



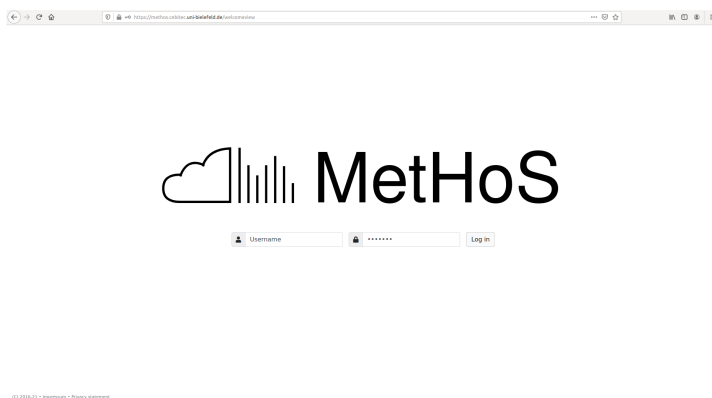
# MetHoS

## A walk through MetHoS

---

A guide through process and analysis of  
mass spectrometry-based metabolomics  
data.

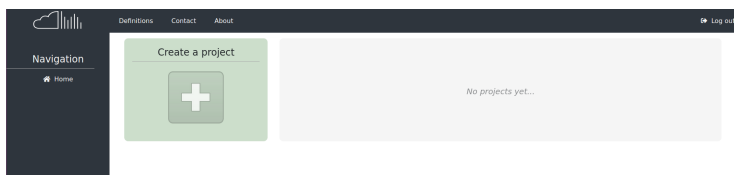
In the following we would like to introduce you some of the functionality of MetHoS - a web-based platform for large-scale storage, processing and analysis of metabolomics data. This walkthrough aims to guide through the system and demonstrate how to reach analysis results from raw data. MetHoS is best displayed using Mozilla Firefox, but other browsers such as Google Chrome can also be used.



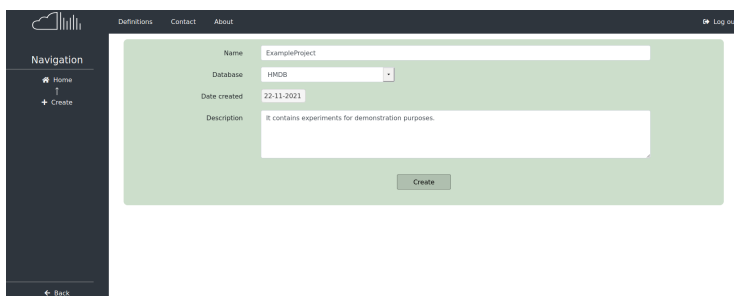
To log into the system, please first enter your username and password and click on the "Log in" button under the logo.



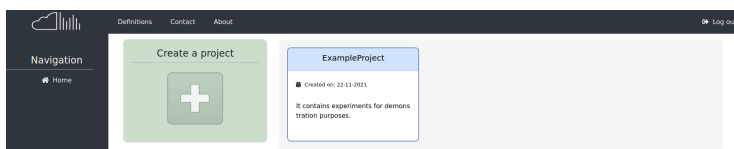
Welcome to MetHoS! On the "Home" page you will notice that you can create a project or navigate to the tabs "Definitions", "Contact" and "About" on the top left corner for more information.



When you create a project you have to give a name, select one of the available databases and write an appropriate description. The database will be used later during the processing in the identification step.



On the main "Home" page you can see all the projects you have created or have access to.



The "Definitions" and "About" tab will help you understand the concepts adopted by the tool and how it operates. There is a brief description on the functionality of MetHoS and which software is integrated in order to achieve parallel processing, distributed storage and distributed analysis.

**What is a project?**  
A project is a collection of tasks the user has to undertake in order to acquire results. It contains many biological experiments which must be processed firstly in order to be analyzed. When creating a project, the user can select among three available databases: HMDB, MassBank and MoNA. Fein. The database will be used later for the identification of metabolite features.

**What is an experiment?**  
An experiment in MetHoS refers to a biological experiment that consists of many biological replicates (.mzML, mzData and .mzXML files).

**What is View?**  
In this section the user has the ability to view the raw results which have been stored into Cassandra database by the XNIME workflows. He is able to view all the information of the identified and unidentified metabolites of an experiment as well as the Biopics of selected experiments and metabolites.

**What is Edit?**  
In the edit section, the owner of the project can edit the project information, specifically the name and the descriptor.

**What is Upload?**  
The upload section is where the user has to upload his experiments by providing a name for his experiment and the replicates (.mzML, mzData and .mzXML files). The replicates need to be uploaded together under the same name of the experiment.

**What is Process?**  
In the process section the user is able to select from his project repository which experiments he wants to process with the integrated XNIME workflows in order to quantify and identify metabolite features. MetHoS currently provides three different workflows for MS1 spectra linked with all three available database and one for MS2 spectra linked with MassBank database.

**What is Analyze?**  
Analyze section provides the statistical tests to analyze the results from the different Xnime workflows. Currently implemented tests are: Principal Component Analysis, two clustering methods, Pearson and Spearman tests as well as some basic statistics and metabolite filtering based on mean and standard deviation values.

**What is Share?**  
The share section gives the opportunity to the owner of the project to allow access to any other desirable user.

**MethoS**  
MethoS is a web-based platform for large-scale metabolomics. It is written in java that utilizes a set of software tools in order to achieve parallel processing, distributed storage and distributed statistical analysis of metabolomics data. It is built with Spring framework and utilizes the D3 javascript library for visualizations.

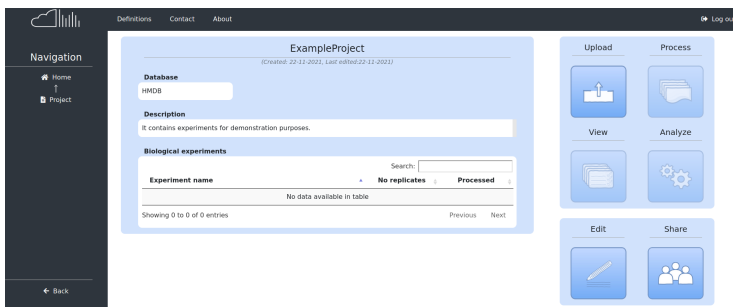
**Architecture Diagram:**  
The diagram shows a central box containing three main components: a yellow circle labeled 'KNIME', a blue circle labeled 'Apache Cassandra', and an orange circle labeled 'Apache Spark'. Arrows indicate dependencies: 'Spring' (green box) points to 'KNIME'; 'KNIME' points to 'Apache Cassandra'; 'Apache Spark' points to 'Apache Cassandra'. To the right of the central box, there are two red boxes: 'Openstack Object Storage' and 'Openstack', with arrows pointing from the central box to them.

**Parallel processing**  
Using pre-optimized XNIME workflows, MethoS is able to quantify and identify metabolite features. Apache Spark is responsible for the distribution of these processing jobs to the nodes of the computer cluster where they are parallelized. Every node in the cluster downloads an experiment from the Object storage, where it is previously uploaded, and processes it with the selected XNIME workflow. When the processing of an experiment is finished, KNIME automatically sends the results to Apache Cassandra and downloads the next one in line.

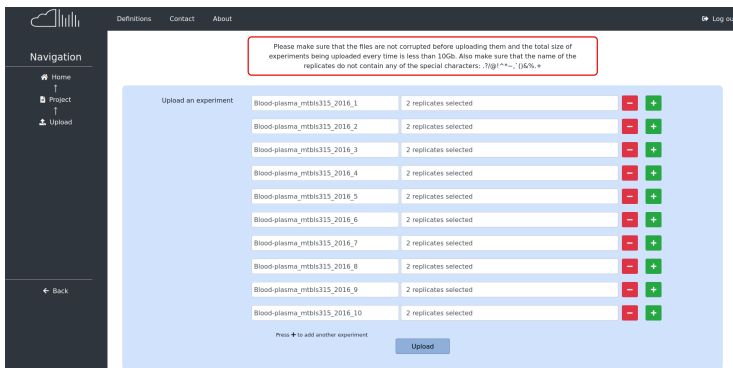
**Distributed storage**  
Apache Cassandra is responsible for the distributed storage of the results of a XNIME workflow across the cluster. The results originated from an experiment are saved in one Cassandra node and copied to two more to ensure reliability and fault tolerance.

**Distributed statistics**  
Apache Spark Machine Learning Library is responsible for the implementations of several statistical tests in a distributed manner.

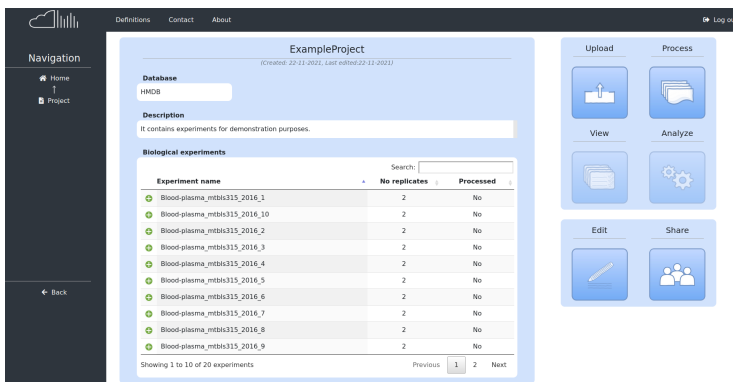
On the main project page you can see the information of the project and some available options. The creator of the project is the owner of the project and the only user who can give or remove access rights of the project with the "Share" button. Also, the owner is the only user who has the right to edit the name and description of the project with the "Edit" button.



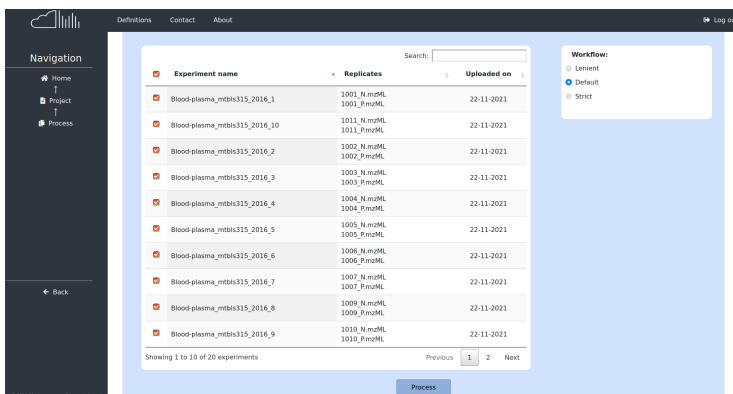
With the "Upload" button you can upload your experiments. Each experiment refers to a biological experiment that may consist of many biological replicates (files). The replicates must be grouped and uploaded together for every experiment. For our example, we have randomly selected 10 blood plasma experiments of the mtbls315 study and 10 urine experiments of the mtbls28 study of the MetaboLights database.



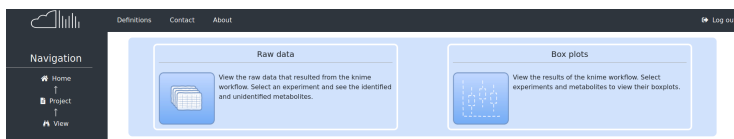
After upload finishes, you will be able to see the list of experiments you uploaded, as well as their replicates, on the main panel of the project. You can also see that as soon as you have unprocessed experiments in the project, the "Process" button becomes available.



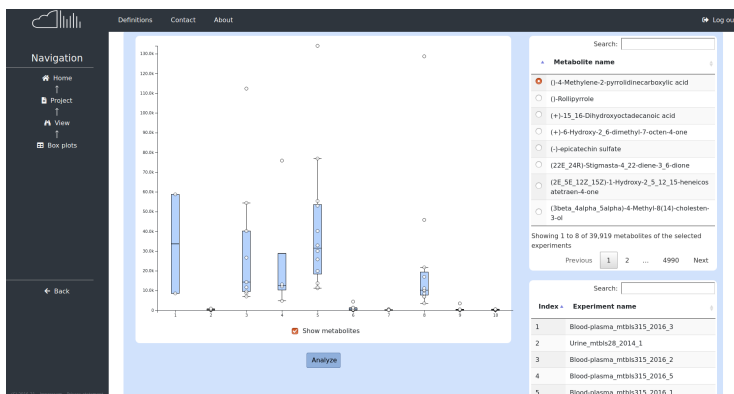
With the "Process" button you can select the experiments you want to process (quantification and identification) and choose one of the currently available workflows.



As soon as the processing is finished, the buttons "View" and "Analyze" become available. With the "View" button you are able to observe the raw data in a form of table or represent them in box plots.



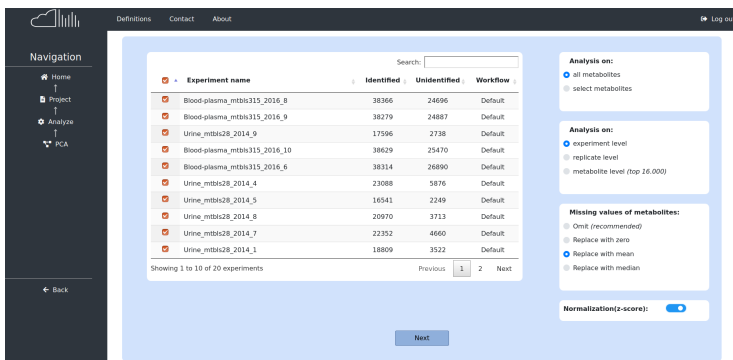
In the Box plots section you can see the distribution of metabolites in the selected experiments. A list of all the metabolites that are present at least once in at least one experiment is presented on the right. You can select any metabolite of this list and observe the boxplots with the outliers. Selecting the checkbox "Show metabolites" will present all the metabolites that are within the quartiles. In this specific example, you can see that the metabolite (-)-4-Methylene-2-pyrrolidinecarboxylic acid has a bigger distribution in the blood plasma experiments than the urine experiments.



With the "Analyze" button you can perform several statistical tests such as PCA, clustering and more.

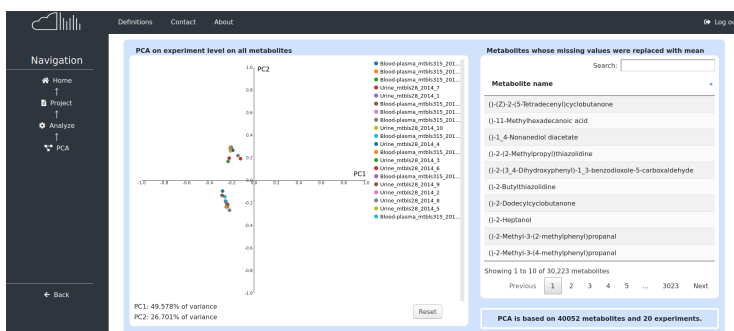


For most of the analysis tests you have the choice to select all or specific metabolites and handle missing values imputation by omitting or replacing them with the zero, mean or median values. Furthermore, you are able to define the depth of your analysis by choosing the level depending on what you want to focus on, e.g. experiments, replicates or metabolites.





Following, you can see the results of a PCA of all 20 experiments on all metabolites in experiment level. The missing values are replaced with the mean and normalized. The metabolites that were replaced by the mean are presented as a list on the right side as well as the total number of metabolites and experiments that participated in the analysis. If you hover over the plot you will see the names of the experiments while also you can zoom in the plot or select any experiment of the list on the right side to highlight it. In this example, we can see that there are two distinct groups, one with all the blood plasma experiments and one with all the urine experiments.



**Thank you very much for your interest  
in MetHoS.**

---

If you have further questions please do not hesitate to contact us.