DESCRIPTION FOR THE GENERAL PUBLIC (IN ENGLISH)

The project aims at developing methods for efficient processing and management of datasets obtained in scientific experiments and containing data of quantitative nature, i.e., observations of traits characterizing studied objects. The methods will be based on the so-called semantic layer of the data, so they will analyse not only the numerical values, but also the meaning of observations, both in the context of a specific research area, and the statistical meaning. The main target is to define of a semantic measure of similarity of scientific datasets that will serve as a basis for their comparisons and searching. The developed solutions will be used for construction of a prototype system for organization and processing of scientific quantitative data.

The problem of automatic processing of scientific data is important and timely because the number of publicly accessible data is rapidly increasing. Data publication is becoming a common element of research process due to requirements of funding agencies and journals, as well as because of numerous advantages for scientific discoveries, and the profits of data citation. Publicly available "open" data are useful only if they can be effectively found, interpreted and integrated with other data for subsequent analyses. Semantic representation allows to formalize and standardize data description, which provides possibilities of automatic processing.

The usefulness of project’s results will be practically demonstrated on exemplary datasets containing results of biological experiments from the area of plant phenotyping, that is, performed to study quantitative plant traits describing their environmental fitness and economical value.