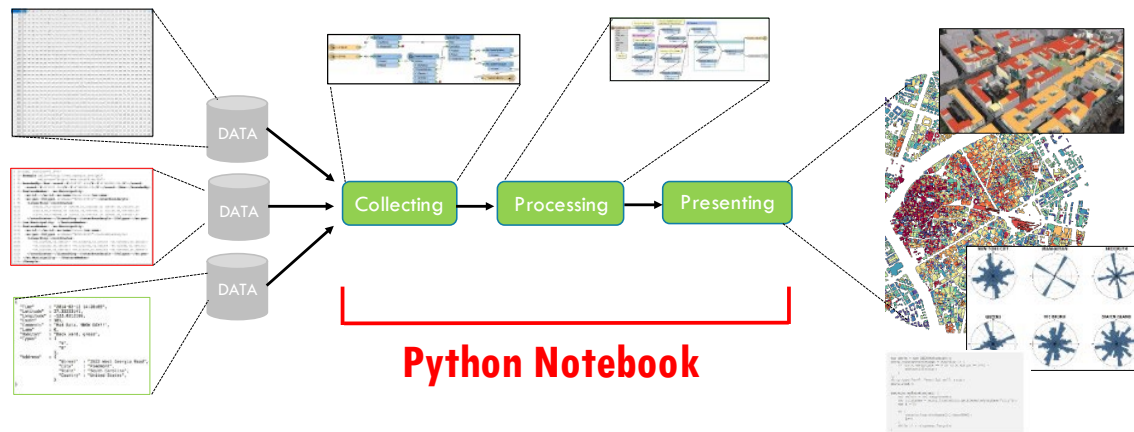


USP-E0363 - Digital Urban Exercise: My data factory

TASK: The course exercise is to create your own “data factory” using Python. Data factory is understood as a data pipeline from raw data to more advanced information content. The topic and maneuvers are defined by student according to one’s own interest and demand level. The simplified idea of this sequential data processing path is visualized in image below.



FORM: The exercise is enclosed in a single *commented* Python Notebook file (.ipynb format). If needed, additional data for the internal use of transformation process can be included as an external file. The only requirement for the pipeline is that at some point of the process geospatial data is used.

The data processing in pipeline can be divided in different phases like shown in above illustration. It is however important to remember that different phases of data flow have not clear-cut boundaries. Technically all components you recognize in your process can most likely further subdivided for alternate requirements and processing needs.

In the design phase of your work important parts are the interfaces between processes: what is the information that is passed for the next stage. Keep in mind that you don’t necessarily have to start from the most elementary pieces of data, but you can pre-process data in some explicitly defined way and only during the process check whether more rigorous definition or advanced processing is needed.

ASSESSMENT CRITERIA: Assessment criteria includes multiple dimensions depending on selected project. The shared criteria for all works are following:

For grade 3 student is expected to pass in a working pipeline code with multiple data sources and outputs.

For grade 4 student is assumed to show advancements in Python implementation, use of several external libraries in creative manner and produce outcome that is not obvious from the original data source or library code used.

For grade 5 student’s shows advanced understanding of coding and data management. Work is packaged in a form that it is able to adjust for multiple source data structures, uses advanced API interfaces and/or is able to process advanced output formats.

Requirements of lower grades must be fulfilled for higher grades.