



HAL
open science

Design and production of a plug-in tool facilitating the submission of microbial meta-omics data to an international repository

Baptiste Rousseau, Véronique Jamilloux, Thomas Denecker, Hélène Chiapello,
Ariane Bize

► To cite this version:

Baptiste Rousseau, Véronique Jamilloux, Thomas Denecker, Hélène Chiapello, Ariane Bize. Design and production of a plug-in tool facilitating the submission of microbial meta-omics data to an international repository. JOBIM 2023, Jun 2023, Nancy, France. . <hal-04139284>

HAL Id: hal-04139284

<https://hal.inrae.fr/hal-04139284v1>

Submitted on 23 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Baptiste Rousseau¹
Véronique Jamilloux¹
Thomas Denecker³
Hélène Chiapello^{2,3}
Ariane Bize¹



DeepOmics Submission, a plug-in tool to facilitate the submission of meta-omics data to the ENA

> DeepOmics

- ❖ **Storage** and interrogation of project **data** and **metadata**, from laboratory reactors and pilots or industrial plants
- ❖ **Import** and **export** data in standard formats (biom, fastq, fasta, CSV...) according to the nature of the data
- ❖ **Statistical analysis** and **data visualization** of 16S metabarcoding and metagenomic data thanks to coupling with the user-friendly application Easy16S - genome.jouy.inra.fr/shiny/easy16S/

> Technological choices

Shiny:

- ❖ Easy and fast to develop, suited for a POC
- ❖ Does not need extra knowledge of *DeepOmics* architecture
- ❖ Facilitates the deployment with an INRAE in-house solution, SK8

Metark:

- ❖ Data-brokering requires specific skills and a precise expertise
- ❖ Ensures the quality of the data and metadata submitted to the ENA
- ❖ Guarantees process stability by absorbing potential changes in the submission process

> Context

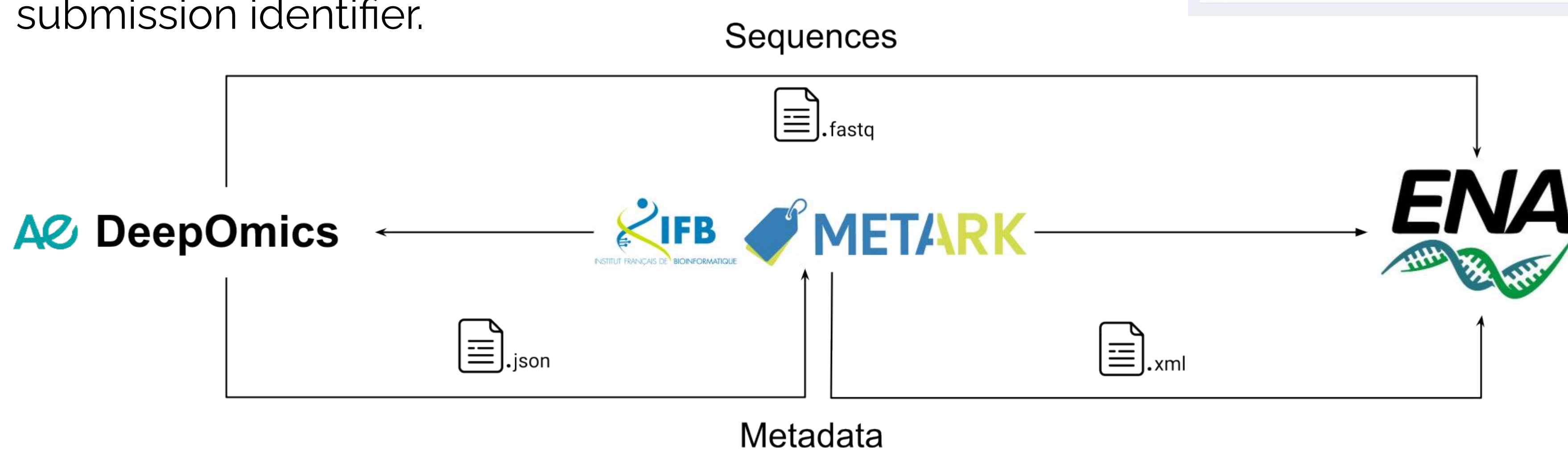
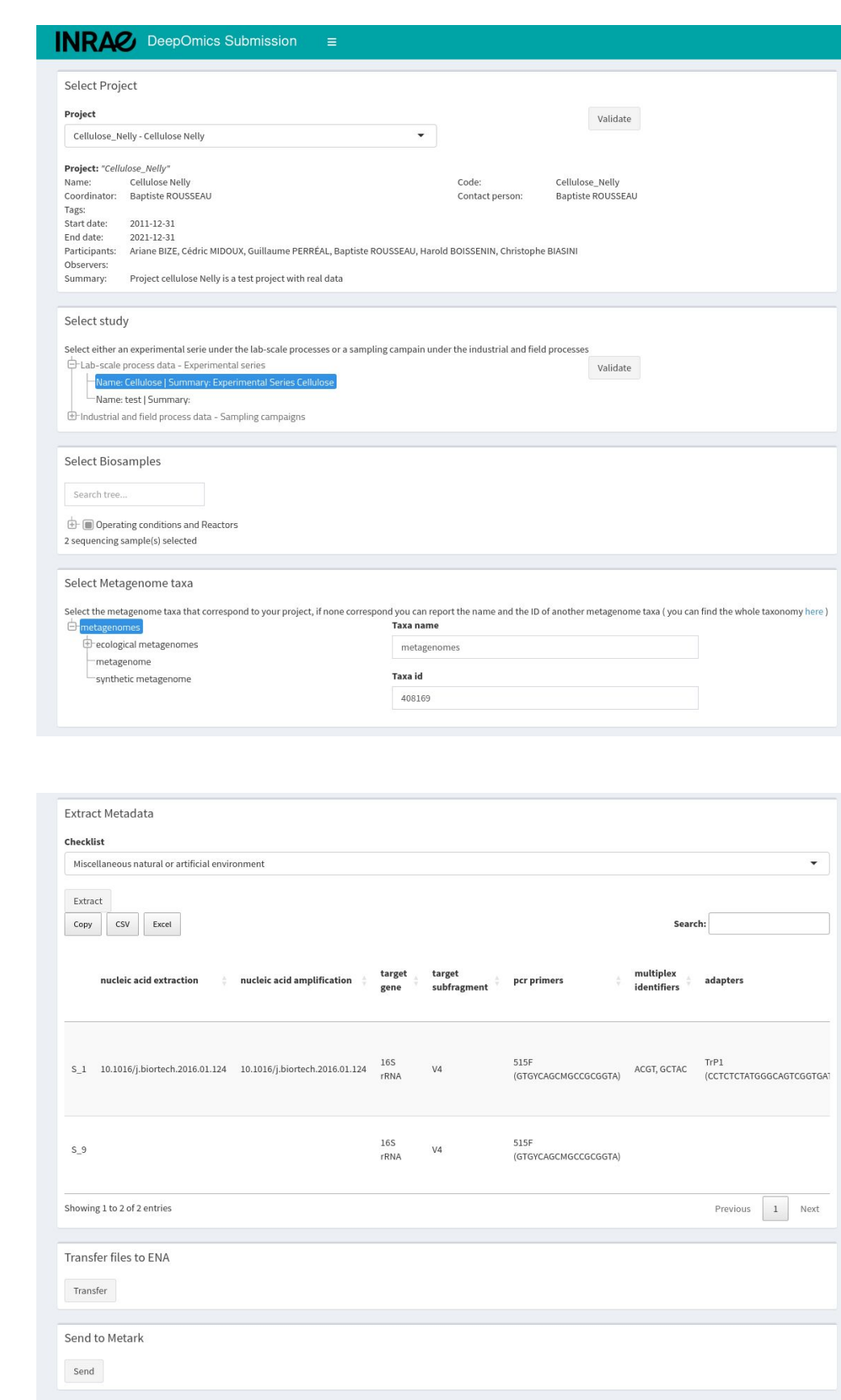
INRAE PROSE has coordinated the development of *DeepOmics*, an information system dedicated to **meta-omics** datasets in the field of **environmental biotechnology**. *DeepOmics* offers the possibility to upload, query and export **16S metabarcoding data** from several environmental biotechnologies, together with many relevant associated metadata, especially regarding operating conditions and process design. To date, *DeepOmics* has all the data and metadata necessary for submission of the datasets to international repositories such as the European Nucleotide Archive (**ENA**), but submission remains a manual and rather complex process.

In collaboration with the French Institute of Bioinformatics (**IFB**), we designed and developed a R-shiny plug-in application, *DeepOmics Submission*, to prepare and automate raw sequencing data submission to the ENA. We will also guarantee the quality of the associated metadata with an IFB data brokering application prototype.

> DeepOmics Submission

A proof-of-concept (POC) application plugging into *DeepOmics* to cover:

- ❖ Mapping of *DeepOmics* concepts onto ENA checklists;
- ❖ Extracting the architecture of a *DeepOmics* project;
- ❖ Enabling the selection of biological samples;
- ❖ Extracting and transmitting to ENA the raw sequence files for these samples;
- ❖ Extracting and transmitting to *Metark* the metadata for these samples;
- ❖ Managing errors and comments from ENA;
- ❖ Finalizing the submission and retrieving the ENA submission identifier.



> Development best practices

Versioning

[git](https://forgemia.inra.fr/prose/deepomics-submission) → <https://forgemia.inra.fr/prose/deepomics-submission>

Virtual environments and dependency



Development tools



> Conclusion

Most objectives of the **POC** have been **completed**. *DeepOmics* is currently in the pre-production phase and is now entering **production**. *DeepOmics Submission* works with the development version of *DeepOmics* and will thus need some modifications to work with the production version, especially regarding user account management. *DeepOmics Submission* will also need to be **deployed**, by relying on **SK8** at this stage. The realization of this POC raised many **conceptual** and **technological challenges**, from **concept mapping** to developing a **headless** and **stateless application**. All in all, it gives rise to **optimism** regarding the prospects of **FAIR data** and contributing to **Open Science**.



Center
Jouy-en-Josas - Antony

¹ Université Paris-Saclay, INRAE, PROSE, 92761 Antony, France

² Université Paris-Saclay, INRAE, MalAGE, 78350 Jouy-en-Josas, France

³ CNRS, Institut Français de Bioinformatique, IFB-core, UMS 3601, Évry, France

ariane.bize@inrae.fr

thomas.denecker@france-bioinformatique.fr