583SM - DATA MANAGEMENT FOR BIG DATA

Aims

The students must learn how to organize, manipulate, and analyze small and big data with a variety methods, techniques, and tools.

Knowledge and understanding: the students must acquire the necessary knowledge to model, import, tidy, transform, query, visualize, and analyze data as well as to communicate the results of the analysis. We take into consideration relational data as well as semistructured and unstructured data.

Applied knowledge and understanding: the students must learn languages and tools for the manipulation, analysis, and visualization of data, such as, for instance, PostgreSQL and BaseX for the management of relational and XML data, R and RStudio environment for data analysis and visualization, Processing for the visualization of data, and R Markdown language for the communication of the results of the analysis.

Making judgements: the students must be able to interpret the experimental results of the analysis and draw effective conclusions relevant to the domain of discourse.

Communication skills: the students must be able to communicate effectively the results of the analysis. This includes both analyst-to-analyst communication and analyst-to-decision-maker communication.

Learning skills: the students must demonstrate that they have learned the ability to choose a sufficiently rich row data set, to analyze the data to extract meaningful information, to draw and to communicate conclusions

Teaching Format

The course is organized in three parts and it is given by two different teachers; in addition, some advanced topics might be covered by invited experts in the field. Classes mainly consist in lectures given by the teacher. Students are also introduced to software resources to download, install, and run for the first time: the teacher will give a brief practical introduction to them.

Assessment

The exam consists of a written test and, possibly, an additional oral examination