Integrating and querying heterogeneous datasets with *AskOmics:* the ***DeepImpact*** case study.

The DEEP IMPACT project aims to combine ecology, biology, plant genetics, and mathematics to identify, characterize, and validate microbial communities, plant communities, and abiotic factors (agricultural practices) influencing *Brassica napus* and *Triticum aestivum* resistance to several bio-aggressors. Datasets from this project include surveys such as agricultural practices (i.e. crop rotation, plowing modes, chemical inputs), and crop yields, soil physicochemical properties, or inventory of parasites and weeds, established on several field plots. They also encompass abundance data of bacterial and fungal microbiota from different compartments of each cultivated plant. Other data, experimental or modeled, are currently being acquired, such as the metabolism of minimal microbial communities built based on field data.

AskOmics (*askomics.org*) is a tool based on semantic web technologies, offering a user-friendly interface for the integration and querying of heterogeneous data. Through this interface, users can query integrated data by creating complex queries iteratively. Using this tool, we have partially integrated the data from the DeepImpact project to provide partners with a simple solution to explore and extract relevant scientific information. More specifically, a data model has been created to integrate field data, climatic data, and soil microbiota data. Through this model, project members can easily construct and share queries enabling the selection and linking of various studied variables from heterogeneous data formats, such as characterizing the microbiota metabolism of each plant based on different abiotic variables. The additional integration of a relevant subset of the NCBI taxonomy's ontology allows users to make hierarchical queries to further improve data exploration. All generated queries can be shared and improved upon between project members to improve reproducibility.