

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>!

Cimics	-XNSII- Com
	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 Julent Apathol Anamenat	
Bienvenue sur le portail d'authentification aux applications INRAE	Support Applications
Veuillez vous authentifier	
Identifiant Mot de passe Voir mes dernières connexions Se connecter	
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.

mics	1	P. IX	Anushi to 21 gives	-
🖷 Hame 🗣 Projects 👘 Anda analysh	25 Ad	ninistration •	& DADAZ	Ø Kasytsk
Projects				
Q last,				+ Avegaged
Project +	Status #	Coordinator g	Detail #	
100 denne AD Inhilh Poletier Bonne penjiett. 4D Inhilhitton diata, Ginam Poletier		LONIO CHIPLEUR	50xtt: 2013-10-01 Exil: 2010-16-31 Contact person: Officer OrdPLEDR 2 separtiveental series 4.1 Sampling campaigns	• Show
400 00_drms_AD_plants		A HIGHE BUT DWEIN: DE CISTERN	Start: 2023-01-23 - Grid Contact present: MININE BIZE DIPUTZAC DE CASTERN El reperimental series & 2 Sempling companyon	• Suw

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.

Deep Omics	P		RIAL	Anuser to 20 powerance of Castrant
# Home	Projects	in these analysis	💁 Administration 👻	I DADA2 I Easylés
Projects				
Q brech.				+ Monepropert.
	Link	k to the recom	nmended bioinformatic pipeli	ine (gitlab 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

entry formats) and the mapping of covariates.

et For	nomia fête ses 6 ans aulourribui l	O · com
Cedric M	doux > Easy16S	
E	Easy16S	☆ Ajouter aux favoris
-0- 90 v	alidations 🚏 7 branches 🧷 4 étiquettes 🗔 7,6 Mo stockage de projet	
	more details for summaryTransfo	[rayment]

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.

mas //	N	RXX	A ASIM PLET DAVEZ	AC DE CASTERN -
🕷 Nome 🗣 Projecta 👘 Mada-analysi	ab Adr	ministration •	& DADA2	8 Easy165
Projects				
Q. Seath_				+ les proprié
Project +	Status O	Coordinator g	Detail e	
100 deine AD_inhib Poirier Deres project 40 Febblier data, lienes Poirier	-	LOIvie CHAPLEON	Start: 1913-19-01 (Frd. 2018-19-31) Contact parameter: Obvior CHARLESH 2 experimental series & 1 Sampling campaigns	· United
400 00 _denso_AD_pleats to dense_AD_pleats		Annal buy traiting the castern	Barts 2021-05-03 End Cartact person: Altitude BUE DWW(224) DK CAST59A O separativential aurios & 2 Sartaling Languigns	• Since

Overview of a user welcome page, with the project table

Project * 00_dermo_AD_inhib_Point Name: Dermo project_AD inhib_Pointer Code: 00_derma_AD_ANB_Pointer Same: Oblee CHMPLEM Cartast pesse: Oblee CHMPLEM Tag: AMBRECK, andorotic diggetive, releasement, tailet, phenol, annonever, tienwate: Oblee CHMPLEM Cartast pesse: Oblee CHMPLEM Tag: AMBRECK, andorotic diggetive, releasement, tailet, phenol, annonever, tienwate: USE USE Same: Same: TOURTER, Angelase GUERNINE, CHRPSTELLE BURGAUL, threoder annonever, tienwate: USE USE Same: Jamin Same: TOURTER, Angelase GUERNINE, CHRPSTELLE BURGAUL, threoder BOUCHEZ, Analade BURE DAREECK DE CHETANA, Oblee CHAPLEM Same: Jamin Same: TOURTER, Angelase GUERNINE, CHRPSTELLE BURGAUL, threoder BOUCHEZ, Analade BURE DAREECK DE CHETANA, Oblee CHAPLEM Same: Same: Tourter and and on the Samon Fasile's SPHQ sock or annoecdu digeston inhibitation try project was supervised by US Oblee Chaplewe. Project: Same: Same: Samon Fasile's SPHQ sock or annoecdu digeston inhibitation try project was supervised by US Oblee Chaplewe. Project: Same: Samon Fasile's SPHQ sock or annoecdu digeston inhibitation try project was supervised by US Oblee Chaplewe. Project: Same: Samon Fasile's SPHQ sock or annoecdu digeston inhibitation try project was supervised by US Oblee Chaplewe.		HØr Projects III. Mela-enelysis	20 Administration +		& DADA3	& Lay11
Name: Deriv proper: AD induction data, Since Pointer Code: out, dema, AD, Adda, Fability Canada Huano Obler: CHAPLERI Canada Huano Obler: CHAPLERI Tags: Ath Edits, assophytic digetility, nicescentri, balch, phenol, summorium, Teiswatte Santa Huano Status Satta Fability Satta Fability Tags: Jatta Fability Satta Fability Satta Fability Satta Fability Satta Fability Patha Hall Satta Fability Satta Fability Satta Fability Jatta Fability Satta Fability Patha Hall Satta Fability Satta Fability Patha Jatta Fability Satta Fability Patha Satta Fability<	oject * 00	_demo_AD_inhib_Poiner				2 8
Same Triategy and Same Triategy	Dertio	propert: Ad inhibition stata, Simon Pointer	Code:	00, Jenu, A0, Johb, Pokier		
age: Antibilition, assentation of digretion, releasement, balance, phenol, permovanem, biowante hart dete: 3EES 10-01 hart dete: 3EES 10-01 hart dete: pathin pathin: pathin: hart dete: Sever PERMENT, Angelow GUENNE, CHRYSTELLE BUREAU, THEODORE BOUCHEZ, ANNAEZ BURE DAVEEAU DE CASTEAN, Obivier CHRYLTON hart dete: Sever PERMENT, Angelow GUENNE, CHRYSTELLE BUREAU, THEODORE BOUCHEZ, ANNAEZ BURE DAVEEAU DE CASTEAN, Obivier CHRYLTON hart downer: The project contains dota from the Sinion Parisier's PHO book on on severable digetion inhibition hy preveit or annaesis. This project was supervised by Dr Divier Chapteou opient's website: - opient's website: - opient's website: -	op/in	CHAPLESH	Cantast person:	Of aller CHAPLEUR		
hart defer 3103-10-01 hart defer 3005-10-31 harts paties harts paties harts denor PDRREEK, Angelan GUEINE, CONVENTION BURGHE, HOUCHES, ANNAE BURGHER DREEKE DE CASTENA, Divine CHAPILON harts operation entry harts operation interpreter Dorthering data from OV Simon Pacine's SPID work on orienterfold digestion inhibition by greekel or annexels. This project was supervised by Dr Divine Chapters. hight's reformed - hight's reformed - hight's reformed - hight's reformed -	100.00	Son, anotrobic digestion, microcosts, balch, phenol, ammoni	um, Teowaste			
Millionia 2005/06/301 Materia patie Materia Second Possible, Angelace GUENNE, CHRYSTELLE BUREAU, THEODORE BOUCHEZ, ANARE BUE DAVEENE DE CASTENA, Onlive CHAPTERN Materia introductionia Materia	1113	19-91				
hata: patile wridgamba: Second POINTER: A registive GUENHIE, CHRYSTELLE BUREAU, THEODORE INCUDIET, ANNALE BUE UNKEEN: DE CANTERN, ONlive CHRYSTELLE tamore: Project contains dool from Or Simon Pointe's Philo acck on an secolds digestion infibition by greated or annaceis. This project was supervised by Dr Divier Chapters. Theorie data were entended in the data weighters: a door of data. vient's wride: - vient's relicier: -	3004-	10-30.				
harticipamia: Seven PORAPER, Angelian GUENAL, CHRYSTELLE BUREAU, THEODORE BOUCHEZ, ANARE DUE D'ANTEAU DE CASTEAU, DRive CHAPTERN Statement: Project container dota from to Senon Posiele's PHO work on an encloke digestion inhibition by greated or antenenia. This project was supervised by Un Divier Chapters. Project and a wern entracid in the data washing as a domo dotatat. State bounds:	pairie					
Statework: enquity: Autorom: This project nontains dob/from Or Simon Painler's PHO sock or anseroble digestion infibition by great or annousle. This project was supervised by Ur Divise Chapters. Specific restricts whends: - Specific restricts: - Specific restricts: -	since	PORTER, Angeles GUENNE, CHRYSTELLE BUREAU, THEODOR	E BOUCHEZ, ARABE BUT D'AREAC DE CASTENA.	Olivine CHAPLEUR		
Accentage : This project nortalies data term for Senior's PHO work or an exercise digestion infibition by greated or annexeds. This project was supervised by the Divise Chapters. These data were established in the data weerbound as a domo dataset. Seniort's vehicles	: (1915)	C.				
togait's websile : togait's websile : converts witchen every end parents :	This p Those	roject contailes data from Dr Simon Pairles's PhD work on ana data were entered in the data wamphowar as a dono dataset.	rrobic digestion inhibition by preval or armonia.	This project was supervised by Dr Dilvier C	tapleon	
hujetts nikooso - zuoriti: aticles empty est patents:	etste: -					
kineriti: urbidan eri parenta: Man decemente 🗮 erita Tama Banarta V 115 100	rforence -					
Washermann . B. Alle Baraneth Hanning H 14 Mil	ntichei empti tal	ř				
And increased. In State Andread of Links by 30 and	anett: 🖪 🕸	(16, Wales/Resaucts, Priesd pdf, 12,9 Mb)				
2014 http://www.mia.put/(4.3.00)	B +	(166-64) fragmenta.pdf (4-3-86)				

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.



			Admisistration *			& DADA2	∂ Eaty16
Project "00_	démo_AD_inhib_F	Poirier * 👻 view dotats		1 .7			3.00 M T
Dias	icale process data		In Instantrial and Seld process da	na -		III Meta-emito a	untime
Froject's input: management	上 Experi	mental series					
ProjectSinput allagement Stopenmantat series	L Experi	mental series					+ Americania
Project's input an agement g to journant a soline	L Experi	mental series	Custost person	Start date		Operaturs	+ Margaret
Projects input anagement Comments series	L Experi	mental series Soperivental seles D 7014 Assession totalities	Custact persons	Start date 2014-03-03	Since PORCE	Operators r, crementatione existence existence existence existence existence existence existence existence existence exist	+ Meetineen

Overview of the tab "Lab-scale process data"

R Home	@ Projects	12 Hote and pti	2: Administration -		& DADA2	Ø Easy11
Project "00_t	Jemo, AD_inhib,	Poirier " 👻 Vew datata				
						No. 10
August of						
deatriel Langing step	/ Samp	ling campaigns o				
deatried L'amplièng stres disolucificatiogical constan	A Samp	ling campaigns o				+ New Campaign
deatrial Gangling step distact real optical scenare Gangling campaigns	A Samp	ling campaigns 🛛	Contact person	Start data	Operators	+ new caregoige
deatrial • Sampling step . Universited optical objection • Sampling campaigne	A Samp	fing campaigns © fans fans fans fans fans fans fans fans	Contact person	Start date 2004-00-05	Operators Stram PO0E5	+ New Constantion

Overview of the tab "Industrial and field data"

oop om gs		-			Autority II	LE RIWEZAC DE CASTERS -
# Home	IP Projects	ileto acorym 🔰 Zo Admin	istration *		& DADA3	₿ Kasy165
Project "00_,	demo_AD_inhib_Pairier *	🛩 View dataila				
<u>h</u> tabə	cale process data	-	Sa Industrial and field process data			emics analysts
Bolinthematics A theorematics III Electropic results Sequencing runs	Sequencing rul	15 0				+ new sequencing
Procedure	Ratione	Sequencer	Status	Date	Uter	Procedure
P PCR conditions 9 Reliceant att	X NH4 45 samples	NAME PROSE POR	Related	2015-01-25	CHRYSTELLE BUNEAU	AH4_publication
E Generating name Bolenformatik Analyses	X Phenol	INFINE PROSE PORT	finitial	2010-05-00	CORSTELLE BUREAU	Photol, gubfication
A Software Bioinformatic workflow			× 1	8		

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*.

Project *00, ال	, demo_AD_inhib_Poirier ™ → Wee intails	Sa Industrial and Skild process data	III Meto-omics analysis
derestary Project Lagan angelent Depertmental water	 Project's input management @ Amilable input Ami	Silewate/food And Pood Processing Hants/large Producers reason reason company reason Company reason Citizer the company's name City Chemandia Codisciser state Uniter the collection state	a, Industrial : Biosemple: 0 + Ass? Company settlety Country Traves
Project "00_	demo_AD_Inhib_Poirier * < successful at	the interface "Project's input mo	anagement"

SIMULT FORER, CHRYSTELLE BLEEAU, ODWIN CRAFLEUR

SINNE FORER, CHRYSTELLE BUREAU, ORMIN CHAPLEUM

Chiney Charlenge

A CRAW CHARLEN

2014-00-03

2014-02-24

L NH4

Prepai





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.

Jmics ///		X		ARMANE BIZE DAWE ZAC DE CASTELIN
# Home 🕪 I	Projects III III III III III IIII IIII IIIII	2e Administration +		Ø DADA2 Ø Easy181
Project "00_demo	_AD_inhib_Poirier * 🐱 view det	when		
La La de a service prove	erss data	in instructurial and field process of	fata	💷 Meto versios analysis
Laboratory		~		
Prejectsinget	Experimental series 705	** Yew details		
a formation and a second se				
22 Cignit Harrison Sol Vis.	A Biosamples o		A Sportfungerplan S	+ Inspect Discoveration + these Discoverages
And a series and as a series	A Biosamples O	Sampling location	Lagent frammergation - N	+ Hepert Desamption + How Desampte
Operating constitutes	Biosamples	Sampling location	Date 2014-10-21.1200	Perceptionangen Perception Description Som off pelot
A chemical series	Biosamples Cede Sum, owa, dwass European Sumples	Sampling location NH4, 2000 + NH4, 2000, H442, 1 NH4, 2000 - NH4, 2000, H442, 1	Dete 2014-00-21 (1200) 2014-00-21 (1200) 2014-05-20 (1000)	Propertitionamples Proceedings Description Som off pelot Nom off pelot
La Colonia francé de la colonia de la coloni	Biosamples Cede S_20,0002,divens E_20,0002,divens S_20,0002,divens	Sampling location 7844, 2000 + 1844, 2000, 1484, 1 1844, 2000 + 1844, 2000, 1492, 1 1844, 2000 + 1844, 2000, 1492, 1	Date 2014-00-21.1200 2014-05-20.1000 2014-05-20.1000	Propertitionamples Proceedings Description Som off public Nom off public Som off public
All Constant of the States of	Biosamples Cede S_01,0002,dayata S_01,0002,dayata S_01,0002,dayata S_01,0002,dayata S_01,0002,dayata	Sampling location 7844, 25000 + 1844, 25000, 1442, 1 1844, 25200 + 1844, 25000, 1442, 1 1844, 352000 + 1844, 25000, 1442, 1 1844, 352000 + 1844, 25000, 1442, 1	Dete Dete 2014-00-21.1200 2014-05-20.1000 2014-05-20.1000 2014-05-20.1000 2014-05-20.1000	Competitionample Competition Competition
A Constant of Action A Constant of Action	Biosamples Cede Sum onka dwate	Sampling location NH41_22000 + NH4_22000, Hull_1 NH44_22000 + NH4_20000, Hull_1 NH44_202000 + NH4_20000, Hull_1 NH4_202000 + NH4_20000, Hull_1	Dete 2894.496-20.10000 2894.496-20.10000 2894.496-20.10000 2894.496-20.10000 2894.496-20.10000 2894.496-20.10000 2894.496-20.10000 2894.496-20.10000 2894.496-20.10000	Constraints
Appendimental and the Anne	Biosamples Cede Sum, cond, doyate Sum, cond, doyate	Sampling location NH41_25000 + NH41_25000_H0[2,1] NH41_25000 + NH41_25000_H0[2,1] NH41_05000 + NH41_25000_H0[2,1] NH41_00000 + NH41_20000_H0[2,1] NH41_20000 + NH41_20000_H0[2,1]	Dete 2894.496-20.10000 2894.496-20.10000 2894.496-20.10000 2894.496-20.10000 2894.496-20.10000 2894.496-20.10000 2894.496-20.10000 2894.496-20.10000 2894.496-20.10000 2894.496-20.10000	

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.				+ Newski
Pearling cantrellar	Reme	Operators	Syndicate	Location	Processe
	las Site_t Hite_t			Oty: Buginty Country Osottlinstes	(8)
	he Site ,2 title,2			Oty. Beginn Country Country	(8)
	In Site 3			City: Register Grantity:	(a)

Overview of the interface "Sampling sites

Liab-scale	e process data	ha tookuttral and field process data	III Neta-ornics analysis
industrial			
in Sampting sites	Biotechnological processes		
Etrebultevological	a territ.		+ New Trans
A Sampling campaigns	Name	Site	Category
	▲ DG_36 00_36	9 _{8.} 1	treatment/atmendels/dignation/yest/catt
	▲ DG_6 06,8	Sec.4	tosatment/accounters/digestion/dry/plag flow anaerics's alignetics
	A DG 2	5hv,1	but/nert/annentbk/digesSco/web/bat/b

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>h</u> lab-sea	le process data		La technicital and field process data			(1) Heta-ornica analysis	
minutrial	A Cample	n campaine -* 40. aludoa *					121
a larging star	- sanipun	Reambailin sin7amoñe					-
Biotechnological processes	Name: Tags:	KD_studge studge, aniambic digestion	Code	AL), stickpr		
Stampling campaight	Ebried prom:	ANNAL BREITWEZACTIE CASTERN.	most d	Mar. 20	121-01-10		
	Openators :	AROME RIZE O'WEZAC DE CASTERA	30mm	52	v_4,5Hc_1,50c_2,5Hc_1,5Hc	5,589.0	
D_shullge	Dummary I						
theory in the	Sampling schemes	analyze .					
hysics-chemical and oPCR							
Descarfingation :	Procedurat	enutly.					
	Nour thata:	went					
2 Maranepen	Other documents	example.					

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*).

1. i.ab-acala peo	sees data	in reductrial and field process	he reductrial and field process data		
utantrial	Sampling campaign "A0	siudge venterals			
 Sompling sites 	Countries countries Bit	and a second second			
I thatanhangkagatal ricenses	Biosamples Ø		L Depart Managine	+ import forwards + free to	
A Complete campoigns	Code .	Easter Institute	Onte	Encircleding.	
	Look	internal personal		fauri durres.	
D_statge	DREA.3	Ste.8>00.605.8	3421-65-05 10(57	anaevolik duitge	
Ludadge	DEAS	Stells > DL 805_8	3421-09-06 (0):57	anaerotik dulige	
D_stadge	04.4.5	SHL 8 > 011, 605, 8 SHL 8 > 05, 305, 8 SHL 8 > 05, 305, 8 SHL 8 > 05, 400, 4	3621-08-05 10:57 3021-08-05 10:57 3021-04-05 12:00	anaestik duige anaestik klutge anaestik klutge	
D_sheetge Transmission trystco-chemical and gPCR Bitsto configuration	06.4.5 06.4.5 06.4.5 06.4.5	Ste. 6 > 00. 605. 6 Ste. 5 > 00. 505. 6 Ste. 5 > 05. 305. 6 Ste. 4 > 05. 405. 4 Ste. 3 > 105. 305. 30	2421-04-05 10:57 2422-04-27 12:00 2422-04-27 12:00 2422-04-27 12:00	anaestiis duitje anaestiis duitje eneestiis duitje anaestiis duitje	
stadge rescentions rescentions latinization latinization	000.4.3 000.4.3 000.4.3 000.4.3 000.00.5	Ste. 8 > 00. 505. 8 Ste. 9 > 00. 505. 8 Ste. 9 > 05. 305. 8 Ste. 8 > 05. 405. 4 Ste. 3 > 105. 3405. 34 Ste. 3 > 105. 3405. 34	2021-08-09 (0):57 2020-08-27 (2):00 2021-08-29 (2):00 2021-08-16 (2):00 2021-08-16 (2):00	anaestik dailje anaestik dailje eterolik dailje ataestik dailje	
B_cheatege (Free resolution hysico chemical and gPCR basic configuration glinitic import glinitic import glinitic import	06.4.5 06.4.5 06.4.5 06.4.5 06.4.5 06.2.5	Str., 8 > 00, 505, 8 Str., 8 > 00, 505, 8 Str., 8 = 05, 305, 8 Str., 8 = 05, 405, 4 Str., 3 > 05, 3405, 38 Str., 3 > 05, 3405, 38 Str., 3 > 05, 3405, 38	2021-05-05 105-77 2023.06-27.0200 2027-04-29.12100 2027-04-29.12100 2027-04-14.12027 2027-04-14.12020 2027-04-14.12020	anaechic dailge anaechic dailge Historic dailge atsechic dailge meachic dailge meachic dailge	

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



hisb-s	cala process data		la todustr	tal and Refel process da	(a))		III Meta-omic	s senségnés
ndustriei	A Campling campa	ian dath chu	ten Maria					
Semplegatas	n Sampung campa	ign anvirant	Age weether	alty				
Botechiological operates	S Monitoring data	management						+ Atlantin
Samplingcamprigm			-	_	_			Deligneering
D_studge	7 illent selected 🗸 🗸	Li Peres selected		gon electus	Marker participants			
Revergies					Precess			
typics-chemical and gPCR		ĉ	ç	ę	ŝ	6	ç.	ę
Data configuration	Monitored Heasters Type	8	20	8	8	B	8	80
Dista Argent		06	90	E.	100	10	00	6
	>	- B		7	¥	2.5	100	
the management					0.0	08	18	0.00
Data display	Antaba (g/L)	1.	10	U #				
E Data display	Avaitate (g/L) Buityvele (g/L)	18	10	0	0	18	18	08
E Date chipley	Acatyles (g/L) Burgveles (g/L) Cepressier (g/L)	18	08	0	08	18	18	08

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.

Li Cal-sca	le process data		in industrial and field process data		III Meta nenica analysia
Existence in the second	Biosa	mple Results			
equecting runs	Q from	5			+ tratition (h
Plot in consideration		2 second reading code	Metabarcoding assessormable nun code	Processing matrice	Mumber of Annotation
Selevantul	2012	A \$ 00 00N2 day000	III NB4_bioinfo	Number of rawseads: 51094,	174
Second and their status	100	Sample 15_00_0000_day600	Workflow-stadiat_Trogs(stationprine);	Number of ASV: 114.	20488
informatic Analyses					
Software	3.5	₫ 5_00_00N2_day000	III NB4_biainfa	Number of new coach: 07003, Post process reads: 22442,	140 Tabel according your
Scinformatic workflow science		Sample 13_00_0000_ptayoon	. Workflow: dads2_brigs (italus;privalst)	Number of Adv. 181.	21442
· Hisknormatic workflow	100	A \$ 00 00N2 day029	III N314 bioinfo	Autobio of new marks servers,	129
Matericantelline		Tampin : 3_00_00003_day020	Workflow - dads2_Frsgs (status private)	Number of 850 228.	26488

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*).

List-m	ale process data	-	in industrial and field process data			III Hets onles analysis	
ninformatica							
Biosamples - Biosample sets #2	A Sequencing run	15 U					
epending runs	Q, Steed)					+ New sequences	
Preceduces :	Runname	Sequencer	Status	Date	Uper	Procedure	
PDI conditione	X NH4	BDM POOST ICH	Factoria -	2010-02-22	CHARGETTICS BUREAU.	Data metilication	
Rolevant set	40 parquire	Press, C MUSE Prise	or a construction of the c	Same	CHARTER BURGER	And Third Street	
Supporting rates	X Phrusi	NAME PARTIE FOR	medieal	2015-01-05	CHRISTELLE BUREAU	Planet publication	
Softward.	An unregides.					1 Manuara 2011	
softenatic weather			(A. 1)	1 A. A.			
oodure							



fileb-sce	in process d	ata.		San (reduct that and	field process data	<		III Meta omica analysis	
Rininformatics		Conversion	n ana a						
dictorrates .	A	Sequencing run	PALINE	where retails					
Il Diosomple results			005						
equancing runs	4	Sequencing sam	ples						
Procedures	1	1 Second							
P PCR conditions	-	and and a second							
🕫 Relevant ur!			Matrix		Targeted some or	Tergeted	Tartet	Avaiaible for	
E Sequencing turns		Code	type	Por Prinners	genantic region	taxoreamic or	whitragments.	bietabarcodleg	Actions
lioinformatic Analyses						functional group		Bioinformatic Ban	
∲ Software		5 00 4052 4-000		Ball CONTRACTORISTING		Arthough		8	-
Bainfornatic warkhow	2	Bangle .	DNA :	STREAD AND A REPORT OF A REPOR	305-1996	Bacteria.	15,191.	. S. THE MARL MAY THE SWAS	
Contraction and the second second		\$_00_00\$1_fay000							-
Bitelaturedea		6 on 1007 August		and successive a state of the second		1444-52			-
iointo matternat	3	S of 10.52 anyons	DNA.	SIN (GRINEAGENGEEDIDGER) .	3451694	Secheta.	M5, 01	1 5. 00. mma_mayoni toria	_ e []
		5,00,00H2,dayses							
		Δ		100000000000000000000000000000000000000					100
schurzh miter	>	S_00_00N2_day029	300	STAR (COCCORANTICAL TRACT) -	225-16966	Hactoria	15,100	1 5 00 0000 megaziten	a 🗌
Pedvaroalimuter		5_00_00H2_day#20						Street and state and state and state	23 M

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.

							De	ep
Project *00_de	uno_AD	jnhib_Poirier" 👻	West details				Digital Environt Platform for Or	nic nental Engine nics Data
Liab-scal	le procest di	sta		he industrial and field proc	ess data		III Neta-omics analy	# 1:
Aninhormatica								
Bioseriptos.	108	Metabarcoding in	formatic runs					
Ellimangle insults equencing runs	0	k boatch					+ sescretatorodin	pelinnellenen
Protections		Code	Sequencing sample	Boisformatic workflow	illion file	Hetrics He	a servicents	action
P PCR conditions 9 Nationant sal 6 Tanaany barrans	3	III NH4_biolafe 1014 febacelor matter amatyris	-	stadic2_bugs (ristusprivate)	B abundanonalion Biter meter	B metricator April: 20000		
kalenhormatik Analysen (h Scitivase	5	III phront_bioinfo Pressel bioloformatic analysis		dadut, bogs (statuspriotis)	aturntanun Silom mitut (Sainemit	B motivation Relations		2
tiokrivensitic workflow.								
C*Dichthermatic enclidion								
17.848 (17.94)								

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*.

1) Laib-acale process data			in Fodustrial and field process state		III Neta omiczatałytie
arteles antelesada antelesada	Q See	mple Results			+ Grute bierefte
Jodures.	\wedge	Sequencing sample code	Netabarcoding bioinformatic run xode	Processing metrics	Humber of Annutation
wertur		L 5_00_09N2_day000 Sample -5_00_0001_day000	III NH4_biomfo Workflow: dada2_Bings (itation;private)	Natifian of our made: 53054, Past process mode: 26685, Auctober of ASV: 124.	174 Total association count : jointe
nmatic Analyses hours elementic workflow ture		A 5_00_08N2_day009 bargin=1_00_0002_day009	III NH4_biolada WalkTios (datat_Wags (datas-private)	Number of new rests: 57822, Post process rods: 57842, Number of ASV: 183	103 Total aerotation count. 23442
interesti wettive		<u>IL 5_00_08N2_day629</u>	illi NB4_bioinfo	Number of row work: 58545, First seconds made 25451	228
abarroubing emotic nan		Sweptic : 8_00_00002_0ay029	WeehTow (#wda2_hogs (datumprivate)	Number of ASE 321	2540
	V,	A 5_00_00N2_day042	III NH4_bioinfo	Number of our math: 116828, Fost process seads, 47128,	390). Totat ensolution count:

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.



	1	2 LQ	A MINHE BUC O'NATAN	C DE CANTENA *
ff Hame 🔰 🗢 Projects 🕴 🖻 Mico analysi	2. ad	ministration +	e PAppa tokens	De tasytë
Projects				_
Q Sect.				+ new project
Project -	4	Ceerdinator #	Getail ±	
100 00_demn_AD_inhib_Poirier Terms project: A5 foldstitus data, Sereen Poirier	-	and the Collection	Start, 2013-01-00. Test: 2000-16-31 Contact person: Oblete CHWUUM 2 experimental series 6-1 Sompling compaigne	0-310W
92000_dems_AD_plants 00.dems_MD_plants	A	Average and down the contern	Start: 2011-01-11 Teve - Contract person: ARMAE REE DOVE[24:THE CATERAL Disperimental series & 2 Sampling campaigns	• Sea

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
T	To delete the corresponding data



3	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
5	To view the project history
B Save	Save button



× Close	Close button	
<u>Close</u>	Close button	
🖿 or 🔀	Indicates a file which can be downloaded from the interface	
>	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 < 1 <mark>2</mark> 3 → H	Pagination bar
• 0 0 0	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	10
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)	18
Overview of an experimental series selected in the list of "Experimental series" (Physico-chemical o	and
qPCR)	19
Overview of an experimental series selected in the list of "Experimental series" (Biosamples)	19
Overview of the interface "Sampling sites"	20
Overview of the interface "Biotechnological processes"	20
Overview of the interface "Sampling campaigns"	21
Overview of the "Biosamples" interface, within a "Sampling campaign"	21
Overview of a sampling campaign the interface "Sampling campaigns" (Physico-chemical and qPC	:R)
	22
Overview of "Biosample results", for the project "00_demo_AD_inhib_Poirier"	22
Overview of the interface "Sequencing runs"	23
Overview of a sequencing run selected in the interface "Sequencing runs"	23
Overview of the interface "Metabarcoding bioinformatic run"	24
Overview of "Biosample results", including the funciton "Create biomfile"	24
Overview of the access to "Application tokens" page	25
Overview of the "Application tokens" page	25
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	M	iddle tab: industrial and field		Right tab: meta-omics
	🖞 Lab-scale process data		la 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		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	sevel 5	bioimormatic 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 (deners, buses, company) Experimental series Operating conditions Replicates	Sampling sites Biotechnological processes Sampling campaigns
Physico-chemical parameters	Data co Dati Data m Data	a import 🔹 💷 a nagement 🔹 💷 a display 🛓 💷
Meta-omics analysis (amplican sequenting data)	Bior Sequencing runs (procedures, PCR of Metabarcoding bioinformatic ru procedure, bromonatic workflow)	samples 🔹 💺 tiv conditions, relevant will IR (software, bioinformatic workflow 主 👼 Robert, methods bev
Biosample results	View by	y biosample 👃 htm

CREATE A PROJECT
 Make sure you are on the project page (click on Home or Projects in the uppest menu har)
2. Create a new Project (private or public) (e.e.
 Create a new Project (private of public) (e.g. domo AD inhib Privite OD domo AD plants)
00_demo_AD_innib_Pomer or 00_demo_AD_plants)
 Click on the created project to reach the project page
CASE 1, LAB WORK, DESCRIBE THE REACTORS
 Example in 00_demo_AD_inhib_Poirler
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)
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.a. number
of compartments, input amounts, temperature, valume,
etc)
 Create Replicates for each Operating condition (e.g.
NHA 00000 ren1 NHA 00000 ren2
NH4_00000_Tep1, NH4_00000_Tep2,)
CASE 2, INDUSTRIAL PLANTS, DESCRIBE THE CAMPAIGNS
 Example in 00_demo_AD_plants
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)
 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 of 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**

