Quality Assessment on Enterprise Linked Data

Authors: *Rauf Aghayev, Vugar Aghayev, Carlos Eduardo Montoya Sanchez*   **Supervisor**: *Jeremy Debattista*
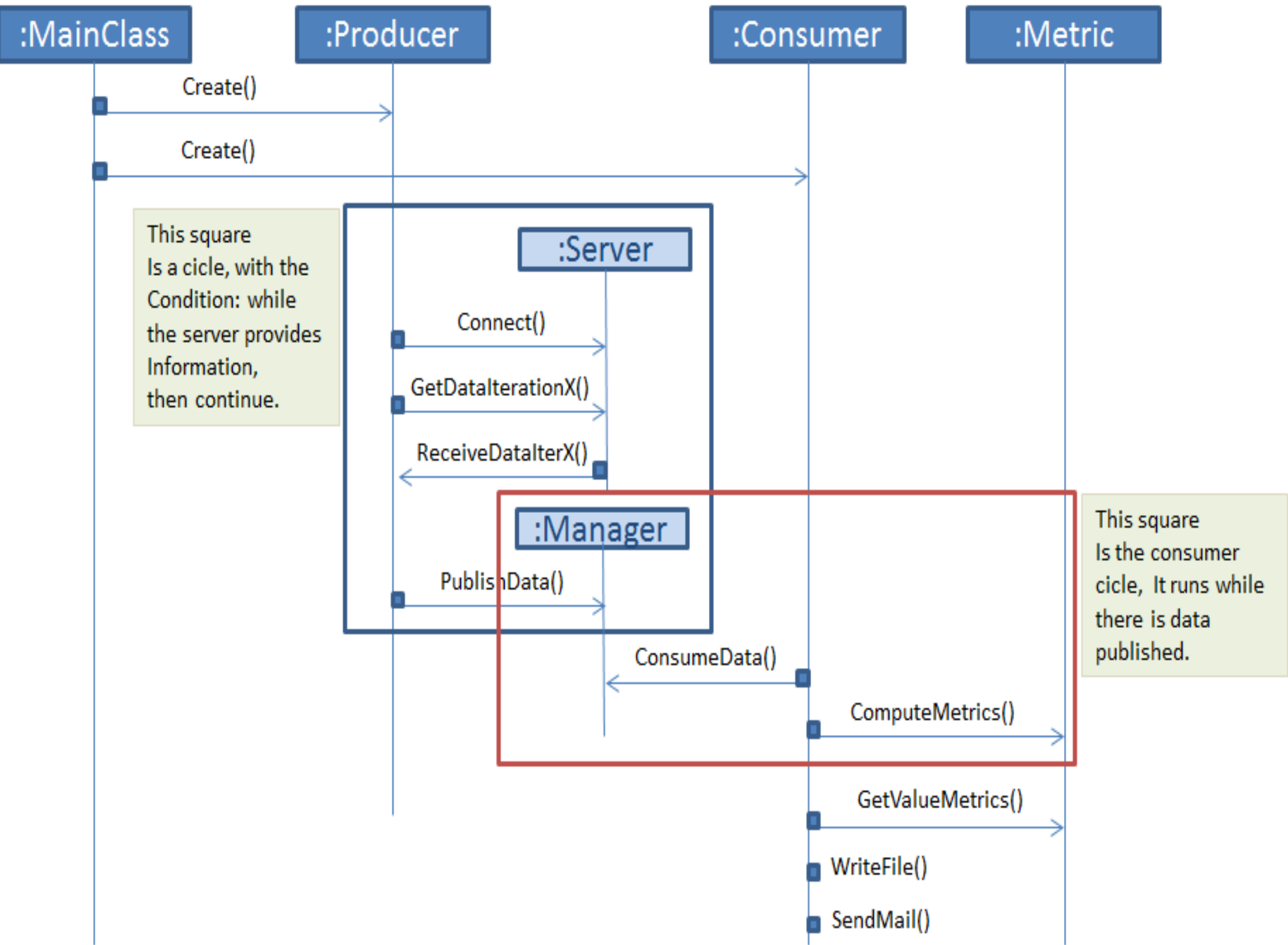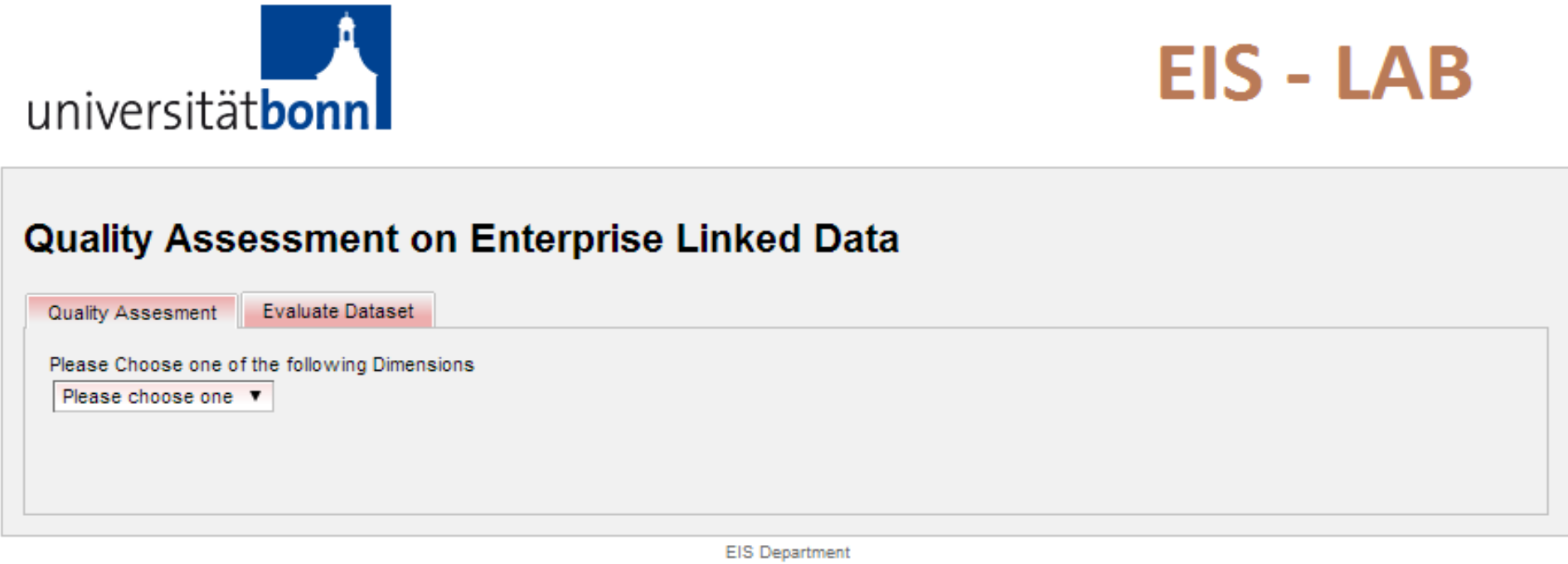
Flow of the program

To understand how it works the processing of the program you should understand the producer-consumer problem. In this lab, the problem is solved in the project eis-lab in the package de.unibonn.iai.eis.qaentlod.io.streamprocessor.

This is implemented to improve the times that takes to stream one SPARQL endpoint, because usually these endpoints contained millions of RDF points of information.

Then the producer in in charge to establish the connection to the SPARQL endpoint, then using SPARQL queries it retrieved all the information by batches of size 10000 triples per time.

While the producer is calling the information, the consumer is running and waiting for all the information that is published by the consumer.

In the consumer are located all instances of the Quality Metrics, so in every lecture that the program does it send to every metric to compute the value of the metrics.

To accomplish this the project use LIMIT and OFFSET.

This process continue while the producer is publishing information, when this one stops, then the consumer received the message that there is no more information pending to be publish. So it continues to end the program.

After ending the consumer, with all the values obtained by the metrics, it goes and save the information of every metric into the File that is design for this.

The final step is that the client sends a mail of confirmation when the process is done. Next is a sequence diagram to explain the process.
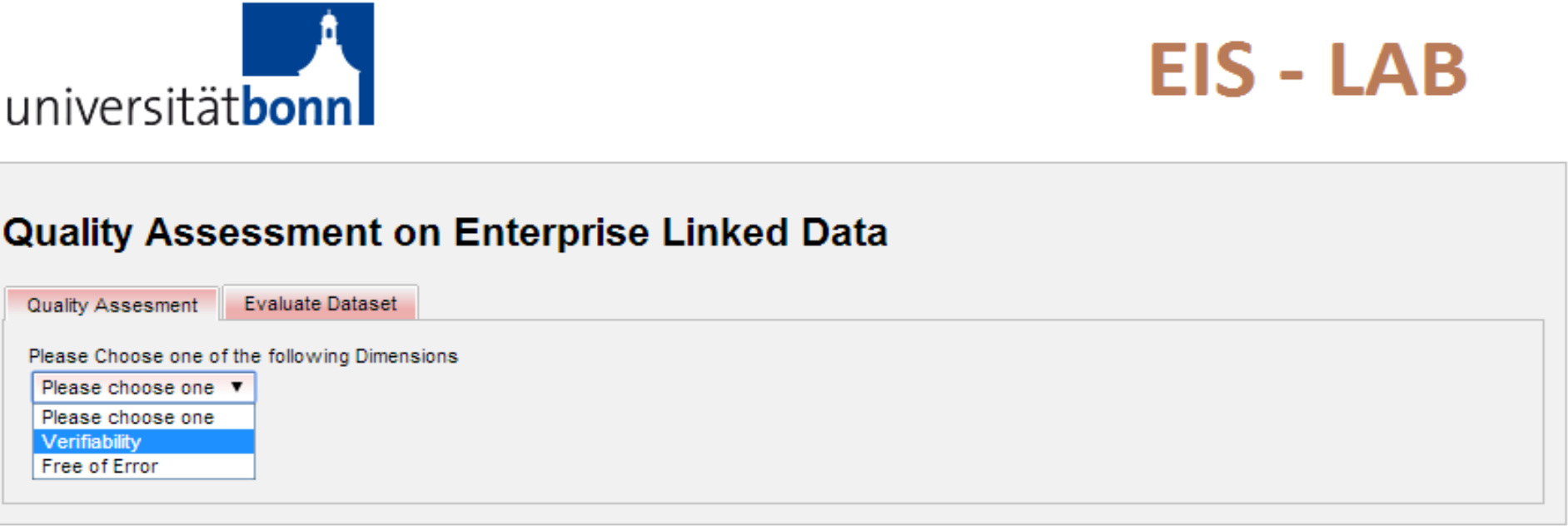

How to Access and use the program

To access to the web application the user need to access on the Web navigator to the link: http://localhost:8080/EIS_LAB_QAENTLOD/index.jsf.

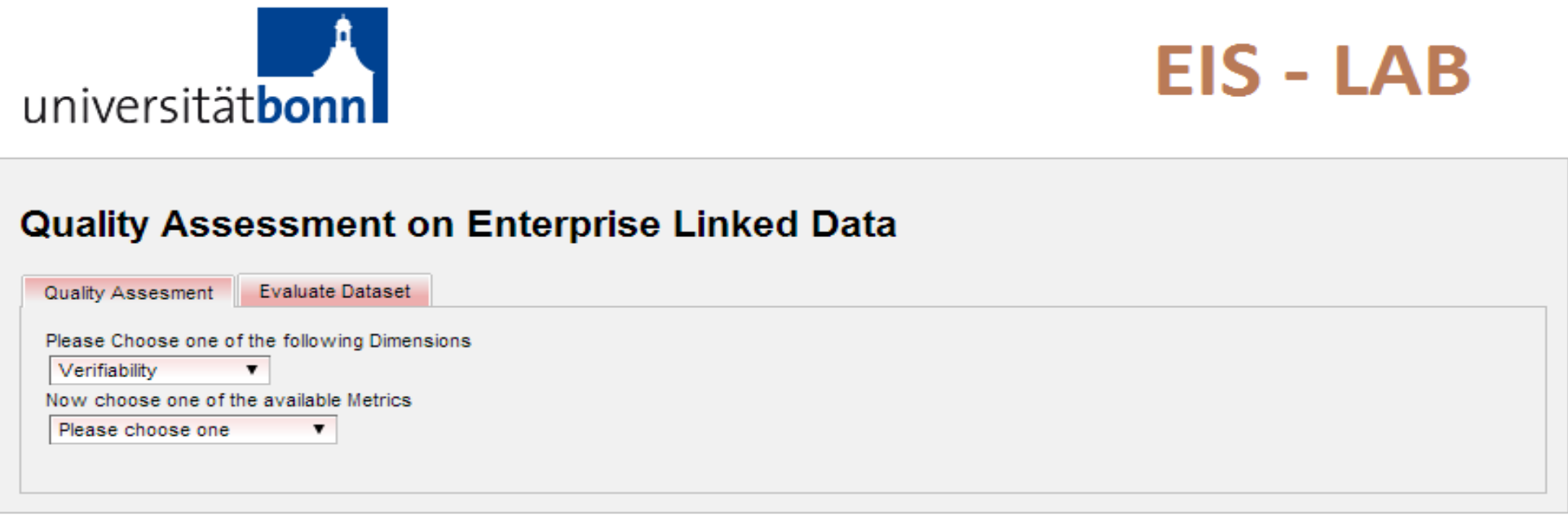Once it is running the user should be able to see the next things.

There are two different functionalities available in the system:
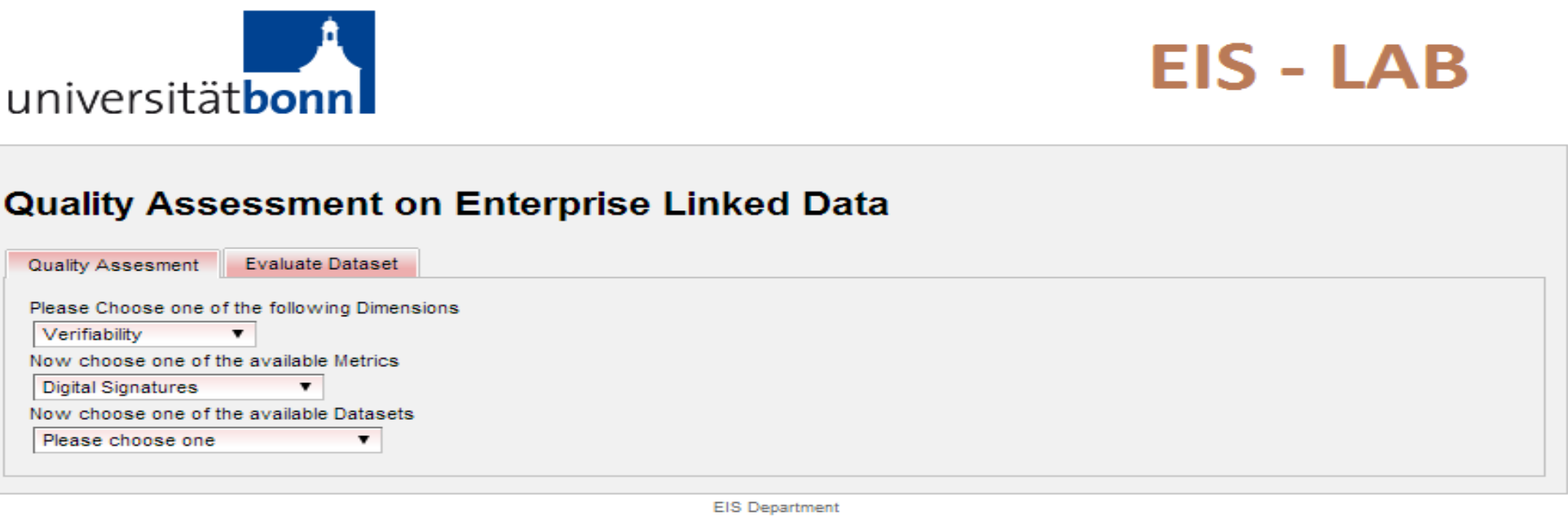
Visualize results

The first tab was created with the aim to visualize the results; to make it work the user should follow the next steps.

First the user should choose one of the dimension that are display on the select panel, the next image show how it works.

When the user has selected one of the dimensions then the system shows a new selectable combo box as shown in the next image.

The user should select one of the available dimensions, once this is done, then the system display a new combo box, to show the available Data Sets that already had been evaluated.
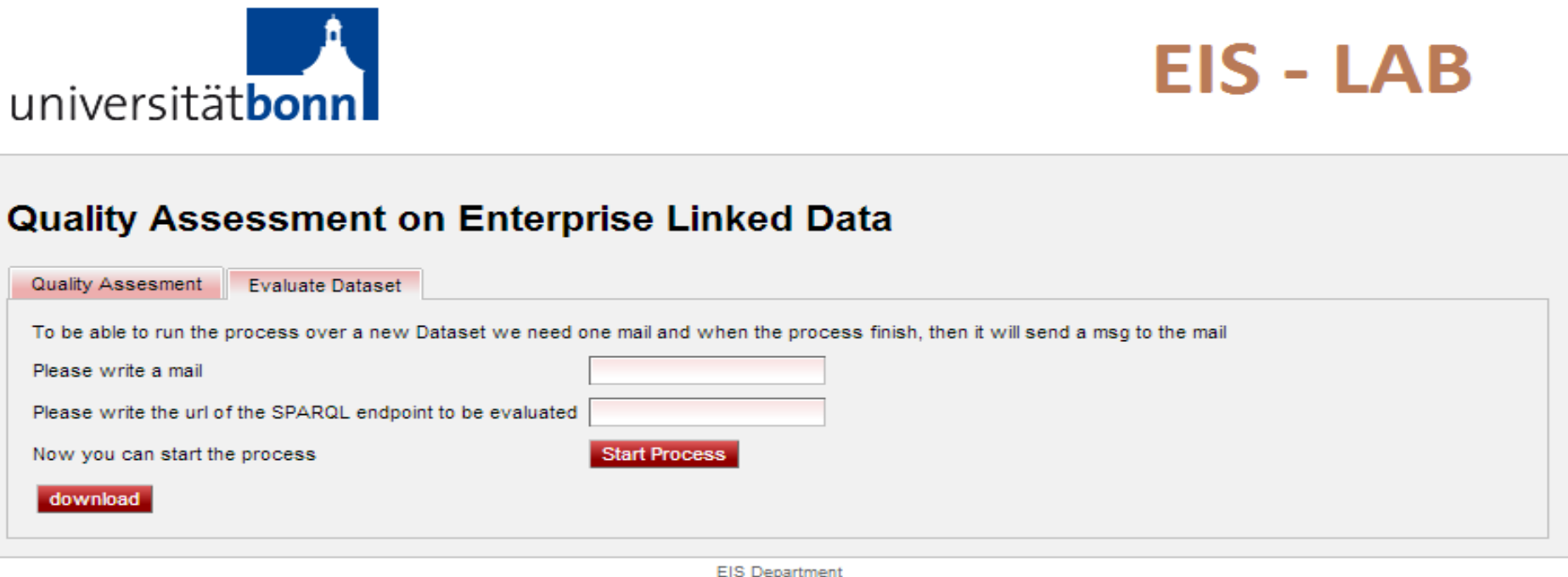

Then the user should select one of the datasets to see the results of this metric over the data set.

Evaluate new Dataset

The second tab provide by the system is in charge to evaluate a new data set or can be used to reevaluate one of the existing datasets.

When the user access to the second tab, the system shows the next screen:

Then the user should to fill the information of user mail and SPARQL endpoint and then activate the "Start Process" button, this is going to start the process of evaluate the data set, once the process is done then the program will send an automatic mail to the specified mail.

Also the user can download the eis-lab.jar to work with it if they want to do it. To accomplish this, the user only has to click on the bottom "download" automatically it will be save to the computer.