

INRAE

DeepOmics user guide

Digital Environmental Engineering Platform for Omics data

Ariane Bize 08/12/2023





Table of contents

| 1. | Introduction: what is DeepOmics? | 4 |
|-----|--|------|
| 2. | License | 4 |
| 3. | Funding and acknowledgements | 5 |
| 4. | How to acess DeepOmics server? | 6 |
| 5. | Which bioinformatics pipeline to use? | 9 |
| 6. | Easy16S, a complementary tool for statistical analysis of microbial ecology omics data | . 10 |
| 7. | File formats for the import of biom files and metrics | . 12 |
| 8. | DeepOmics key concepts | . 13 |
| 8 | 1 Lab-scale process data | . 13 |
| | Project's input management | . 13 |
| | Experimental series | . 13 |
| | Operating conditions | . 13 |
| | Replicates | . 13 |
| 8 | 2 Industrial and field process data | . 14 |
| | Sampling campaigns, Sampling site and Biotechnological Process | . 14 |
| В | iosamples | . 14 |
| 9. | DeepOmics structure and key functionalities | . 14 |
| 9 | 1 Projects | . 14 |
| 9 | 2 Lab-scale process data | . 17 |
| 9 | 3 Industrial and field process data | . 20 |
| 9 | 4 Meta-omics analysis | . 22 |
| 10. | Query in DeepOmics | . 24 |
| 11. | Useful browsing tips | . 26 |
| 1 | 1.1 Physico-chemical and qPCR | . 26 |
| | Data configuration | . 26 |
| | Data import | . 26 |
| | Data management | . 26 |
| | Data display | . 26 |
| 1 | 1.2 Meta-omics analysis | . 27 |
| | Sequencing runs | . 27 |
| | Biosamples | . 27 |
| 1 | 1.3 Common browse buttons | . 27 |
| 12. | Informatic structure | . 30 |



| List of illustrations | 31 |
|---|----|
| Memo on DeepOmics functional structure | 32 |
| A recap on DeepOmics structure | 32 |
| Left tab: labscale | 32 |
| Middle tab: industrial and field | 32 |
| Right tab: meta-omics | 32 |
| Quick start / Memo on how to enter a new dataset into DeepOmics | 1 |
| Main steps of the procedure | 1 |
| Summary table | 1 |
| Project, reactors and processes | 1 |
| Physico-chemical data | 1 |
| Amplicon sequencing data (e.g. 16S rRNA gene metabarcoding) | 32 |
| A recap on the different xlsx templates | 33 |



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 for 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 accepts both **single-end or pair-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: <u>https://deepomics-info.hub.inrae.fr/</u>.

In the near future, we intend to enrich DeepOmics by developing new features. New types of metaomics 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

All rights reserved. In the future, DeepOmics may be released under the GNU Affero General Public License (AGPL).



3. Funding and acknowledgements

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

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.







Lbe







4. How to acess DeepOmics server?

DeepOmics server is located at the following address: <u>https://deepomics.inrae.fr/</u>

If you depend from Renater federation (national research and education network in France), you can connect to DeepOmics with your Renater login and password. You will then directly access to DeepOmics public projects in read-only mode.

If you are interested in creating your own projects and entering data, please contact us!

If you do not depend from Renater federation (e.g. private entities, abroad academic entities), please <u>contact us</u>!

| Gringes | - XNSIDE COM |
|---------|-------------------------|
| | |
| | Authentication required |
| | Photo-log is. |
| | |
| | |

Overview of the sign-in interface

Click on *Log in*.

You are directed to the authentication interface of Renater federation. Select your organization, and then enter your usual login and password.



Overview of the sign-in page in Renater federation (step 1).

| RÉPUBLIQUE FRANÇAISE Junet Autor Autor Antonio | |
|--|----------------------|
| Bienvenue sur le portail d'authentification aux applications INRAE | Support Applications |
| Veuillez vous authentifier | |
| | |
| J'ai perdu mon mot de passe | Mentions legales |

Overview of the sign-in page in Renater federation (step 2)



You are now connected to DeepOmics IS through your personal account.

| mis Contraction | 1 | P.L. | Anufit Bat gwelad t | ECASTIAN T |
|--|-------------|------------------|--|-----------------------------------|
| Hame Projects data dealy | in 🔰 🕹 Ada | ninistration * | of DADA3 | Ø Easyte |
| Projects | | | | |
| Q lasety. | | | | More property |
| Project - | Statue # | Coordinator g | Detail # | |
| | | | | |
| 400 00 denne AD indalle Poirier Brenn project AD indallition statu, Simun Poirier | - | A OWNER COMPLEUN | Start: 2013-10-01. Exil: 2010-10-01 Contact person: Oficio (2009-10-00 2 oppriveental onio 6.1 Sampling companys | • Steel |

Overview of a user welcome page

You can start using the interface. When connected, you have access to the public projects of DeepOmics, including 2 demo datasets:

- 00_demo_AD_inhib_Poirier: a demo dataset for lab-scale experiments; it contains data from Dr Simon Poirier's PhD work (2013-2016) on anaerobic digestion inhibition by phenol or ammonia (2 experimental series, with 48 biosamples for each). This project was supervised by Dr Olivier Chapleur (INRAE-PROSE).
- *00_demo_AD_plants*: a demo dataset for full-scale processes; it contains data related to the feeding and sludge sampled from 6 various AD plants, (one time point per plant, total of 20 biosamples). The data has been anonymized.



5. Which bioinformatics pipeline to use?

To analyze amplicon sequencing data, processing them with bioinformatics tools is required before data interpretation.

In DeepOmics IS, we provide the link <u>https://forgemia.inra.fr/cedric.midoux/deepomics16S</u>) to the pipeline we recommend for <u>16S rRNA gene metabarocoding analysis</u>, for 4 main reasons:

- this pipeline is based on state-of-the-art tools, including DADA2¹ and FROGs²;
- the output files generated by the pipeline are directly compatible with import into DeepOmics;
- if all the users employ the same bioinformatics pipeline, it will promote data homogeneity and intercomparability;
- the pipeline is suitable for an information system, enabling the addition of datasets by batches or individually, without the need to run the pipeline again on the whole database.

As stated in the introduction, **both single-end and pair-end data** are supported presently.

| Jacp/ Similas | () | | A ALMANE & LE DIWIELAG DE CASTINA |
|------------------|------------|--------------------------------------|-----------------------------------|
| # Home | I Projects | 🕫 mas-analym 🔰 🕹 Administration 👻 | e DADA3 e Easy158 |
| Projects | | | |
| Q isect. | | | + Monepropert |
| | Linl | k to the recommended bioinformatic p | ipeline (aitlab page) |



Overview of the Gitlab page dedicated to deepomics16S bioinformatic pipeline

¹Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581–583 (2016). <u>https://doi.org/10.1038/nmeth.3869</u> ² Escudié, P, Auer, L, Bernard, M, Mariadassou, M, Cauquil, L, Vidal, K, Maman, S, Hernandez-Raquet, G, Combes, S, Pascal, G. FROGS: Find, Rapidly, OTUs with Galaxy Solution, *Bioinformatics*, Volume 34, Issue 8, 15 April 2018, Pages 1287–1294, <u>https://doi.org/10.1093/bioinformatics/btx791</u>



6. Easy16S, a complementary tool for statistical analysis of microbial ecology omics data

In DeepOmics, we also provide the link to a complementary tool named Easy16S, which can be used to lead classical multivariate and other statistical analysis of biom files generated by DeepOmics. Easy16S code is available on Gitlab (<u>https://forgemia.inra.fr/cedric.midoux/easy16s</u>).



Easy16S (<u>https://shiny.migale.inrae.fr/app/easy16S</u>) is a user-friendly and free access shiny web application that enables the dynamics visualization of count data in microbial ecology (biom or other



Overview of the Gitlab page dedicated to Easy16S

The plots generated in the interface can be downloaded as images. Moreover, the code used to produce each plot can be displayed, copied and pasted in an external text file, to keep trace of the analysis.





Overview of Easy16S interface (using one of Easy16S demo dataset)



Overview of the code display functionality in Easy16S



7. File formats for the import of biom files and metrics

If you are using the pipe-line recommended for DeepOmics, the biom file will be in the correct format and the metrics file in the correct format will automatically be generated.

The biom file must be provided in format biom-1.0.

https://biom-format.org/documentation/format_versions/biom-1.0.html

A metrics file must also be provided. It must be a tab-separated file with the following headers:

"Sequencing sample code": the code for the biosample/fastq (same as in DeepOmics and as in the biom file).

"Input reads": number of raw reads for each sample

"Post-process reads": number of reads after the preprocessing, for each sample

"Nb ASV": number of amplicon sequencing variants in each sample.

In this metrics file, each line corresponds to one biosample.

An example of such file is provided below:

| Sequencing sample code | Input reads | Post-process reads | Nb ASV |
|------------------------|-------------|--------------------|--------|
| DIG_3_f_MAX | 43006 | 33621 | 56 |
| DIG_3a_s | 63190 | 31427 | 192 |
| DIG_3a_s_MAX | 71649 | 37428 | 237 |
| DIG_3b_s | 81791 | 42241 | 207 |
| DIG_4_f | 108888 | 80711 | 236 |
| DIG_4_s_MAX | 68089 | 38158 | 243 |



8. DeepOmics key concepts



8.1 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.



8.2 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.

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).

9. DeepOmics structure and key functionalities

9.1 Projects

The DeepOmics data warehouse is organized into *Projects*. Each connected user can create projects and indicate the project *Coordinator* (himself or a third person). The project *Coordinator* can in particular define:

- the project *Status*: *private* (all the project data are private) or *public* (all the project data are public); the project *Status* can be modified at all time;
- the list of *Participants*: all participants of the project have the same rights, namely reading and writing data;
- the list of *Observers*: observers have access to the project in read only mode.

| mics | and a | K B A | A ASIAN DISC DWYEZ | |
|--|-------------|--------------------------------|---|--------------|
| # Home Projecta III Mole analy | an Adr | ninistration * | & DADA2 | & Easy16 |
| and with the fi | | | | |
| Projects | | | | |
| Q bart. | | | | + les prijet |
| Project + | Status S | Coordinator a | Detail e | |
| 100 00 deine AD inhib Poirier Deen project AD fehibilise data, lienen Poirier | - | & Olivier CHAPLEON | Start: 1913-10-01 First: 2018-10-31 Contact parson: 1994ar CH49LEDM 2 experimental sector & 1 Sampling ratiopages | · Une |
| 100 demo AD plants | | A MOME BUT IT MEDIC DE CASTERN | Nort: 2011-03-11 End Cardad period: Allohe BUE DVW224CDK CASTERN | B.Stev. |

Overview of a user welcome page, with the project table

| | * 00_demo_AD_inhib_Poiner > | | | | |
|--|--|---|--|----------|-----|
| ator I | | | | | 2 8 |
| | Demo project: AD inhibition data, Simon Politier | 60de) | 00, Jenu, AD, Johb, Pokier | | |
| -thung 6 | ORVIN: CHAPLESH | Cantact persons | Officier CHAPLEUR | | |
| ight i h | inhibition, asserobic digestion, microcosm, batch, phenol, anniormen | n,TiFowashn | | | |
| tart slots 1 3 | 1813-19-91 | | | | |
| ad dates 3 | 2016-10-30 | | | | |
| hataasi ji | putrit | | | | |
| uticiaents 3 | Since FORER, Angelias GUERNE, CHRYSTELLE BUREAU, THEODORE | | | | |
| tustwere: + | erupty. | | | | |
| | This project contains data from the Simon Paikles's PhD work on an error These data over entered in the data watshoose as a done dataset. | olo digestion inhibition by greensi or arminesia. | This project was supervised by the Orlivier Or | tambrot. | |
| ogent's website : | | | | | |
| nijett's microsco - | | | | | |
| ciontific articles — e est patients : | antal) | | | | |
| ther documents: | 2016, Weinsteinauch, Physiol.pdf 12,9 Mill | | | | |
| | 2016, MTE, Americana, and (4.3-86) 2018, Datasetistic (Princip), Americana M, 11.3-861 | | | | |

Within a project, 3 main modules are available, accessibles through tabs, from left to right:

- 1. a tab dedicated to *Lab-scale process data* (or small reactors or pilots)
- 2. a tab dedicated to Industrial and field process data (from full-scale processes)
- 3. a tab dedicated to Meta-omics analysis and data

Within each tab, a dynamics menu bar is available on the left, whose item list is contextualized according to the current screen.



| # Hame | I Projects | 1 If Inste analysis 1 | As Administration * | | | & DADA2 | ∂ Eaty16 |
|---|--------------------|-------------------------|------------------------------------|--------------------------|-----------|-------------------------------------|---------------------------------|
| an admin out | HALF STREET AND ST | Poirier * 👻 view dotate | | | 7 | | |
| -built | scale process data | | he tertantrial and Seld process da | da . | | III Meta-emic | a analysia |
| | | | | | | | |
| beratory | 10 m | | | | | | |
| Project's leput | 上 Experi | imental series | | | | | |
| Project's Input an agenvent | L Experi | | | | | | Anternation |
| Project's Input an agenvent | - N | | Custact person | Start date | | Operators | + Mortena |
| keratory Projects input anagement Cognomentation | - N | | Contact person | Start date 2014 03-03 | Simus POR | Operators ex, crangiture europau | - |

Overview of the tab "Lab-scale process data"

| # Home | @ Projects | III. Alata analysia | 2s Administration • | | & DADA3 | Ø Lasy16 |
|--|--------------------|--------------------------|-------------------------------------|--------------------------|--------------------------|----------------|
| Project "00_t | demo, AD_inh/b_ | Poirier " 🗸 ver datala | | | | |
| the Labor | cale process slata | | In technicital and field process de | te | III Meta o | mici analysis |
| | | | | | | |
| 2.1.(0 | - | a 44/4 2020/04/2010/04/1 | | | | |
| deatriel #Sampling stres | 🎒 Samp | ling campaigns 🛛 | | | | |
| a Sampling stres Histochriological oceanes | A Samp | | | | | + New Language |
| | | | Contact person | Start dels | Operators | + Row company |
| a Sampling stres , the technological occurates | | | Cardact person | Start data 2004-01-05 | Dpenators Simue POOLE | |

Overview of the tab "Industrial and field data"

| # Home | 🖙 Projects 👘 👘 | into analysis 🕴 🎍 Admin | istration * | | & DADA3 | & Kasyis |
|--|-------------------------|-------------------------|--------------------------------|--------------------|---------------------------|-------------------------|
| Project *00, | demo_AD_inhib_Pairier * | ₩ View datalia | | | | |
| <u>h</u> lab-s | cale process data | | Industrial and field process (| fata | 10 Marca | emics analysts |
| lixinformatics | | | | | | |
| L Monarryska B Biosampie results ogsensling runs | Ca form | 15 0 | | | | + newsequencing |
| Billionample results | | Sequencer | Satur | Date | Uter | + Terrestan |
| B Electropie results equincing runs | Q forth | | Statun Festival | Data 2015-01-25 | User Cherstelle Buneau | Resources to the second |

Overview of the tab "Meta-omics analysis"



9.2 Lab-scale process data

This tab is adapted to process reactors for which highly detailed information is available (operating conditions, nature and amount of the influents, etc), and/or if the experimental design includes replicates.

It contains functionalities related to:

- *Project's input management*: for lab-scale process data, the pool of inputs added to the reactors must be defined in the first place; inputs are categorized according to a controlled vocabulary and they include all types of residual bioresources (waste, sludge, biomass, effluent), as well as gas, chemical compounds, matrices, pure microbial strains, etc. The inputs are defined at the project level.
- *Experimental series*: the experimental series are structured into *Operating conditions* and, within each operating condition, into *Replicates*.

| Annihiti in) Coperimental union Annihiti in | ct's input management 🛛 | Sin Industrial and field process data Discounts:/Food And Pood Processing Hasts/Jargo Po Ports * resulted food biscounts Company name Enter the screaming Visaria | roducers, industrial (Contparty websity | ereks analysts Blosaenple: 0 🔶 Juli |
|---|--|--|---|--|
| Project Aspar senger sent Dispertmental webs Copertmental webs | pats work Ooktand bod processing series anached lood binnessin minited lood binnessin minited ser paste series skalge | reaction from bioreaste Company ruana | Company activity | Simample: 0 🔶 Add |
| - 52 Ho - 12 H - 12 - 12 - 12 - 12 - 12 | worke worken food processing worke a transmissionen, bekentlich waarheid food blemaste waarheid food blemaste | reaction from bioreaste Company ruana | Company activity | Blosemple: 0 🔶 Autr |
| - 54 | a Domestre Nood wante AD skudge pellet Norikol Sciences Norikol Compositio Hangen gas | City Chenaudin Callectionidau Unity collectionidati. Convenits | Country France Year 😧 | |
| Over | | n the interface "Project's inpu | t management" | |
| . Et Lats scale process data | | In technicial and field process data | 18 Nota | onici analysia |

| a contra | Search and a second search | | a reconnariae and a process of | | in the one and it |
|---|----------------------------|-------------------------------|--------------------------------|------------|--|
| aboratory Project's input anagement | ம் Expe | erimental series | | | |
| D Capacity and A cardina | Q ber | D) | Contact person | Start dete | + Reveale |
| | 8 | 1 NH4 Arstronia inhibition | Colorier Completion | 2014-03-03 | Simul FORER, CHIMITELE BUREAU, Shive CRAFLEDR |
| | | D Phonoi Phonoi antibition | ≜diserce#008 | 2014-02-24 | Singer PORCE, CHRYNELLE BUREAU, ODARC CHARLEUR |
| | | | 0.603 🗉 | 1604 | |

Overview of the interface "Experimental series"





Overview of an experimental series selected in the list of "Experimental series" (Operating conditions)

The *Replicates* are visible at the bottom of the interface for *Operating Conditions*.



Overview of an experimental series selected in the list of "Experimental series" (lower part: Replicates)

Within the *Experimental series* interface, you can notice, in the left menu bar, a group of functionalities dedicated to *Physico-chemical and qPCR* data (*Data configuration*, *Data import*, *Data management*, *Data display*).



Overview of an experimental series selected in the list of "Experimental series" (Physico-chemical and qPCR)

In the left menu bar, the *Biosamples* functionality displays all the biosamples of the considered *Experimental series*. *Biosamples*, created by the project users, are the virtual equivalent of a real experimental sample and they are defined in order to subsequently upload the associated amplicon sequencing data.

| W Home 🕸 | Projects: III Imma analysis | Le Administration * | | & DADA2 & Easy1 |
|--|--|--|--|--|
| | | | | |
| Project *00 demo | o_AD_Inhib_Poirier * 🗸 Yiw data | 14.1 | | |
| - Figher so_perio | "Tip"huno"Louise | m. | | |
| 2) Left scale pro | weeks data | Son invituativial and field process of | data | III Meta-ornics analysis |
| | | | | |
| aboratory | L H | | | |
| Prejectsinger | b Experimental series * <i>NH</i> | 4 Yiww datalis | | |
| Aboratory Projects input Inside the input Examples of the input Ex | Experimental series *NH Biosamples | 4 Yew datalla | Lagest Binarryke | M 🕂 Ingert Dissanguiss 🕈 Novellinsanguis |
| Prejects input tanaproved | | 4 v Yiwe details Sampling boation | La fayor filmeriyan | 👻 🕂 Hoped Dissonation 🕈 Hord Housenple Description |
| Projects input management & Equations to takes | A Biosamples | | Name and a state of the state o | CONTRACTORISTICS ESCONTRACTORIST |
| Projects input management & Equations to takes | A Biosamples Cede | Sampling location | Date | Description |
| Projects input management & Densitivental sorties Bio Colessiting conditione & Colessiting conditione | Biosamples Cade | Sampling location 1944, 2000 - Hits, 2000, Hug2, 1 | Data 2018-10-21.11.000 | Description Som cell pellot |
| Projects input management & Dentimental series Biomanytes Trystes Chemical and gPCR | Biosamples Cade S_21,0002,dispase E_22,0002,dispase | Simpling location 1944, 2000 - Hint, 2000, Hall, J. 1944, 2000 - Mint, 2000, Hydr, J. | Date 2014-00-21.12:00 2014-06-20.10:00 | Description. Som off paikt Vision off paikt |
| Projects input management & Departmental series Bio & Celenstring conditione | Biosamples Cede S.m. 0994, dayam E.21, 0994, dayam S.m. 0944, dayam | Sampling location 1844_2000 Hird, 2000, Hird, 1 1844_2000 Hird, 2000, Hird, 1 1844_2000 Hird, 2000, Hird, 1 1844_00000 Hird, 2000, Hird, 1 | Deta 20154-00-21 12:00 20154-06-20 10:00 20154-06-20 10:00 | Description. Som off pailet Non-call pailet Som call pailet |

Overview of an experimental series selected in the list of "Experimental series" (Biosamples)



9.3 Industrial and field process data

This tab is adapted to data from full-scale processes (no replicates, usually less detailed information on the process, parameters and influents). It is structured into *Biotechnological processes*, which are associated to *Sampling sites*. The process type is defined according to a controlled vocabulary. Moreover, a form specific for each type of process is available to describe the process parameters (currently available: wet and dry anaerobic digestion, activated sludge; the other processes have a basic form until further developments).

| ti Lab-sca | de proceso data | - Mit finduntrial and the | il process data | ill Neta-onic | z anatysis |
|--------------------------------|-----------------------|---------------------------|-----------------|--|------------|
| tedustrial Incharging star | Sampling sites | | | | |
| Elistectmological processes | Q Seath. | | | | + New Di |
| W Sentilut contrellor | Reme | Operators | Syndicate | Location | Processe |
| | has Silv_1 Hite_5 | | | Oty: Bugiory Country Osofficies | (8) |
| | he Site ,2 title,2 | | | Oty. Beginn Country Country | (8) |
| | In Site 3 | | | City: Register Citatetry: | (a) |

Overview of the interface "Sampling sites

| ,biab-so | elle garacess data | he solution and field process data | III Neta-ornics analysis |
|--------------------|----------------------------|------------------------------------|--|
| industrial | | | |
| in Sampting sites | Biotechnological processes | | |
| L tiotadovinginal | Q bench. | | + New Proces |
| Sampling compaigns | Name | Site | Category |
| | ▲ DG_36 96_35 | 19 ₈₋ 3 | treatment/atmendatic/dignation/sec/catt |
| | ▲ DG_6 05,8 | Sec.3 | to salement, 'arisen shen, 'd'ago shen, 'd'ago shen an anni she idigentar |
| | ₫ DG_2 06,1 | Site_1 | trainent/assentix/digitize/web/batch |

Overview of the interface "Biotechnological processes"

Sampling campaigns can then be defined and each of them is associated to one or several Sampling Sites.



Overview of the interface "Sampling campaigns"

| <u>th</u> Lab-sea | ie process data | | In Industrial and Rold process | data | | 11 Heta-onvica analysia |
|------------------------------|------------------|---|--------------------------------|-----------|----------------------------------|-------------------------|
| helicetrial | A Samplin | g campaign "AD_sludge" | | | | |
| a langing atos | •• Sanipun | Reambarbi un-amoide | | | | |
| Biotechnological vocesses | Name: Tags: | KD_sluttys studge, assiantbit digestion. | | Coder | AD_stickpe | |
| Stampling campaight | Eprind provid | ANNAL BIZE ITWEEZACTIE CASTERN. | | met date: | #121-81-10 | |
| | Openators . | AROANE RIZE O'WYEZAC DE CASTERIA | | Tibes) | 58v_4,59v_1,50v_2,59v_3,58v_5,58 | 0.0 |
| D_shullge | Dummary I | | | | | |
| the amples | Sampling schores | 100405/ | | | | |
| hysics-chemical and gPCR | | | | | | |
| Date configuration | Proceduras | earing), | | | | |
| Itatainpart | Nour thata: | weitth | | | | |
| | Other documents! | examply | | | | |
| E Deta mariogeneent | | | | | | |
| 🛃 Data diquiny | Atlan | | | | | |

Overview of a selected "Sampling campaign" associated to several "Sampling sites"

Similar to the *Experimental series* interface, you can access to the *Biosamples* related to a selected *Sampling campaign* in the left menu bar.

And you can also notice, in the left menu bar, the group of functionalities dedicated to *Physico-chemical and qPCR* data (*Data configuration*, *Data import*, *Data management*, *Data display*).

| ≜ Lab-scale pro | cont data | Senstantist and field process | ninea - | III Heta-ontics analysis |
|--|----------------------------|---|--|---|
| Industrial | Sampling campaign "AD | studge " Vewithtally | | |
| la Sompling sites | Jamping campaign into | anonge of the tran | | |
| A tixtucheckapeal pricesses | Biosamples 0 | | | + manual la constante + mon la constante |
| H Sampling campaigns | Code | Sampling location | Oute | Description |
| | | | | |
| D_statige | 100_0_3 | Ste_8>01_605_8 | 3451-69-09 10025 | anaesothic duritge- |
| Tenengter | 06.5.3 | She.4> 00_605_8 Mrs_5>05_305_8 | 2021-05-05 10:57 | anaentiik dudge |
| Transmerer Trysko chemical and gPCR | | | | |
| Transmose hysico chemical and gPCR | DEAS | Mw_N+05_305_3 | 2021-04-27 (2:0) | anaestik sludge |
| Efficiency of the second secon | 08,3.3 08,4.3 | \$he_5>05_305_8 \$f8_4>05_405_4 | 2023, 64 (27 12/00) 2027-04-20 12/00 | anaentia statije enerotia statije |
| AD_stadge AD_stadges AD_stad | 06.5.5 06.4.5 06.6.5 | 506_3 > 05_305_8 502_8 > 05_405_4 506_3 > 105_3405_3a | 2021-04-27 12:00 2021-04-25 12:00 2021-04-14 12:07 | anaentin stadige etoerobic stadige arteenstik stadige |

Overview of the "Biosamples" interface, within a "Sampling campaign"



| 上 Lab-scala ; | process data | | in Anderson | tal and field process da | (4) | | III Meta-omic | u setudynia |
|--|------------------------|------------------|---------------|--------------------------|--------------------|-------------|----------------|---------------|
| Industrial | Sampling campa | ing that she | lge v Mem dat | | | | | |
| In Semplegatus | n Sampung campa | ign nv, sevo | ge www.ma | 210 | | | | |
| A Boltenvological processes | S Monitoring data | management | | | | | | + Atlantini |
| A Sampling camping on | A COLUMN AND A COLUMN | 11 Perm selected | | goor selection . | teleine aufin than | | | |
| AD_studge | | | | | | | | |
| A Rewardstee | | | | | Process | | | |
| | | | | | | | | |
| Physics-chemical and gPCH | | ĉ | S. | ÷. | 2 | - | 10 | 5 |
| | Monitored Newsone Type | C 90 - 13 | Chart | Case C | 1.90×1 | 4+ 00.4 | 5 - DG , S | 1 × D6, 4 |
| O Data configuration | Monitored Newsone Type | (.)961.00 «.1.3 | 3 * D6_304_3 | pail, pa + it, | 3 - 06, 3806 | 4=06,400,4 | | 1 > D6, 504, |
| O Data varfiguration O Data import | | e_1 = 06_106_3 | 12×06_200_2 | 1. 50% 50 ° C | 31-00,300 Jb | 4+05,405,4 | 3 - DG (20G, 3 | s> 06, 404, 4 |
| Ditksonfigunities Ditksonfigunities Ditksonfigunities | | 1.90,00+1 | 3-00,300,3 | at pole, pole | 4C 904C 90 41 | 4+ DC 400,4 | 5,001 | 0.906,904 |
| Ditaisanfigunitat | | | | | | 01,4 | 5,000 | |
| Physics-chemical and spira Data configuration Data configuration Data import If then an exagence the Data display | Acatalar (g/L) | 18 | 18 | 0 🖩 | 0. | 0 | 1 | 0 |

Overview of a sampling campaign the interface "Sampling campaigns" (Physico-chemical and qPCR)

9.4 Meta-omics analysis

This tab is dedicated to amplicon sequencing data. The left menu bar contains 3 main groups of functionalities.

The first group, *Bioinformatics*, enables the visualization of all the *Biosamples* from the *Project* (all *Experimental series* and all *Sampling campaigns* together), and their associated amplicon sequencing results.

| solutionatics | | | | | |
|---|----------|---|--|---|----------------------------------|
| Tiesanske Filmanske coarts | Biosa | mple Results | | | |
| idmected units | Q, Seers | 6.) | | | + Create Secretia |
| Photoduras | | Sequencing sample code | Metabarcoding bioinformatic run code | Processing instrics | Humber of Armstation |
| PCR10mBlane Relevant sal | - 5 | A \$_00_00N2_day000 | III NB4_bioinfo | Number of rawoeshits: kites4, Post process reads: 29495, | 174 Selal genetation must |
| Sequeraling rans | | Sample (S_III_EUR2_dayNU) | Workflow- studiet, Jrogs (status prinate) | Mambie of ASVE 114. | 20488 |
| | | | | Namber of one couch: STUD. | 10 |
| A COLUMN TO A C | | \$ 5 00 00X1 double | | | |
| Nofernatic Analyses Software Bioinformatic workflow | => | ▲ 5_00_00N2_day009 tample : 1,10_0001_day008 | III NB4_Ideinfo Weikflow dada2_bigs (dottargeriede) | Past process ready 22442, Asureon of Adve.181. | Taital annotation on at 20002 |

Overview of "Biosample results", for the project "00_demo_AD_inhib_Poirier"

The second group is dedicated to the *Sequencing runs*. The *Sequencing runs* can include *Biosamples* from all the considered *Project*. In the results of a *Sequencing runs*, the raw *fastq* files are in particular available for download, in a compressed format (*.gz*). Several functionalities are dedicated to the



metadata required to describe the *Sequencing runs* with relevant parameters (*Procedures, PCR conditions, Relevant url*).

| Lisb-mai | le process data | in the | Industrial and field process d | yta | III Here | omics analysis |
|--|---------------------|-----------------|--------------------------------|------------------|----------------------|--------------------|
| deformatica Biocurptes | Sequencing run | 5 0 | | | | |
| Boards moto | Q Senti | | | | | + New sequencing |
| Proceduces | Runname | Sequencer | Status | Date | User | Procedure |
| PDI conditions Relevant urt | X NB4 40 samples | INBAE PROSE FOM | Dilibert | 2015-01-23 | CHRINEFFELLE BURKEAU | Vera Toky gotpu |
| Sequencing turn deformatic Analyses | I Phenal | NAME PROVE HOM | rested | 2015-01-15 | CHRISTELLE BUREAU | Press, publication |
| Soforen Skiefernatic waithw | | | 0.0 | 1 • 1 • 1 | | |
| enteoo | | | | | | |



| thist-sce | is process o | lata. | | San bredsortchiel avoid | field process data | < | | III Meta ornica analysia | |
|---|--------------|--|------------|--|--------------------|----------------------------------|-------------|------------------------------------|---------|
| Bininformatics | X | Sequencing run | If and a P | | | | | | |
| Biosargine | A | sequencing run | PWI 199 | where retails | | | | | |
| Il Diosorpie ietulta Inquancing runa | д | Sequencing sam | ples | | | | | | |
| Procedures | | | | | | | | | |
| PCR conditions | | A hearth | | | | | | | |
| Ø Relevant sat | | | Matrix | | Targeted gene or | Terprind | Target | Avaiathie for | |
| K Sequencing rum Bahrformatic Analyses | | Code | type | Per Poloners | genomic region | taxonomic or functional group | withagments | Hetabartodleg Bioinformatic Ran | Actions |
| ∲ Software | | Δ | | | | | | A | - |
| Okinformatic workhow recordure | 2 | 5_00_00N2_day000 hample: .5_00_0200_day000 | DNA | NUM ICTOROPICALITORITORIASTI . | 315-1894 | Archaen, Bacterla | 10,04 | . S. M. MAL MARKED SAME | |
| Biointernatic workflaw | | Δ | | | | | | | |
| II Hetabercoding, ioinformatic rus | 3 | S_00_00N2_day009 hample1 %_co_unit2_dayout | 260. | SUR (CCCCPCARTCH(TTHACT), SUR (CPCMCARCAGEDICSCO), | 385-594 | Archani, Bacheta | 35.99 | S. Jan, Janes, Jayoni beng | ,ea |
| Sequencing uniples | | 5 06 00N2 day029 | - | SUMPLECOCOFCAATTEHETTBAGT) | | Archana . | 140.000 | 0 | - |
| Sequencing sample execution | > | 5_00_0052_day029 5ampte: 5_00_0052_day536 | -0104 | SUM (COCOCAATTON THACT) , SUM SCHONCAGE MOCCOCOCOCOCO | 335-1696 | Bacleria | 15,596 | \$ 5,00,0000,00y025.5mm | a l |

Overview of a sequencing run selected in the interface "Sequencing runs"

The third group, finally, is related to the *Bioinformatic Analyses* of the raw reads. Again, several functionalities ensure the adequate description of the tools and pipe-lines (*Software, Bioinformatic workflow*) used for the analysis. For each sequencing run, at least a biom file and a specific metrics file are loaded, and they can be subsequently downloaded from the interface.

| | | | | | | | Digital Environm Platform for On | nic mental Engine nics Data |
|---|-----------|--|-------------------|----------------------------------|-----------------------------------|---------------------------------|-------------------------------------|-----------------------------------|
| Project *00_c | terno_AD, | | Mesa detalla | in industrial and field proc | ess data | | TH Neta-order analys | 61). |
| Bioinformatica | | | | | | | | |
| A Blonarriptos | 118 | Metabarcoding inf | formatic runs | | | | | |
| Ell Tanareple results Sequencing runs | 0 | L foot(h. | | | | | + less netatorcolin | peternalisme |
| Presentario | | Code | Sequencing sample | Bioinformatic workflow | illion file | Hetrics file | samments | action |
| PCR conditions Balement cal Elegenerichgraphi | 3 | 100 NH4_bioinfe 1004 filosoformatic analysis | - | statlad Joogs (statusgehoate) | B abundaronation (altern metal | B methodates Algebra tamante | | |
| Beinhematic Analyses 9b Software | | III plarnet_bioinfe Pressil kintefermativ analysis | | dadwa_bogs (statusprivate) | aturdanin biom mitut harmai | B metricanie Referi facence | | 20 |
| Elsénhernetic unektion. procedure | | | | 4 | D | | | |
| | | | | | | | | |

Overview of the interface "Metabarcoding bioinformatic run"

The bioinformatic results from individual samples can be combined into a new *biom* file in the *Biosample results* interface, by selecting the desired *Biosamples*.

| | le process data | | in industrial and field process date | | III Meta-centra analysis |
|--------------------------------|-----------------|--|--|--|-----------------------------------|
| Informatics Nonartyles | Biosa | ample Results | | | |
| Court (in main | Q Sea | | | | + Qualichian/He |
| roordures. | \wedge | Sequencing sample code | Netabarcodieg bioinformatic run code | Processing metrics | Humber of Annutation |
| wewant.urt | | <u>I</u> S 00 00N2 day000 | III NH4_hiomfa | Number of convents: 53894, Post process mode: 26486, | 174 Total association count : |
| spectrum | | Sample - \$_00_0002_tby/00 | Workflow: didd_(Hegs (distantisticate) | Austriller of ASVL 174. | 30455 |
| risrmatic Analyses | | Walter Land College | 100000000000000000000000000000000000000 | Number of row results: 17822, | 140 |
| ictown . | > | A 5_00_06N2_day409 barryin: it un rents staynos | III NH4_biolate Warkflow: dodat, frigt (status private) | Past process mode 22442, Number of ASV, 183 | Total association count: 23942 |
| sinformatic workflow inform | | | | | |
| Antonestic workflow | | <u>II</u> 5_00_08N2_day629 | III NB4_bininfa | Number of raw wards 58545, Post prices words, 25453, | 22m Tatal accordance cannot |
| Walternaling derivation an | | Sweepin : 9_00_00962_0890029 | WeekTow (deda2_hoga (datumprivate) | Norther of ASE 328 | 3940. |
| | | A 5 00 00N2 day042 | III NH4 bioinfo | Number of our could: 116526, First process made: 47726, | 200 Total according counts |

Overview of "Biosample results", including the funciton "Create biomfile"

10. Query in DeepOmics

DeepOmics data can be queried and modified through an API (Application Programming Interface).

https://deepomics-api.prose.inrae.fr/

Application tokens can be obtained directly on DeepOmics server, by clicking on the *App tokens* link below the *Logout* button.



| En al anti- | 1 | RIA | | C DE CANTENA .* |
|--|-------------|-------------------------------|---|-----------------|
| ff Herne Projects II Mice analysis | 2e Ad | ministration * | e PAppa token | Ø tasyt |
| Projects | | | | |
| Q Such. | | | | + new project |
| Project + | Status B | Coordinator ± | Detail ± | |
| 100 00_demo_AD_inhib_Poirier Same project: A5 Inhibition date, Simon Public | 4 | 🚨 Citaler CitAPUDUR | Thart; 2015-03-07 Feed: 2000-10-31 Contact person: Oblive CrWUUUM L'experimental seties & L'Sompling compaigne | 0 stow |
| 60 denne_AD_plants | 4 | ANNUE RUE D'ANEZAC DE CARDENA | Start: 2011-01. Then: - Contact person: AniAME 001: DNAT (ACCHE CARTERN In experiencental series & 2 Sempling campoigns | • Stin |

Overview of the access to "Application tokens" page



Overview of the "Application tokens" page

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



11. Useful browsing tips

11.1 Physico-chemical and qPCR

Data configuration

How to add a new monitored parameter?

Select one parameter in the left panel tree and fill the form.

How to edit or delete a monitored parameter?

Select one Monitored Measure in the right panel tree. If data has already been entered, it is impossible to: **Delete** Monitored Measure or compartment with data and **Edit unit**

Data import

Generate a CSV or XLSX template to import your data

- 1. In the left panel, select at least one monitored parameter
- 2. In the right panel, select at least one **location** (for lab-scale processes, the possible locations are the compartments of the reactor replicates; for industrial and field processes, the possible locations are the processes).
- 3. In the middle panel, select your template type
- 4. Click on "Generate template" below
- 5. Once the template is filled, **upload** it in the middle panel, at the bottom of the page

Data management

Select at least one Monitored Data and one Replicate/Process

Data display

This interface enables the creation of **Graph collections**, which will automatically updated according to newly entered data in the considered **Experimental series** or **Sampling campaign**.



11.2 Meta-omics analysis

Sequencing runs

Each sequencing run gathers biosamples from the project. It can combine biosamples from distinct experimental series, sampling campaigns, and inputs. In DeepOmics, the raw sequencing data and/or the data processed with a bioinformatics pipe-line can be uploaded.

To analyze amplicon sequencing data, we strongly advise to use the <u>pipe-line</u> developed specifically for DeepOmics. Indeed, it is convenient as the obtained outputs are directly compatible with import into DeepOmics. Moreover, this pipe-line is well adapted to microbial variants comparison in a data warehouse framework (relying on DADA2 for the clustering, which produces Amplicon Sequencing Variants, ASVs). Finally, using always the same tool will favor homogeneity across the DeepOmics data warehouse, which aims at promoting meta-analysis. However, if you prefer to use your own pipe-line, it is possible, provided that the data are formatted in the correct way for upload into DeepOmics (see format in section 7).

DeepOmics is intended to store and request meta-omics data analyzed beforehand with bioinformatics pipe-line. It is not oriented towards the visualization and statistical analysis of meta-omics data. For such tasks, it is possible to use <u>Easy16S</u>, a distinct user-friendly web application, freely available on INRAE-MIGALE bioinformatics platform.

Biosamples

The biosamples are created within each experimental series (laboratory data) or sampling campaign (industrial data). In the tab dedicated to meta-omics analysis, all the biosamples from the project are visible (all experimental series and all sampling campaigns); it is however only possible to visualize them, not to modify them.

11.3 Common browse buttons

| Widget | Function |
|--------|---|
| 0 | Tooltip: a message pops up when the mouse pointer hovers over this symbol |
| ľ | To edit the corresponding data |
| Î | To delete the corresponding data |



| 8 | To delete the corresponding item |
|-------------------------|---|
| × | To delete the corresponding item |
| * | Mandatory field |
| • Add a tag | To add a missing element |
| + New series | To enter new data of the indicated type (current example: a new series) |
| + New biosample | To enter a new biosample (individually, by filling a form on the interface) |
| + Import biosamples | To import biosamples (batch mode, using the xlsx template) |
| 🛓 Export biosamples 🛛 🗸 | To export biosamples (xlsx file) |
| + Create biomfile | To generate a biomfile from the selected biosample results |
| 3 | To view the project history |
| Save | Save button |



| | Platform for Omics Data |
|-------------------|---|
| × Close | Close button |
| <u>Close</u> | Close button |
| 🖹 _{or} ዾ | Indicates a file which can be downloaded from the interface |
| S | Indicates that the required data are complete |
| 8 | Indicates that some required data are still missing |
| ✓ View details | Expand button |
| ∧ Hide | Collapse button |
| > | Expand button |
| ~ | Collapse button |
| Q Search | Search bar |
| | Check box |



| H 4 1 2 3 > H | Pagination bar |
|---------------|--|
| • • • • | Pagination bar (for Operating conditions) |
| < > | Left and right pagination buttons (for Operating conditions) |

12. 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. 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/).



List of illustrations

| Overview of the sign-in page in Renater federation (step 1) | 7 |
|---|------|
| Overview of a user welcome page | 8 |
| Link to the recommended bioinformatic pipeline (gitlab page) | 9 |
| Overview of the Gitlab page dedicated to deepomics16S bioinformatic pipeline | 9 |
| Link to Easy16S, a userfriendly tool for microbial ecology data analysis | |
| Overview of Easy16S interface (using one of Easy16S demo dataset) | 11 |
| Overview of the code display functionality in Easy16S | 11 |
| Simplified model scheme of DeepOmics concepts | 13 |
| Overview of a user welcome page, with the project table | 15 |
| Overview of a project description | 15 |
| Overview of the tab "Lab-scale process data" | 16 |
| Overview of the tab "Industrial and field data" | 16 |
| Overview of the tab "Meta-omics analysis" | 16 |
| Overview of an input in the interface "Project's input management" | 17 |
| Overview of the interface "Experimental series" | 17 |
| Overview of an experimental series selected in the list of "Experimental series" (Operating condition | ons) |
| | 18 |
| Overview of an experimental series selected in the list of "Experimental series" (lower part: | |
| Replicates) | |
| Overview of an experimental series selected in the list of "Experimental series" (Physico-chemical | |
| qPCR) | |
| Overview of an experimental series selected in the list of "Experimental series" (Biosamples) | |
| Overview of the interface "Sampling sites" | |
| Overview of the interface "Biotechnological processes" | |
| Overview of the interface "Sampling campaigns" | |
| Overview of the "Biosamples" interface, within a "Sampling campaign" | 21 |
| Overview of a sampling campaign the interface "Sampling campaigns" (Physico-chemical and qPC | |
| | |
| Overview of "Biosample results", for the project "00_demo_AD_inhib_Poirier" | |
| Overview of the interface "Sequencing runs" | |
| Overview of a sequencing run selected in the interface "Sequencing runs" | |
| Overview of the interface "Metabarcoding bioinformatic run" | |
| Overview of "Biosample results", including the funciton "Create biomfile" | |
| Overview of the access to "Application tokens" page | |
| Overview of the "Application tokens" page | |
| Schematic view of DeepOmics informatics structure | 30 |



Software

Bionformatic workflow procedure Bioinformatic workflow Metabarcoding bioinformatic run

Memo on DeepOmics functional structure

A recap on DeepOmics structure

Data display

| | Level 1 | Projects | (e.g. 00_demo_AD_inhib_Poirier) | | |
|---------|--|----------|--|---------|-------------------------|
| | Left tab: labscale | Μ | iddle tab: industrial and field | | Right tab: meta-omics |
| | 🛓 Lab-scale process data | | 🖿 Industrial and field process data | | III Meta-omics analysis |
| Level 2 | LAB SCALE PROCESS | Level 2 | INDUSTRIAL AND FIELD PROCESS | Level 2 | META-OMICS ANALYSIS |
| Level 3 | Project's inputs (buffers, inocula, feeding,) | Level 3 | Sampling site (e.g. Site_1) | Level 3 | Bioinformatics |
| Level 3 | Experimental series (e.g. NH4) | Level 3 | Biotechnological processes (e.g. DG_1) | | Biosamples |
| | Operating conditions (e.g. NH4_000500) | Level 3 | Sampling campaigns (e.g. AD_sludge) | | Biosample results |
| | Replicates (e.g. NH4_00500_rep1, NH4_00500_rep2) | | Biosamples (on which meta-omics analysis have | Level 3 | Sequencing runs |
| | Biosamples (on which meta-omics analysis have | | biosamples (on which inecarbings analysis have been/ will be performed) | _ | Procedures |
| | been/ will be performed) | | Physico-chemical and qPCR data (e.g. CH4, VFAs,) | | PCR conditions |
| | Physico-chemical and qPCR data (e.g: CH4, VFAs,) | | Data configuration | | Relevant url |
| | Data configuration | | Data import | | Sequencing runs |
| | Data Import | | Data management | Level 3 | |
| | Data management | | Data display | Sever 5 | Bioinformatic analyses |



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

Main steps of the procedure

Project, reactors and processes

Summary table

| | Lab-scale process data | Industrial and field process d | | |
|--|--|--|--|--|
| Design | Inputs (denters, boxes, company) Experimental series B Operating conditions Replicates | Sampling sites Biotechnological processes Sampling campaigns | | |
| Physico-chemical parameters | Dat: Data m | enfiguration a import 🏦 im anagement 🛔 im a display 📕 im | | |
| Meta-omics analysis (amplican sequenting data) | Bior Sequencing runs (procedure, PCR o Metabarcoding bioinformatic ru procedure, bromonatic workflow) | | | |
| Biosample results | View by | biosample 👃 htm | | |

| CREATE A PROJECT |
|--|
| 1. Make sure you are on the project page (click on Home or |
| Projects in the uppest menu bar) |
| Create a new Project (private or public) (e.g. |
| 00_demo_AD_inhib_Poirier or 00_demo_AD_plants) |
| 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 |
| 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 |
| Make sure you are in the up-middle panel: |
| Industrial and field process data |
| 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

- 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)
 Click on Data configuration (left menu bar)
- Select on the left the parameters corresponding to your experiment, set their unit, and validate)
 Click on Data import (left menu bar)
- 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

 Fill in the xlsx template, for instance in Excel application (external to DeepOmics)
 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

- 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)
 Click on Samples in the left menu bar
- Click on Import samples (top right) and subsequently on Download template (xisx)
- 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
- 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

 Make sure you are in the right panel (Meta-omics analysis)

- 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.
- 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

- 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).
- 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.
 Towards the bottom of the left menu bar. click on
 - Towards the bottom of the left menu bar, click o Sequencing sample metadata
- Export the xisx template, fill it in (e.g. in Excel application, external to DeepOmics), and import the filled-in template on the same DeepOmics page.
- Click on Import fastq (towards the bottom of the left menu bar)
- Export, fill-in and import the xisx template, similar to above.
- 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



A recap on the different xlsx templates

- 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**

