INRAE

DeepOmics user guide
Main concepts

Digital Environmental Engineering Platform for Omics data

Bize Ariane
08/12/2023

v 2023-12-08

Table of contents

1. Introduction: what is DeepOmics?

DeepOmics is an information system (IS) dedicated to meta-omics data from environmental biotechnology processes, such as wastewater treatment or anaerobic digestion. It enables the management of data from samples that originate either from full-scale processes, or from laboratory or pilot reactors.

It intends to support the production of FAIR data, thereby promoting data valorization, exchange and reuse. Through its wide use, it will enable data mining and facilitate biostatistical meta-analysis. It could foster innovation by accelerating the development of a microbial management of environmental processes.

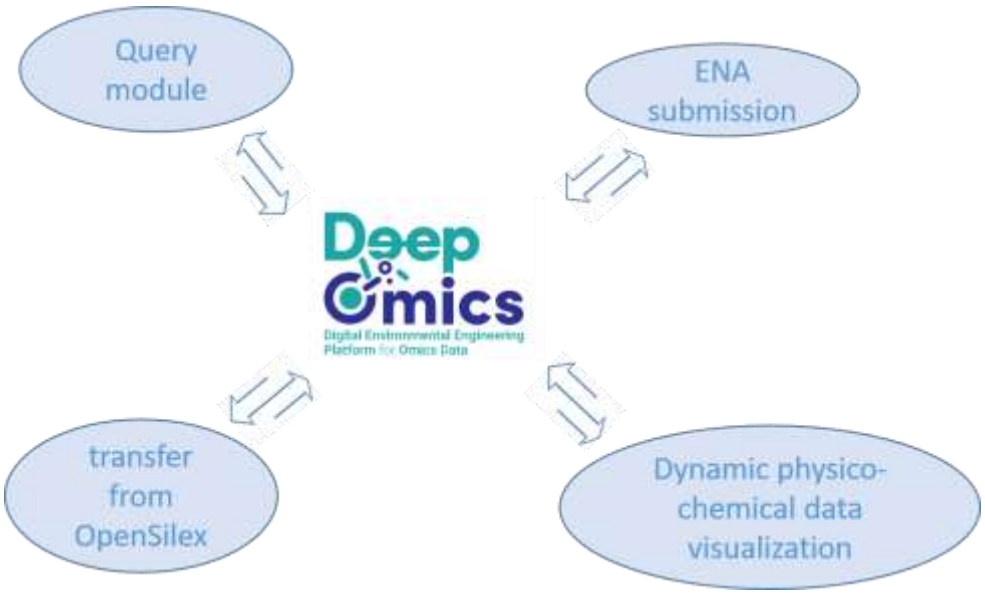
In the present version, DeepOmics enables the storage of **amplicon sequencing data** (typically **16S rDNA metabarcoding** data but not limited to them) as well as very rich data describing process design, operating parameters and physico-chemical monitoring measurements. The data stored in DeepOmics can be exported in standard formats (csv, biom, fastq, …). It covers both **single-end and paired-end data**.

For lab-scale and pilot processes, DeepOmics presently covers reactors with up to 3 compartments. Batch processes are more easily described in DeepOmics, but semi-continuous and continuous processes can also be entered with some limitations.

For full-scale processes, the current version of DeepOmics mainly covers wet and dry digestion, as well as activated sludge. The other types of processes can still be entered with a more limited and standard description.

A documentation website is available: https://deepomics-info.hub.inrae.fr/.

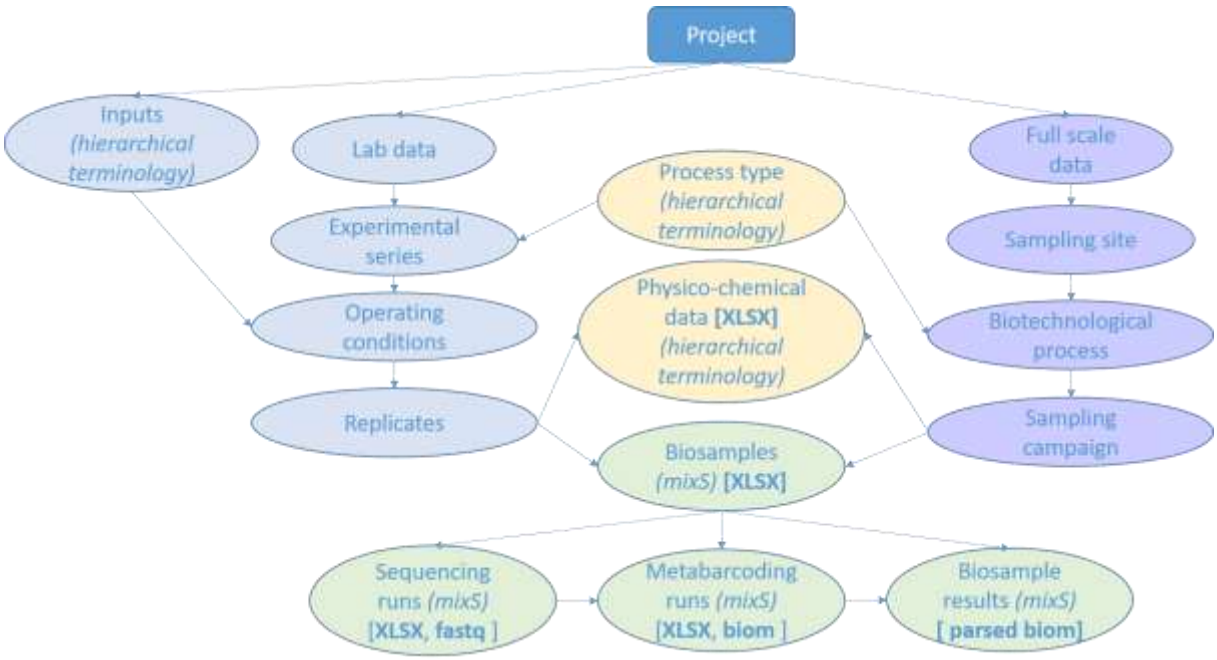
In the future, DeepOmics will be connected to various functional modules.



DeepOmics functional modularity in the near future

2. License

3. DeepOmics key concepts



Simplified model scheme of DeepOmics concepts

3.1 Lab-scale process data

Project's input management

The inputs are the compounds and elements which will be used in the reactors (feeding, gas, inocula, buffers, matrix, pure microbial strain, …). They are described according to a controlled vocabulary and they are defined at the scale of the project.

Experimental series

They represent a consistent batch of experiments led into reactors or pilots, with a project. They are structured into operating conditions and replicates.

Operating conditions

Each reactor can be composed of 1 to 10 distinct compartments.
A given operating condition can be defined at the level of the reactor (if it is identical for all compartments) or at the level of each compartment.
In each operating condition, you will be able to define replicates.

Replicates

Reactors which were subjected to the exact same treatment; they are grouped by operating conditions.

3.2 Industrial and field process data

Sampling campaigns, Sampling site and Biotechnological Process

You need to start by creating a **'Sampling site'**, which corresponds to the industrial site from which the samples originate (e.g: a wastewater treatment plant).
Secondly, you need to create a **'Biotechnological process'**, with the precise description of the process and reactor from which the samples originate. Indeed, a single industrial site can gather several processes (e.g. activated sludge, anaerobic digester), hence the relevant ones.
Finally, you need to create a **'Sampling campaign'** and you will then be able to enter your data.

3.3 Biosamples

Biosamples represent physical samples for which you plan to perform meta-omics analysis. It is advised to create them in DeepOmics before the acquisition of the corresponding meta-omics data.
You can add multiple biosamples at a time by clicking on **'+ Import biosamples'** (batch mode, through the filling and upload of a template).
Alternatively, it is possible to create biosamples manually, one at a time, by clicking on **'+ New biosample'** (interactive mode).

3.4 Meta-omics analysis

Sequencing run

It gathers a set of biosamples that were sequenced during a same sequencing run, associated to relevant metadata and raw fastq files.

Metarcoding run

It gathers a set of sequenced biosamples that were processed together for the bioinformatics analysis, associated to relevant metadata and biom file.

Biosample results

It shows all the metabarcoding run results, parsed by biosamples, for the whole project.

Biosample results can be selected to create a new biom file.

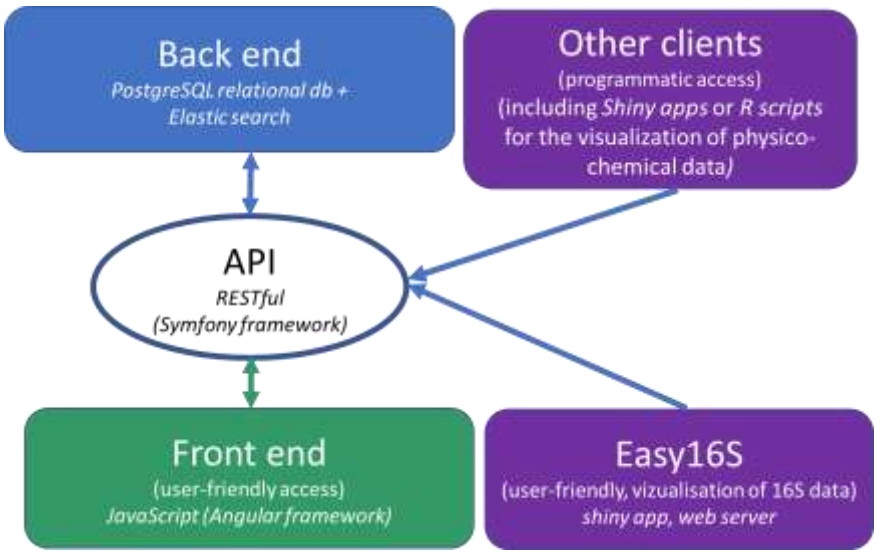
4. Query in DeepOmics

DeepOmics data can be accessed through an API (Application Programming Interface).

https://deepomics-api-test.solapp.inrae.fr/

No user-friendly search interface is available at the moment, but it is planned to develop one.

5. Informatic structure



Schematic view of DeepOmics informatics structure

DeepOmics Information System is an n-tier web application: the user interface is a single page application built with the Angular framework. It accesses the data using a RESTful API. Data are stored in a PostgreSQL relational database as well as in an indexation and search engine (Elasticsearch, open-source version). Easy16S is an interactive R shiny interface based on two main R packages, shinydashboard and phyloseq. Easy16S is currently deployed on the INRAE-MIGALE server (https://migale.inrae.fr/).

6. Funding and acknowledgements

DeepOmics was originally developed by the Information Systems Division of INRAE, under the coordination of INRAE-PROSE unit (https://www6.jouy.inrae.fr/prose_eng/), in collaboration with INRAE-LBE (https://www6.montpellier.inrae.fr/narbonne_eng/Laboratory-of-Environmental-Biotechnology/Welcome), INRAE-OPAALE (https://www6.rennes.inrae.fr/opaale_eng/) and INRAE-MaIAGE, MIGALE platform (https://migale.inrae.fr/).

DeepOmics recently benefited from the financial support of the division **Microbiology and the food chain of INRAE** (2020-2021), and of the **3BCAR network** (2022-2025).

We are grateful to Prof. Jo De Vrieze (CREAS, KU Leuven, Belgium) and Dr Claudia Etchebehere (Microbial ecology laboratory, Clemente Estable Biological Research Institute, Montevideo, Uruguay) for helpful discussions.

List of illustrations