**Project specifications**

The project aims to make an analysis of a large amount of data.

* In this project we work on Belgium open data,we have to select databases for:Population ,External immigration and Number of crimes

Those data must be open linked data , available in the open data cloud , in the open data of the belgium government or in the european portal.

* We want to understand the link between those databases in order to understand the causes behind the increase of the number of crimes in the regions of Belgium over the time.
* Visualization is an efficient method to do data analysis , so we have to invistigate the most efficient visualization techniques in order to display our data in graphs (Piechart,columnchart ,Areachart,Barchart,maps,....) then it will be easy to take the right decision about the main cause of the Crimes increase.
* Finally we have to build an authorized user interface that enables us to see our visualized data.

**Question 1 : Why do we use linked open data ?**

* More structered(triplet subject,predicate,object)
* Machine readable (RDF form)
* More useful through semantic queries

So we have to master semantic technologies in order to work with linked data :

* semantic web
* Web ontology language ,Description logic and RDF schema
* Resource Description framework
* SPARQL protocol and RDF query language
* Linked data

**Question 2** **:How can we convert raw data to linked data (RDF form) ?**

* First of all we have to define our vocabularies describing names of our classes and the links between objects. So we use the well known software of defining semantic ontologies which is "Protégé"
* After defining RDF vocabularies ,we convert the data to RDF ,we pick one of the RDF serialization forms RDF turtle for example.
* In order to convert our raw data (excel file) to RDF turtle we can use an efficient tool for cleaning data and converting it to RDF form which is LodRefine.We have to make our RDF graph by defining nodes and the links between them and that by using the rdf vocabularies that we have already defined with Protégé. Finally we can export the Data to RDF turtle.

**Question3 : how can we query RDF data ?**

* We query RDF data by using SPARQL which is the query language of RDF. So we have to master SPARQL queries in order to have the right wanted results from our RDF database.
* We have to stock our rdf data in an rdf database and then we write some sparql queries in order to get some results from our rdf database.
* we use Apache Jena Fuseki which contains an RDF database and enables us to have a sparql endpoint in our localhost. For example , we display :Population, External immigration and number of crimes of Brussels-Capital from 2000 to 2013.

**Question 4: How can we visualize the data from our SPARQL endpoint ?**

* There is an easy java script library that enables us to visualize the data from the sparql endpoint which is **sgvizler.**
* we have to put our **endpoint** which is in our localhost and to write **the query** for the result that we want to display and finally we pic the **type of the visualization** that we want to do (Piechart for example)

**Question 5 : Is there other methods of visualization and mapping ??**

**Question6: How can we develop the user interface or a dashboard ?**

**Question 7: Finally how can we write our decisions after the visualization ?**