GATE and xml-archive protocol
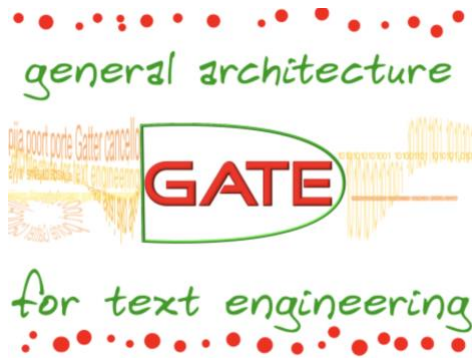
GATE download page: https://gate.ac.uk/download/
GATE Tutorials: YouTube Tutorials Point

NOTES
- Ctrl + Z to undo does not work…in fact, I haven't found a single feature that lets you undo. Even under Keyboard Shortcuts, there is no documentation to undo a mistake. This is a terrible design and a very unforgiving software.
- Name your documents reasonably. For example, "<year published> – <title>".

PROTOCOL
1. Open GATE by clicking on the application icon.
2. Right click on **Datastores** > **Create Datastore** > **SerialDataStore.** Then, select a folder to be your datastore. I kept mine simple and called a folder on my Desktop "datastore". A datastore is how you will save the progress of your annotated GATE files. It will also save your corpus or corpora.
3. If you already have a datastore, right click on **Datastores** > **Open Datastore** > **SerialDataStore** and then select your datastore folder. A datastore will then appear underneath datastores. For example,


4. Double click on the datastore icon. Now you can populate the datastore with GATE corpora or GATE documents. Start by right clicking on **Language Resource** > **New** > **GATE Document**. Click on orange folder on the center-right of the popup window. Select your document you want to annotate. Then, type in "utf-8" for the encoding. Click **OK**.
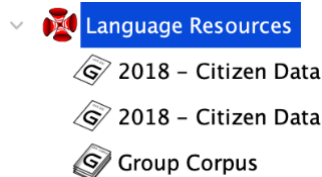
5. Now create a GATE Corpus. Right click on **Language Resource** > **New** > **GATE Corpus**. Click on the red pencil and paper icon on the right of the popup window and select your folder. Click **OK**.
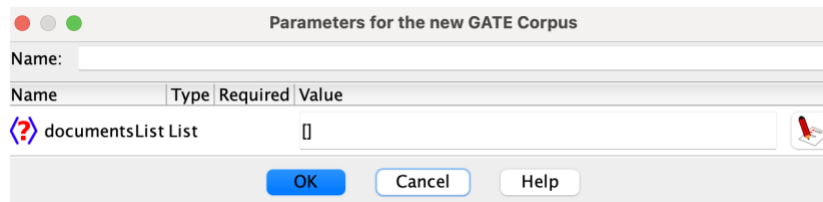

6. On the left-hand toolbar you'll see which ever GATE documents or corpora you uploaded.

7. Double click on the Corpus and click on this icon to add a file to the corpus. Click on this icon to remove a file. Click on this icon to open a file and start annotating it. Once you have a file, open the file.

8. **Click** on Text if it's not already selected and click on **Annotation Sets** > **Annotations List**. Select the dropdown arrow next to **Original markups** and select each tag. All the tags' annotations will show up in the box under the text. For example,
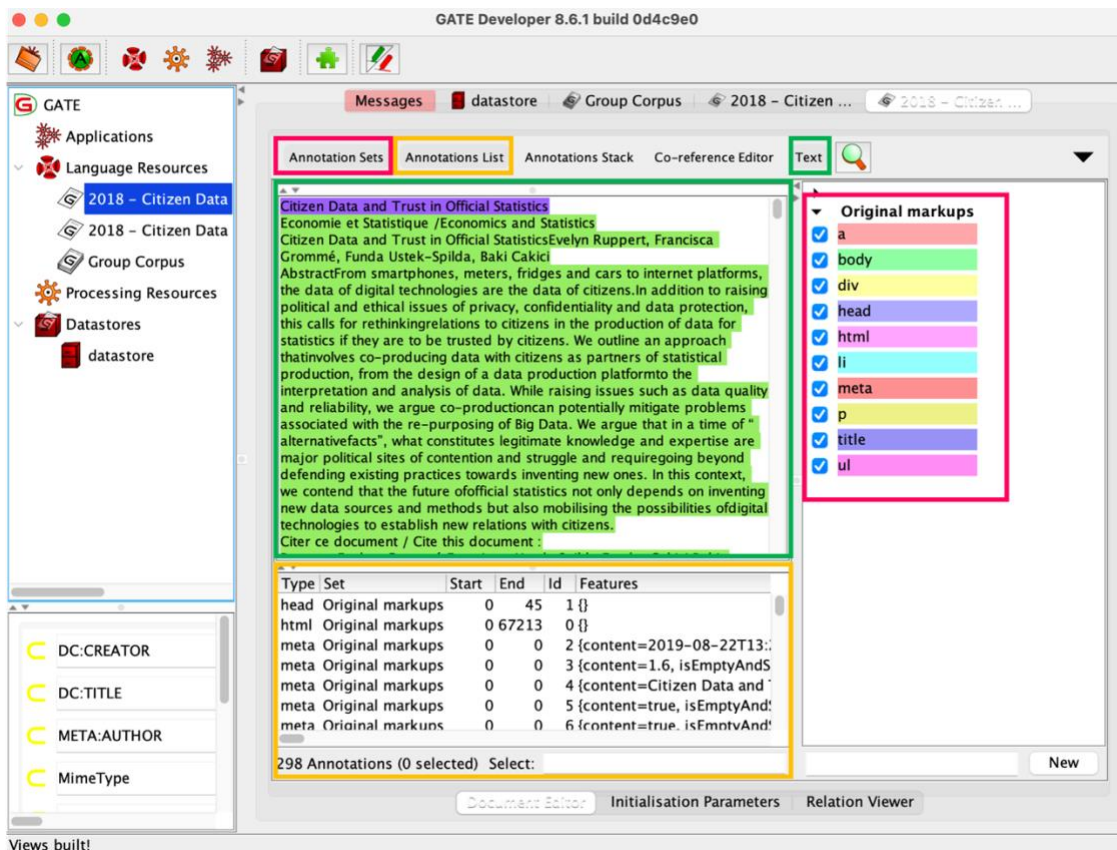
9. Delete all the original markup annotations by clicking on an annotation, typing Cmd + a, right clicking, and selecting **Delete Annotations**.
10. To create a new tag, highlight a sentence and then right click on the sentence. This box will appear:
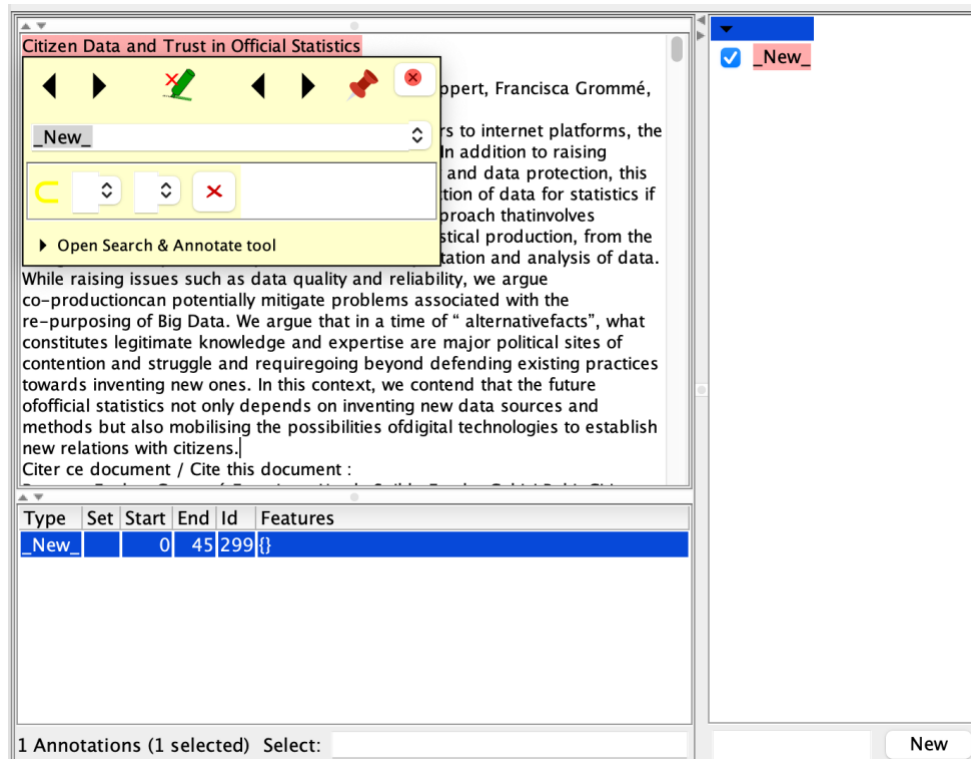
11. Type in the name of the tag and click the red x ⊗ in the corner. To change the color of the tag, right click on the tag in the right-hand box > **Change colour**.
12. Once you are done marking up your document, you need to do two things. If you are not finished marking up and you want to continue you later, you can right click on the **GATE document > Save to its Datastore**. Then, do the same to the corpus it came from: right click on the **GATE corpus > Save to its Datastore**. *This saves your progress*.
13. If you are fully done marking up your document. Follow step 12 and then finish by right clicking on the **GATE document > Save as… > GATE XML (.xml)** to create the .xml file.
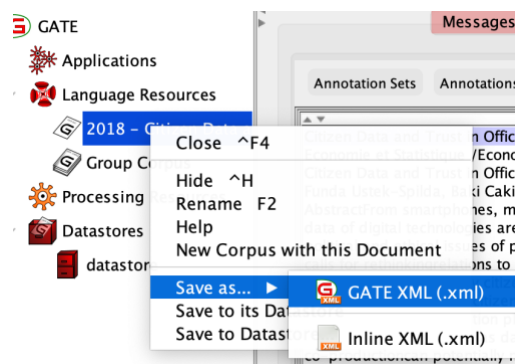
14. Finally, place the **xml file** into the xml-archive git repo. **Git add** it, **git commit** it, and then **git push** it for others to see/use.
15. If you want to run the python file on it, be sure to change the **filepath** and **summary sheet file name**.