The article describes a process of designing software for the aggregation of data (macroeconomic and statistical indicators) from distributed heterogeneous sources and their analysis based on the previously developed ontology of innovation activity and economic potential. The software includes a data aggregation system (supporting the user’s markup process of PDF, HTML, and XLS documents and texts for further automated collection), an ontology automatic replenishment system and a system of semantic search for data subsets according to certain criteria[1].

The semantic web is built on ontologies. And the ontologies are usually used to form knowledge systems. Unlike the semantic search that uses machine learning, the ontologies use manually declared links and relations between terms. Such an approach allows engineers to extract hidden knowledge from the data set but it also requires a lot of input work (ontology formation) before the system can spot anything useful [1].

The user uploads a table of information in a file of one of the following formats: xlsx, html, xml, pdf. The system parses the file in search of a table. When a table is found it is displayed and the user is prompted to map all the rows and columns in the table to classes and their properties in the target ontology respectively. Once the mapping gets done, the system inserts the data into the target ontology. The system automates the information insert and thus simplifies the process of information preparation before the analysis. But the system still requires repetitive manual work (preparation of the file, mapping so on).

The topic of integration of heterogeneous data is very crucial in the realm of data analysis. Many works are written on the approaches applicable to the integration of different ontologies. The overall approach is to design a global ontology.