human) annotator checks this information beforehand this automated annotation might be very helpful. Here are some other examples from *La Argentina Manuscrita* to give you a sense of georeferencing in action, and to help guide your own annotation practice:

1) 'que los de Buenos Aires descubrieron por tierra el año de 605' (del Rio Riande et al. 2019a, Chapter 2).[12] Buenos Aires is in the Indias gazetteer, so I find the place easily:

However, our document doesn't refer to the city of Buenos Aires, but to the port of Buenos Aires. This is a detail that we can add in the free text comment section (see **Figure 8**).

2) 'la boca de este gran Río de la Plata, a quien los naturales llaman Paraná Guazú, que quiere decir río como mar' (del Rio Riande et al. 2019a, Chapter 1).[13] This example includes a name in both Spanish and in its Guaraní indigenous form, *Río de la Plata* and *Paraná Guazú* respectively. *Recogito* is unable to find the Guaraní name, since it
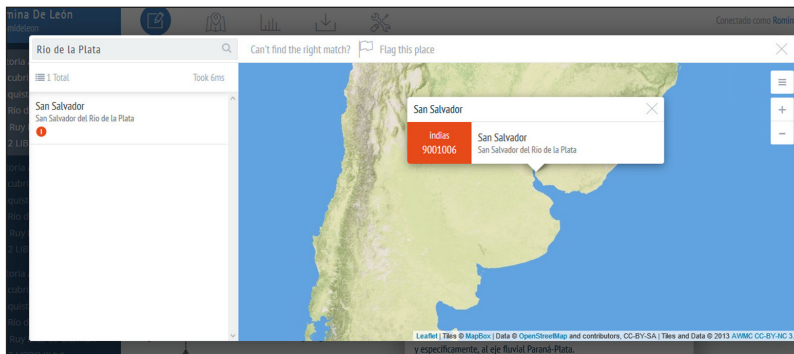


**Figure 7:** San Salvador del Río de la Plata.



**Figure 8:** Buenos Aires.

---

[12] In English: 'that the people from Buenos Aires discovered by the year of 605'.
[13] In English: 'The source of this big Plata river, that the indigenous call Paraná Guazú, that means river as sea'.

wasn't used in the historical-geographical dictionaries on which our specific gazetteer is based. (It doesn't even appear in Geonames.) We know from the text that the author is referring to the source of the river. Thus, after this intellectual step, we can make the match to San Salvador (as in San Salvador del Río de la Plata).

3) In some cases, the place has changed its name sometime over the last century. This is the case for Cabo de Santa María, which is nowadays known as La Paloma:

Again, the human annotator must know this information beforehand in order to make a decision. This information that relates the old and modern names can be added in the comments.

If you want to know more about semantic annotation in Recogito, have look at the video tutorial in YouTube, 'Annotate texts in Recogito/Anotar textos en Recogito' (del Rio Riande et al. 2019b):



**Figure 9:** Paraná Guazú.



**Figure 10:** La Paloma.
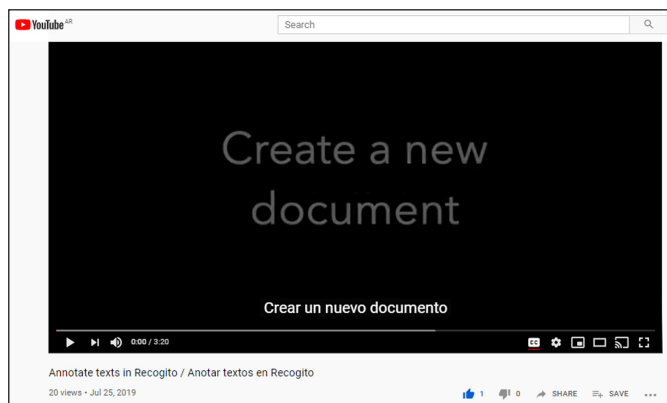


**Figure 11:** Create a new document, video tutorial.

### Step 5: Relations

There is one other kind of annotation that you can perform in Recogito. This is known as *relational tagging*, by means of which you can create a connection between entities, or relations between two existing annotations.

To mark relations between entities, switch Recogito's annotation mode to *Relations*, and then simply click on the first annotated entity and drag the pointer to the second. A dotted line will appear connecting the two annotations, along with a text box: you can fill this in to describe (or tag) the relationship. The line also has an arrow, which indicates the direction of the relationship. This is crucial for relationships that are hierarchical, as in, for example, *isPartOf* or *isDaughterOf*.

The relations created in Recogito can be exported in two formats: a basic CSV, and two separate tables for nodes and edges.[14] Both can be visualised in network analysis software such as Gephi.[15] If you are using the simplest option, just remember to change the denomination of the columns from *from_quote* to *source*, and *to_quote* to *target*, and simply upload the spreadsheet as an edge table (selecting the option *create missing nodes*). If you want to use the nodes and edges format instead, to have more control over your network visualisation, please bear in mind that, when downloaded in this format, each relation receives a different ID – and so the data will need consolidation before being processed in Gephi.[16]

In the example shown in **Figure 12**, we are marking the relationship between different South American ethnic groups: the Guayanás and the Guaraníes, also known (by Ruy Diaz) as Arachanes, and their enemies, the Charrúas:

When you have finished annotating entities (such as places or people) in your text document, and you have also marked the relations between them, your Recogito annotation screen will probably look something like this:
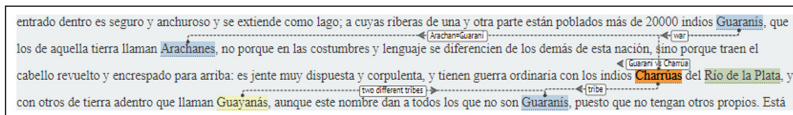


**Figure 12:** Relations.



**Figure 13:** Annotated text (semantic annotation and relations).

---

[14] In Graph Theory, an edge is a visual representation of a relation. It is a line that connects two nodes.

[15] Gephi: https://gephi.org/.

[16] For more information about annotating relations and the workflow with Gephi, please see https://github.com/pelagios/pelagios.github.io/wiki/Recogito-Tutorial:-Download-Options-for-Text.

**Step 6: Visualise the geographic annotations**

Annotations can be visualised and read as a continuum in the *Map View* option (upper menu). Just click on the arrow and move from site to site:

**Step 7: Download the text with the annotations**

While you can do all your annotation in Recogito, the platform has the great advantage of enabling you to download both your annotated texts and/or annotation data in a variety of different formats. This means that you can explore your annotations or do further editing of your text in other applications.

In the example below, we are downloading our annotated text in TEI-XML, the standard mark-up language for digital scholarly editions in Digital Humanities:



**Figure 14:** Visualisation of geographic annotations.
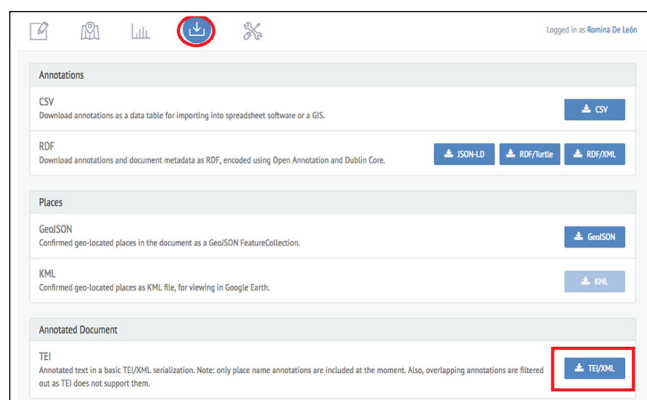


**Figure 15:** Download options: TEI-XML.

**Figure 16:** *La Argentina Manuscrita* TEI file downloaded from Recogito.



**Figure 17:** Minimal edition of *La Argentina Manuscrita*.

As well as UTF-8 plain text, Recogito now allows you to annotate TEI-XML documents. However, this functionality is still very much in beta mode. As a result, when you download your text, you will need to revise and rework the TEI (e.g. annotations currently overlap).[17] Nevertheless, since you can annotate entities (such as places and people) so easily in Recogito, and given the fact that there is also some basic encoding support for the text header and body, it is a great platform from which to start working on a digital edition and/or learning the basics of TEI encoding.

In the case of *La Argentina Manuscrita*, we downloaded, revised and corrected Recogito's TEI, and exported it to GitHub pages using Ed/Jekyll in a minimal edition format.[18]

If you want to know more about the download options in Recogito and the export in TEI-XML, watch this video tutorial (del Rio Riande et al. 2019c):[19]

### Step 8: Share the annotations (and credit)

Recogito allows you to work on your own or in a team. You can modify your sharing options, add collaborators and share your annotations in the tool section (upper menu). You can also follow up, backup your work or delete it.

---

[17]  The TEI guidelines for encoding a digital edition are available at: https://tei-c.org/release/doc/tei-p5-doc/en/html/SG.html.

[18]  About Minimal Editions with Ed/Jekyll, see: https://github.com/minicomp/ed.

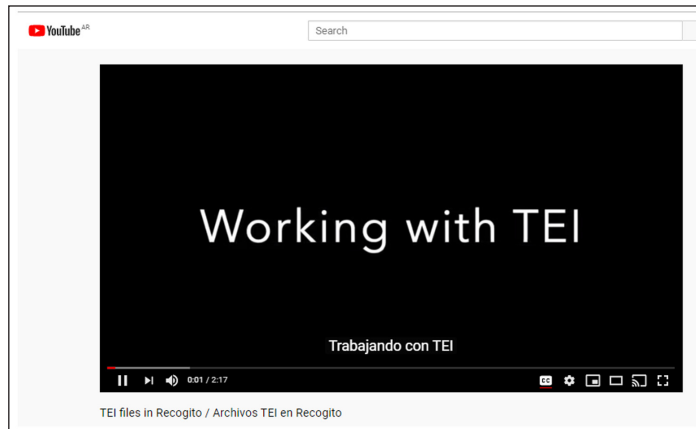[19]  Video accessible at: https://youtu.be/AgE3KcrxTfw.

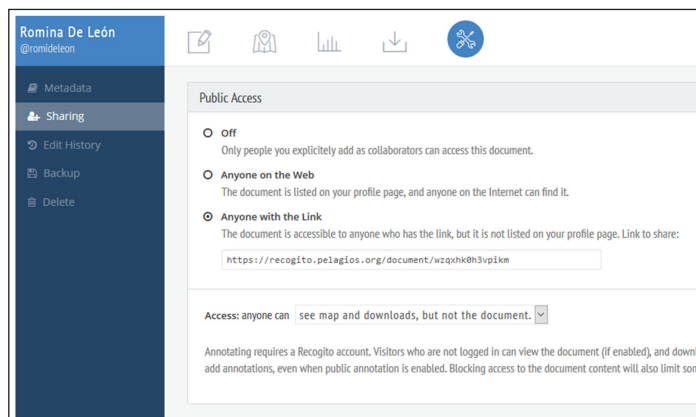**Figure 18:** Working with TEI files in *Recogito*, video tutorial.
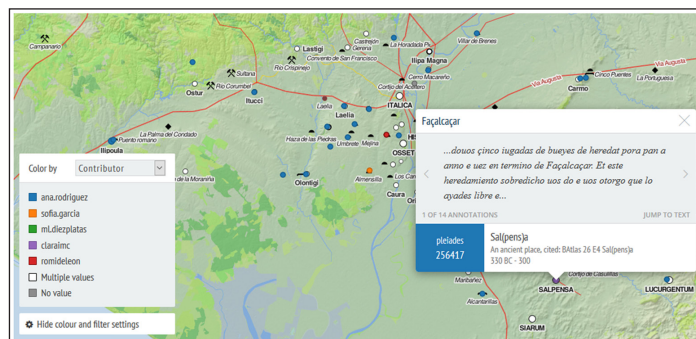


**Figure 19:** Sharing options.



**Figure 20:** Map view.

Remember that if you choose to leave your document open on the Web, it will be searchable by any engine and Google will index it. If you are doing teamwork, remember to use a proper name and email that give credit to all your collaborators. You will also be able to see your collaborators in the map view:
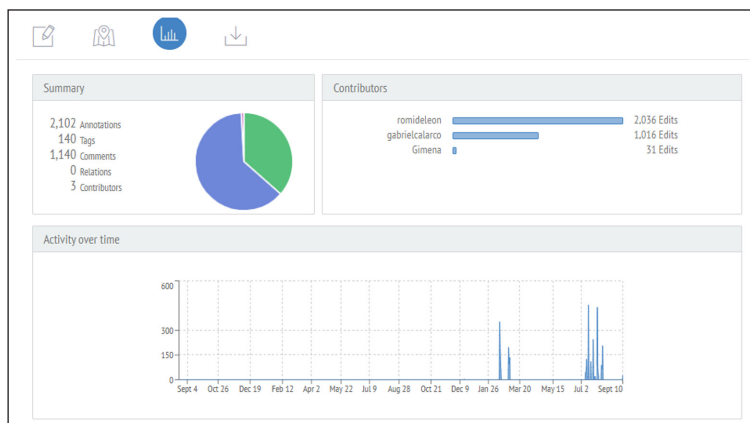
**Figure 21:** Stats.

And finally, don't forget you can also follow all the annotations in the Stats section:

## Author Information

Gimena del Rio Riande is a researcher at the Instituto de Investigaciones Bibliográficas y Crítica Textual of the National Scientific and Technical Research Council (CONICET. Buenos Aires, Argentina) and Professor at the University of Buenos Aires. MA and PhD in Romance Philology (Universidad Complutense de Madrid). Her main academic interests deal with Digital Scholarly Edition, the use, and methodologies of scholarly digital tools as "situated practices", Open Scholarly Communication and Open Research Practices in the Humanities, and the interaction of the global and the local in the development of academic disciplines. She has been working since 2013 with different DH communities of practice in Latin America and Spain, especially in Argentina, where she organized the first Digital Humanities Conference in 2014 and many other relevant events for this field.

Valeria Vitale is a Research Fellow at the Institute of Classical Studies in London, and the Chair of the Pelagios Network. She holds an MA and a PhD in Digital Humanities, and she specialises in spatial humanities and Linked Open Data, often investigating the synergies between the two. She has collaborated with several digital projects including The Pleiades Gazetteer, the Heritage Gazetteer of Libya, Cross-cultural After Life of Classical Site. She was Education Director for the Pelagios project and member of its Investigative Team. She teaches Digital Approaches to Cultural Heritage at the University of London.

## Videos

Videos from all the tutorials in this collection have been archived in a special playlist on the Digital Modern Languages YouTube channel at: https://www.youtube.com/playlist?list=PLfa B2f0CyBdu1OluU36KKLKXpKWDco62q.

Here is a list of videos associated with this tutorial:

- Annotate Texts in Recogito/Anotar Textos en Recogito https://www.youtube.com/watch?v=zx--3clSGM4
- TEI files in Recogito/Archivos TEI en Recogito https://www.youtube.com/watch?v=I7HM_HINMZU

## References

Brumfield, Ben. Interim Report on the LatAm Gazetteer of Colonial Latin America. 2018. https://medium.com/pelagios/interim-report-on-the-latam-gazetteer-of-colonial-latin-america-713613fb592c [Accessed May 14th 2020].

del Rio Riande, Gimena, Romina De Léon, and Nidia Hernández. *Annotate Texts in Recogito/Anotar Textos en Recogito*, 2019b. DOI: http://doi.org/10.5281/zenodo.3464 568 [Accessed May 14th 2020].

del Rio Riande, Gimena, Romina De León, and Nidia Hernández. *La Argentina Manuscrita.* 2019a. http://hdlab.space/La-Argentina-Manuscrita/ [Accessed May 14th 2020].

del Rio Riande, Gimena, Romina De Léon and Nidia Hernández. T*EI files in Recogito/Archivos TEI en Recogito*, 2019c. DOI: http://doi.org/10.5281/zenodo.3464576 [Accessed May 14th 2020].

Berman, Lex, Merrick, Ruth Mostern, and Humphrey Southall (eds). *Placing Names: Enriching and Integrating Gazetteers*. Bloomington: Indiana University Press, 2016. DOI: https://doi.org/10.2307/j.ctt2005zq7

Simon, Rainer, Elton Barker, Leif Isaksen, and Pau de Soto Cañamares. 'Linked Data Annotation without the Pointy Brackets: Introducing *Recogito 2*'. *Journal of Map & Geography Libraries*, vol. 13, no. 1, 2017, pp. 111–32. DOI: https://doi.org/10.1080/1 5420353.2017.1307303

Simon, Rainer, Valeria Vitale, Rebecca Kahn, Elton Barker, and Leif Isaksen. 'Revisiting Linking Early Geospatial Documents with *Recogito*'. *e-Perimetron*, vol. 14, no. 3, 2019, pp. 150–63. http://oro.open.ac.uk/68009/ [Accessed May 14th 2020].

Vitale, Valeria, and Rainer Simon. The Complete Recogito Tutorial. 2019. https://github.com/pelagios/pelagios.github.io/wiki [Accessed May 14th 2020].

# Typesetting queries

1. Figures 1, 2, 3, 5, 6, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20 and 21 are not cited within the text. Please advise where it should be cited.