This is a documentation of a temporary REST service for the data curator.

To execute the REST service, through a temporary web interface, follow these steps:

1. Open the “data\_curator.py” script with a Python editor (Spyder is recommended) and change the IP in the last line of the source code to match your computer’s IP.
2. Make sure that the “pSS\_reference\_model.xml” is inside the same folder with the script.
3. Make sure that the folder “templates” is also inside the same folder with the script.
4. Make sure that the folder “uploaded” is also inside the same folder with the script.
5. Run the “data\_curator.py” file.
6. Make sure that the REST server is successfully created.
7. Open a browser and call the service as follows: 127.0.0.1:9000/upload
8. The temporary interface of the data curator’s REST service will then come up.
9. Select a dataset to upload (only .xls or .xlsx formats are accepted). The dataset will be stored on the folder “uploaded/” and after the execution of the service it will be automatically erased.**\***
10. Select a method for outlier-detection (mandatory).
11. Select a method for data imputation (optional).
12. Press the “Apply” button.
13. A folder “results/” will be automatically generated, including three documents: (i) the curated dataset, (ii) the data quality assessment report, and (iii) the data standardization report.
14. The files will be also automatically downloaded in a .zip folder.

**\*** The dataset needs to be in a tabular format (where the number of rows is equal to the number of patients and the number of columns is equal to the number of features, with the first row including the features labels). Since the data standardization process is exclusively dedicated to the primary Sjogren’s Syndrome (pSS) domain, any attempt to run it using an irrelevant dataset would be pointless. However, the produced data quality assessment report and the curated dataset will not be affected.

For data protection purposes, the anonymized datasets that were used in the manuscript which involves the source code of the data curator are not published.

Contact e-mail: [bpezoulas@gmail.com](mailto:bpezoulas@gmail.com)