INRAE

# DeepOmics user guide Quick start

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, …). Currently, it is limited to **single-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 webpage is available: https://deepomics-info.hub.inrae.fr/.

In the near future, we intend to enrich DeepOmics by developing new features. New types of meta-omics data (e.g. shotgun metagenomics, metatranscriptomics) and process types (bioelectrochemical systems) should be covered. Moreover, additional invaluable functionalities should be included such as a userfriendly search interface and the integration with complementary tools (easier sequence submission in the European Nucleotide Archive (ENA), coupling to Easy16S, a userfriendly tool for the interactive statistical analysis of count data from microbial communities, https://shiny.migale.inrae.fr/app/easy16S, doi attribution, etc).

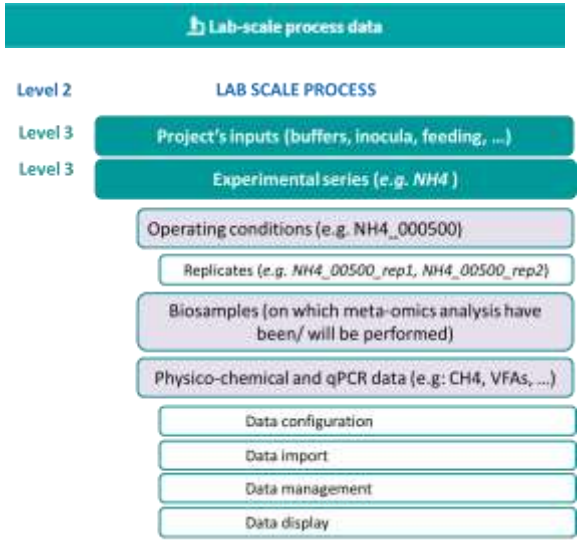
# 2. License

# 3. Memo on DeepOmics functional structure

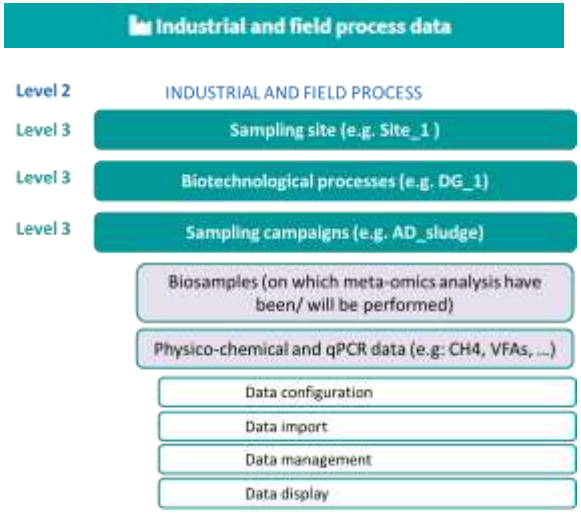
## A recap on DeepOmics structure

**Level 1** — **Projects (e.g. 00_demo_AD_inhib_Poirier)**

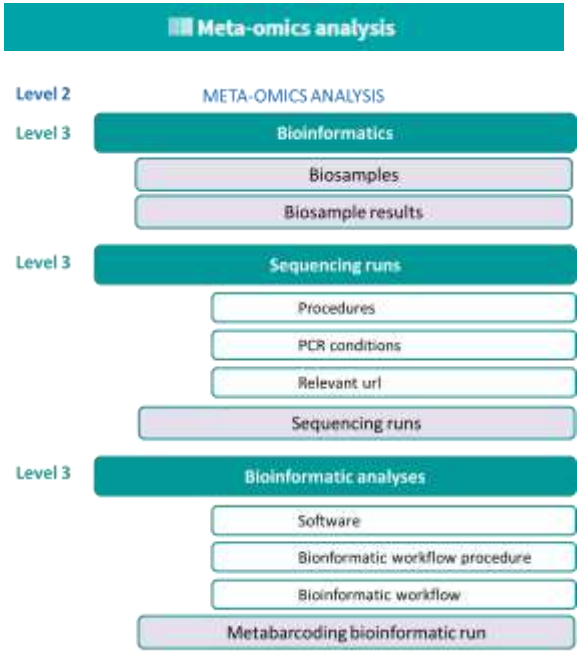
### Left tab: labscale




### Middle tab: industrial and field



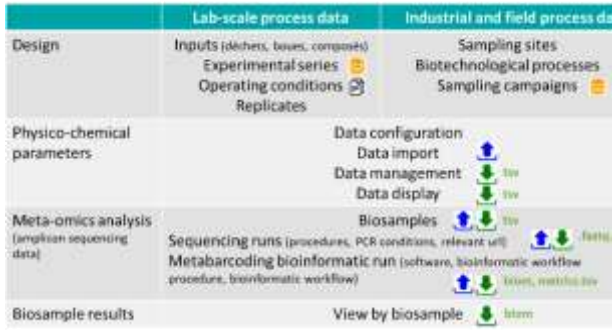

### Right tab: meta-omics



# 4. Quick start / Memo on how to enter a new dataset into DeepOmics

## Main steps of the procedure

### Summary table




### Project, reactors and processes

**CREATE A PROJECT**

1. Make sure you are on the project page (click on Home or Projects in the uppest menu bar)
2. Create a new Project (private or public) (*e.g. 00_demo_AD_inhib_Poirier* or *00_demo_AD_plants*)
3. Click on the created project to reach the project page

**CASE 1, LAB WORK, DESCRIBE THE REACTORS**

1. Example in *00_demo_AD_inhib_Poirier*
2. Make sure you are in the up-left panel: Lab-scale process data
2. Create the Inputs used in the project (buffers, substrates, sludge, chemical compounds, etc)
3. Create a new Experimental series (*e.g. NH4*)
4. Create Operating conditions (*e.g NH4_000000, ...* ) and describe in the form the precise conditions (*e.g. number of compartments, input amounts, temperature, volume, etc*)
5. Create Replicates for each Operating condition (*e.g. NH4_00000_rep1, NH4_00000_rep2, ...*)

**CASE 2, INDUSTRIAL PLANTS, DESCRIBE THE CAMPAIGNS**

1. Example in *00_demo_AD_plants*
2. Make sure you are in the up-middle panel: Industrial and field process data
2. Create a new Sampling sites (*e.g. Site_1*, left menu bar)
3. Create a new Biotechnological process (*e.g. DG_1*, left menu bar)
4. Create Sampling campaigns (*e.g. AD_sludge*, left menu bar)

### Physico-chemical data

**DESCRIBE THE ANALYTICAL PARAMETERS AND UPLOAD DATA**

1. Make sure you are either in an Experimental series (lab) or in a Sampling campaigns (industiral) (you can navigate by clicking on the up and left menu bars)
2. Click on Data configuration (left menu bar)
3. Select on the left the parameters corresponding to your experiment, set their unit, and validate)
4. Click on Data import (left menu bar)
5. Select on the left the desired parameters, in the middle the desired template type and on the right the desired reactors/processes
6. Generate the xlsx template
7. Fill in the xlsx template, for instance in Excel application (external to DeepOmics)
8. Import the data by selecting the filled xlsx file, at the bottom of the same DeepOmics page

**CREATE GRAPH COLLECTIONS TO VISUALIZE THE DATA**

1. Clic on Data display (left menu bar)
2. Clic on New graph collection
3. Chose a name and create New graphs

Amplicon sequencing data (e.g. 16S rRNA gene metabarcoding)

**CREATE BIOSAMPLES**

Biosamples correspond to samples for which you plan to acquire / have acquired amplicon sequencing data

**WARNING: chose identical names for the biosamples, fastq files and sample names in the biomfile**

1. Make sure you are either in an experimental series (lab) or in a sampling campaign (industrial) (you can navigate thanks to the up and left menu bars)
2. Click on Samples in the left menu bar
3. Click on Import samples (top right) and subsequently on Download template (xlsx)
4. Fill in the template with approriate information, e.g. in Excel application (external to DeepOmics). You can find some help on the DeepOmics page, by clicking on ?Show help to fill in template, or directly in the template, on the 2d sheet
5. Import the Biosamples by selecting the filled-in xlsx file, at the bottom of the same DeepOmics page

NB: For experimental series (lab), a biosample can also be created for Inputs. In this purpose, make sure you are in the interface Lab-scale process data (left pannel). Click on Project's input management towards the top of the left menu bar. Then, click on the desired Input (e.g. a sludge). Finally, click on +Add, on the top right of the form, next to "Biosample".

**CREATE A SEQUENCING RUN**

1. Make sure you are in the right panel (Meta-omics analysis)
2. Start by creating the description of your Procedures (extraction, amplification, library, sequencing, etc). It can be a document, an url or a doi. Also describe the PCR conditions (left menu bar). This step is not mandatory but it is advised. These functions are available in the left menu bar. The described procedures will be available for selection (through their name) in the template created afterwards.
3. Click on Sequencing runs (left menu bar) and add a New sequencing (top right). Fill in the form and Save.

**CREATE SEQUENCING SAMPLES AND IMPORT SEQUENCING RESULTS**

1. Make sure you are in a sequencing run (if required, click on Sequencing runs in the left panel and click on the desired sequencing run).
2. You should see: Manage sequencing samples. Download the xlsx template, fill it in (e.g. in Excel application, external to DeepOmics) and import the filled-in template on the same DeepOmics page.
3. Towards the bottom of the left menu bar, click on Sequencing sample metadata
4. Export the xlsx template, fill it in (e.g. in Excel application, external to DeepOmics), and import the filled-in template on the same DeepOmics page.
5. Click on Import fastq (towards the bottom of the left menu bar)
6. Export, fill-in and import the xlsx template, similar to above.
7. On the same page, add the fastq files (raw sequencing data)

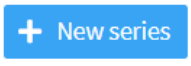
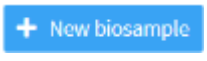
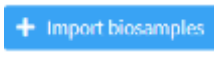
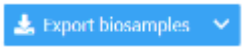
NB1: the Biosamples of a Sequencing run can be selected in the whole Project (all Experimental series and all Sampling campaigns)

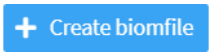
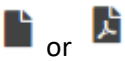
NB2 : when clicking on Biosamples (top of the left menu bar), you can see the list of all the Biosamples of your project

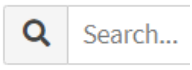
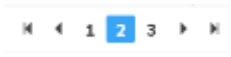
## 5. A recap on the different xlsx templates

1. Template for **physico-chemical or qPCR data** (to import the data corresponding to such monitored parameters)
2. Template for **biosamples** (samples likely to be sequenced)
3. Template for **biosamples included in a given sequencing run**
4. Template for **sequencing metadata** associated to the sequencing run (basically the same template as above, to further fill in)
5. Template for **fastq files**

## 6. Common browse buttons

| Widget | Function |
|---|---|
|  | Tooltip: a message pops up when the mouse pointer hovers over this symbol |
|  | To edit the corresponding data |
|  | To delete the corresponding data |
|  | To delete the corresponding item |
|  | To delete the corresponding item |
|  | Mandatory field |
|  | To add a missing element |
|  | To enter new data of the indicated type (current example: a new series) |
|  | To enter a new biosample (individually, by filling a form on the interface) |
|  | To import biosamples (batch mode, using the xlsx template) |
|  | To export biosamples (xlsx file) |

| | |
|---|---|
|   | To generate a biomfile from the selected biosample results |
|  | To view the project history |
|   | Save button |
|   | Close button |
| Close | Close button |
|  or | Indicates a file which can be downloaded from the interface |
|  | Indicates that the required data are complete |
|  | Indicates that some required data are still missing |
| View details | Expand button |
| Hide | Collapse button |
|  | Expand button |
|  | Collapse button |

| | |
|---|---|
|   | Search bar |
|  | Check box |
|   | Pagination bar |
|  | Pagination bar (for Operating conditions) |
|  | Left and right pagination buttons (for Operating conditions) |

## 7. 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 greatful 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.


Deep
Omics


INRAE


CARNOT
3BCAR


PROSE
PRocédés biOtechnologiques
au Service de l'Environnement
université
PARIS-SACLAY


Lbe


OPAALE


MaIAGE
migale