WE1S "import" module
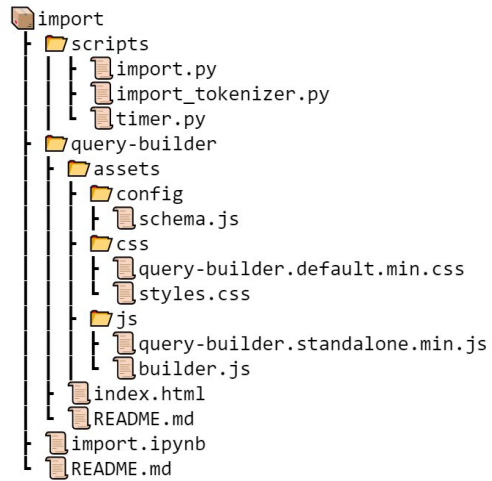
The WE1S `import` module is the main starting point for using the WE1S Workspace (see S-2). Use this module to import data into your project. Imported data is stored in the `project_data` folder. The import pipeline attempts to massage all data into a collection of JSON files, stored in the `project_data/json` folder. These JSON files contain both data and metadata as a series of key-value pairs that conform to the JSON file format. The names of the keys (also called "fields" or "properties") conform to the WE1S manifest schema *and* certain metadata required by the tools in the WE1S Workspace. After import, the text content of your data will be stored in the JSON files' `content` field. If your documents have not been tokenized previously, you can do so at this stage to extract lexical features to a `features` field and/or word counts to a `bag_of_words` field."

The import notebook requires either a zip archive of your data or a connection to a MongoDB database where the data is stored. Zip archives may take one of the following forms:
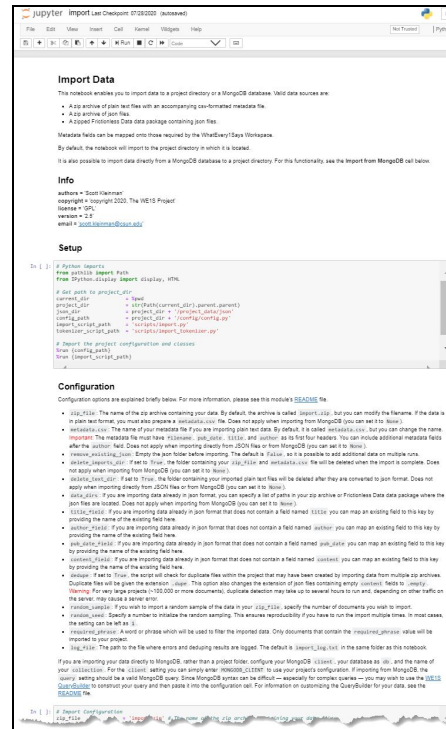
1. A zip archive containing plain text data *and* an accompanying CSV file with relevant metadata.

2. A zip archive containing data already in JSON format.

3. A zip archive of a Frictionless Data data package containing data already in JSON format.

If your metadata does not contain the field names required by the WE1S workspace, the import module allows you to map your metadata's field names onto the WE1S field names.

Module structure

Screenshot from `import.ipynb` (larger)

Further Information:

* M-15 (about Jupyter notebooks)

Main Jupyter notebooks in this module:

* `import.ipynb`

Code source: [TBD] (MIT License)