

Atelier Technique#17

Données bioinformatiques de diversité

11-13 mars 2024

Station Biologique de Roscoff

Présentation des plateformes ABIMS et SEBIMER

Erwan Corre, Patrick Guido Durand

Outils

Recommandations, exemples d'utilisations,
gestion de métadonnées, publications dans des
entrepôts, visualisation, moteurs de workflows,
génération de rapports d'analyse





CNRS • SORBONNE UNIVERSITÉ
Station Biologique
de Roscoff



Biogenouest



INSTITUT FRANÇAIS DE BIOINFORMATIQUE



ABIM⁴S

Analysis and Bioinformatics for Marine Science

Ateliers techniques ODATIS/PNDB

11-13 03 2023



Activities



Computing and storage Infrastructure

Bio-analysis

E-infrastructure

Software Engineering

Data Management



Expertises

Support

Training

User Communities

1100 Accounts but 400 active users



CNRS • SORBONNE UNIVERSITÉ
Station Biologique de Roscoff

50%



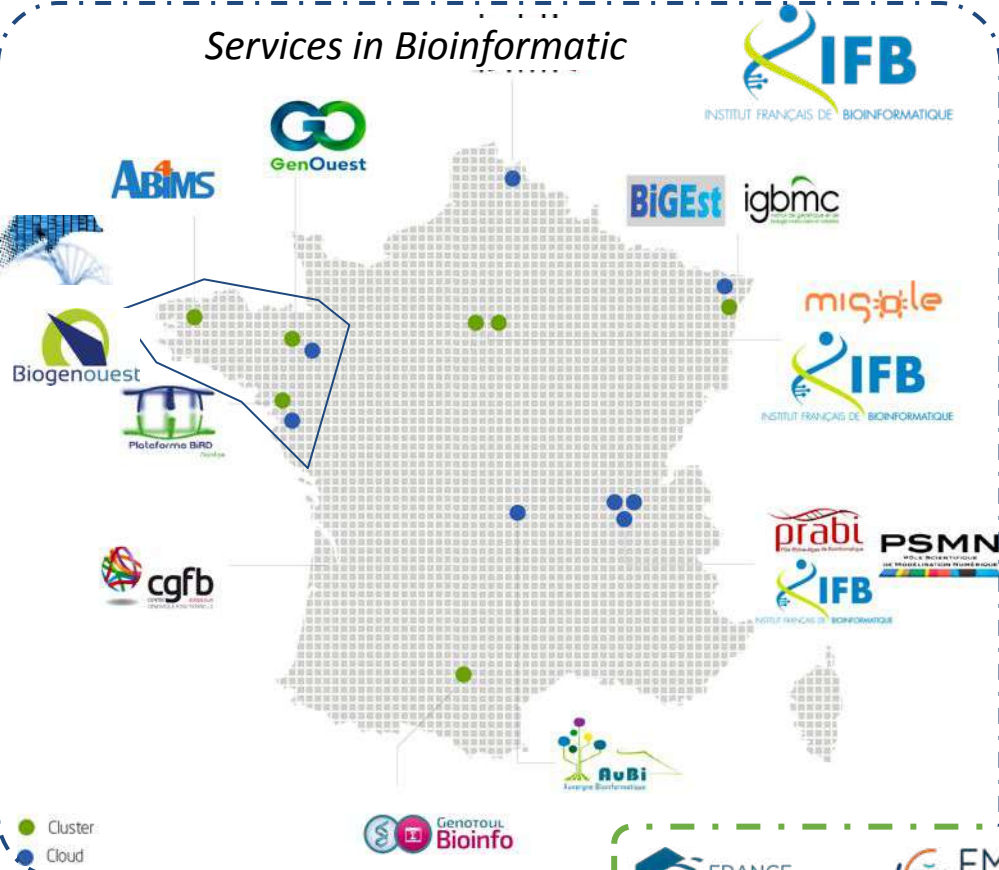
50%



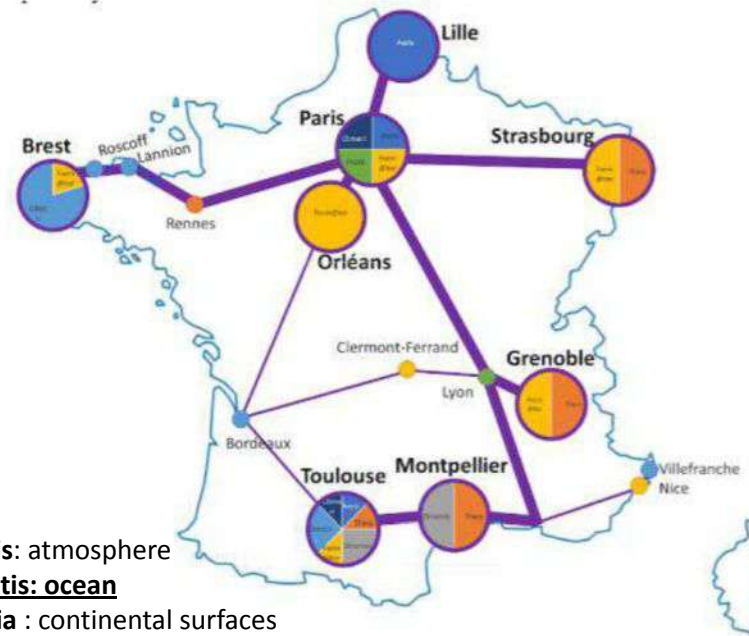
Between 12 and 20 (2024) peoples : 8 permanents = 4.5 FTP PF

National infrastructures integration

Services in Bioinformatics



Earth observation data



aeris: atmosphere

odatis: ocean

theia : continental surfaces

formater : solid earth

G. Le Corguillé : comité de direction



M. Hoebeke: Chargé de Mission ODATIS.

6 PIA-PEPR-PPR as scientific or work package leaders

- **MUDIS4LS**: Mutualised Digital Spaces for FAIR Life Sciences ([infrastructure](#) - [e-infrastructure](#) - [marines data management](#))
- **AO-EMBRC** : Augmented-Observatories of EMBRC ([e-infrastructure](#) - [marines data analysis](#))
- The **EMERGEN** consortium : conducting genomic surveillance of SARS-CoV-2 variants in France. ([e-infrastructure](#))
- **ABROMICS** : a digital platform on antimicrobial resistance to store, integrate, analyze and share multi-omics data ([e-infrastructure](#))
- PPR : **FUTURE-OBS** : Augmented Observatory for Coastal Socio-ecosystems ([marines data management](#))
- PEPR : **ATLASEA** : Atlas des génomes marins : des données massives à l'innovation. ([infrastructure](#) - [e-infrastructure](#) - [genomic data management / analysis](#))

+ EMBRC-Fr, IdeaLg, Oceanomics , IFB/RENABI parts 1,2

9 ANR projects as partners

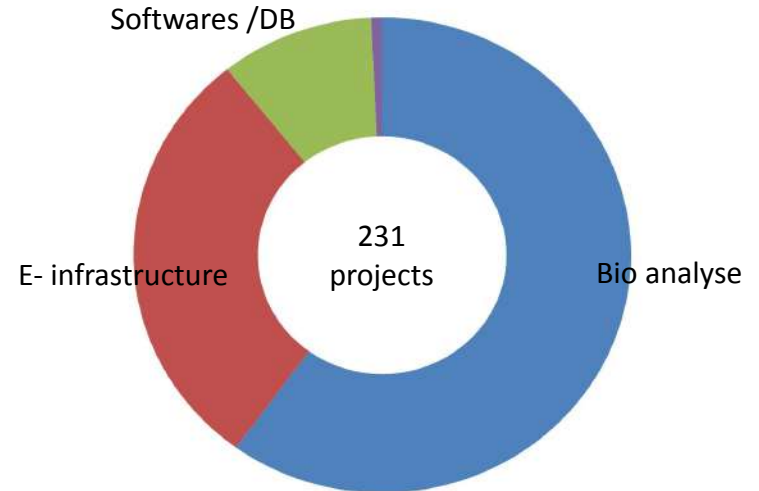
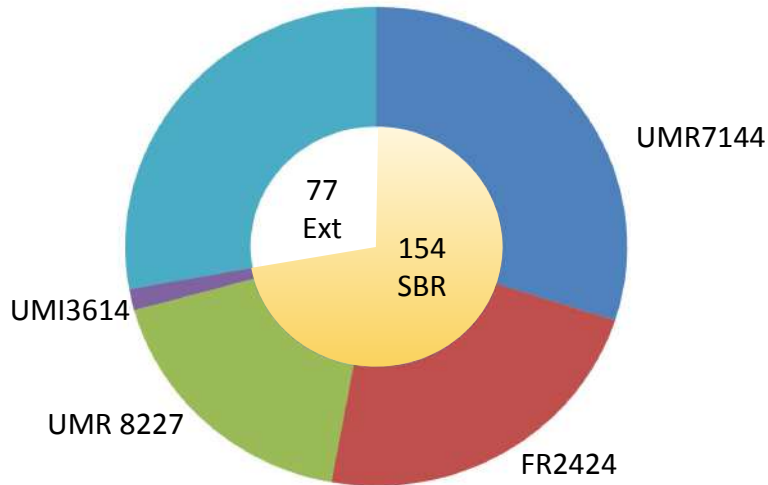
2 National in-situ observation projects

- SNO PHYTOBS
- SNO BENTHOS

European projects

- EOSC Life (2019-2023)
([Workflow management](#))
- EMBRIC H2020 (2015/2019)
([Observation data management](#))
- ELIXIR
([Marine metagenomics](#), [Diversity](#), [Galaxy](#))
- ESFRI EMBRC-Europe project
([e-infrastructure](#))

	Collaboratives projects	Publications	Acknowledgment	Trainees thematic schools
2017-2022	231	84	101+	492 +480
2023	49 (on going)	17	12	63 +100



Metabarcoding analysis

- SOMLIT ASTAN time series
- TARA Ocean analyses
- Tonga volcanic arc
- Several holobionts analysis

Metagenome analysis

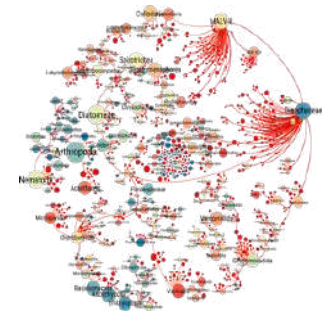
- Metagenome Virus. A. C. Baudoux (Roscoff)
- Metagenome viral Oysters. JM. Escoubas (Montpellier)
- Metagenome Brown algae holobiont
- Venise Laguna

Diversity / population analysis:

- Larvomix. AquaNIS F. Viard (Roscoff)
- RAD seq : Jaera, Ectocarpus T. Broquet (Roscoff)
- Fishes Coastal Rocky environment

Reference database

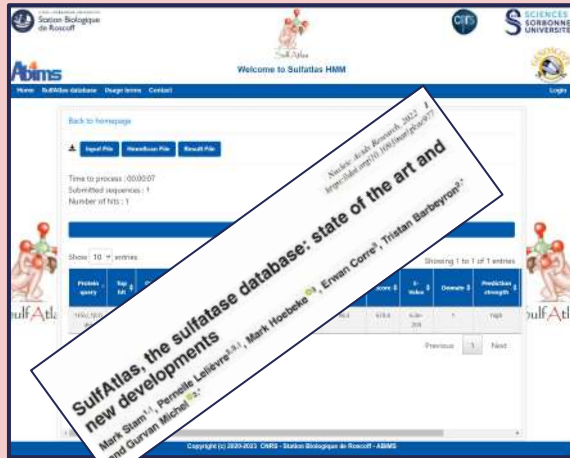
- Marine Micro Eukaryotic transcriptomes reference database (Roscoff- Genoscope-EBI)
- Eukbank: Global eukaryotic diversity (Roscoff- Tara consortium)
- Coastal Invasive species



Software Engineering: panel of activities (requirement collection, specification, development, testing, deployment) producing software applications in response to specific needs for which no shrink-wrapped solutions are available.

Primarily : **Web applications** exposing the contents of **databases** (visualise / search / export / curate).

Sequence-oriented



SulfAtlas, the sulfatase database: state of the art and new developments
 Mark Stam, Pernelle Lalleve, Mark Hoebeke, Erwan Corre, Tristan Barbeyron

SulfAtlas HMM

Pernelle + UMR8227 + CEA

<https://sulfatlas.sb-roscoff.fr/sulfatlashmm>

Hybrid



MyGOD

Charlotte A. + UMR7144 + BGO + MIO

<https://test.sb-roscoff.fr/mygod>

Biodiversity-oriented



Benthobs-B

Jukka + FR2424/SO + SNO BenthObs

<https://benthobsb.sb-roscoff.fr>

Local

Observatoire Astan

Gen4Bio

National

FUTURE OBS

SNO (s)

AO EMBRC

IS2 MUDIS4LS

European

WGEi EMBRC Europe

EMO-BON

MARCO BOLO

Consolidation of issues and practices
Capitalising on skills through pooled recruitment

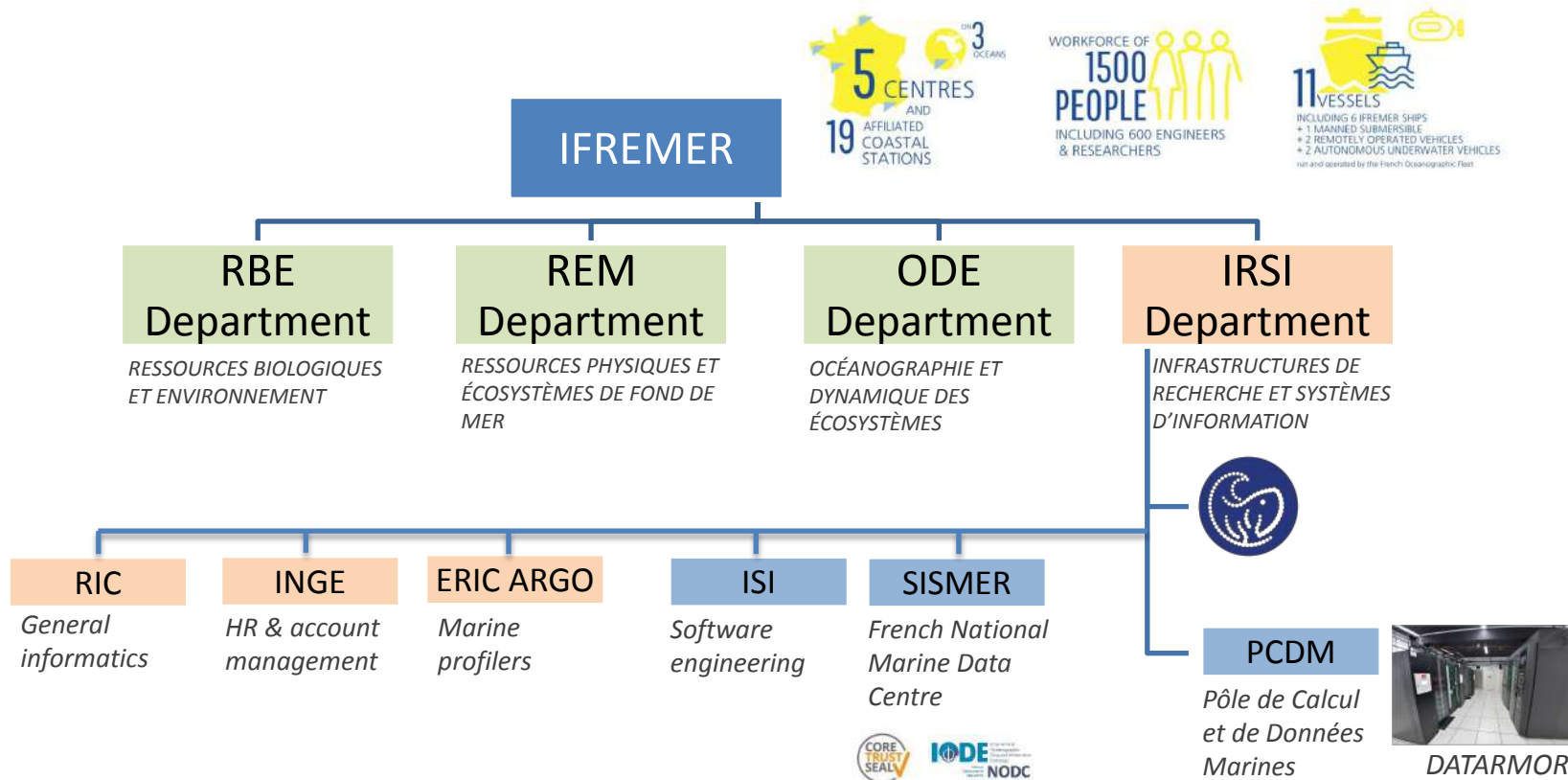


SeBiMER

Bioinformatique marine



Ifremer organization: 4 research departments



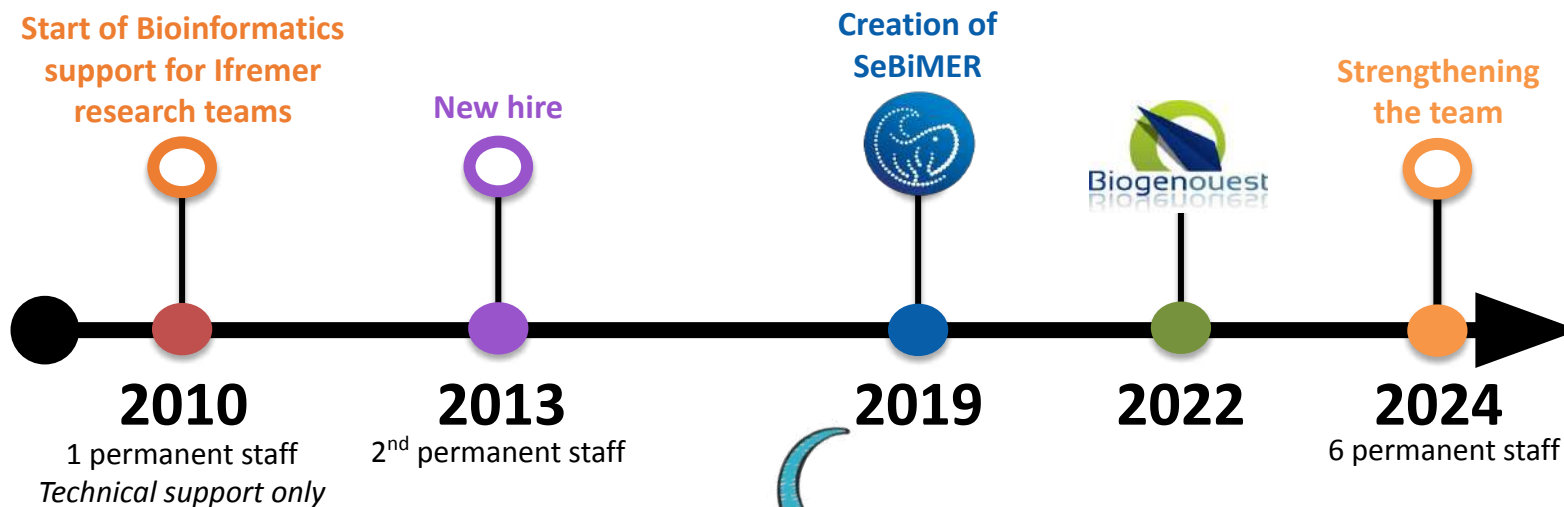
Location



SeBiMER
Bioinformatique marine

Ifremer, Centre Bretagne, Plouzané

A very brief history of SeBiMER



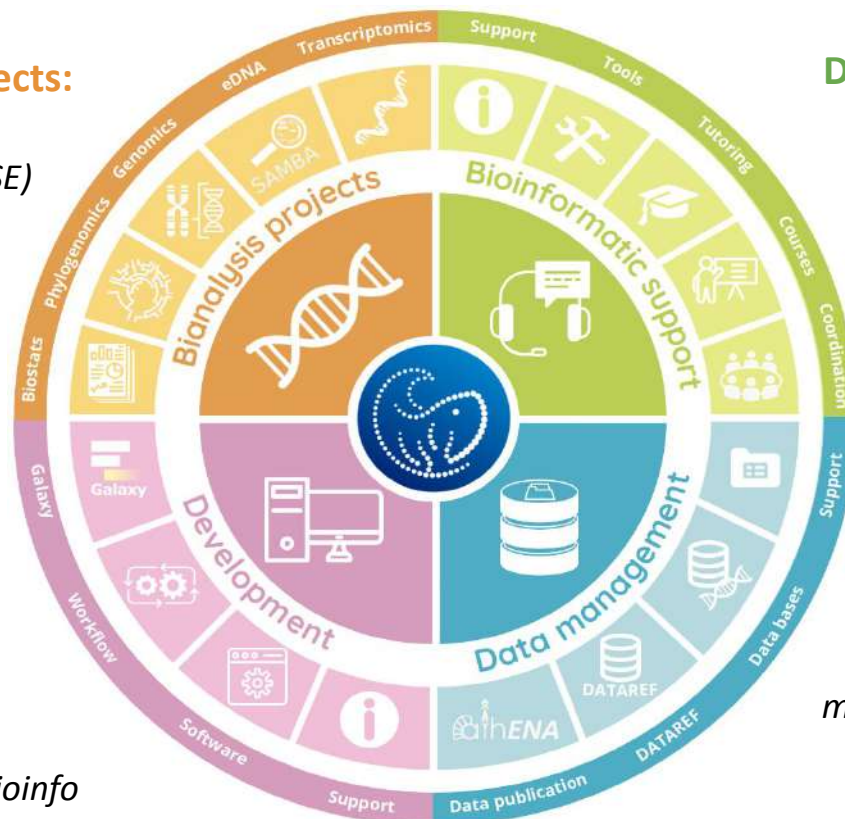
Original missions:

- Pooling the bioinformatics activities of Ifremer
- Active collaboration with research teams
- Focus on transversal bioinformatics activities

SeBiMER missions

Partners of research projects:

- 60 projects since 2019
- European projects (FAIR-EASE)
- National projects (ATLASea)
- eDNA Observatory (ROME)



DATARMOR supercomputer

- 260 bioinfo accounts
- 580 softwares / 203 public data sets
- Galaxy and cmdline access
- Training sessions on HPC uses, FAIR data management, Metabarcoding, Genome assembly, Transcriptomics, Nextflow
- Bioinformatics teaching (Master 2, University of Rennes; with ABIMS and Genouest)

Free and open-source softwares, pipelines & training materials:

<https://gitlab.ifremer.fr/bioinfo>



Bioinformatics data management protocol & tool



SeBiMER team



**Patrick Durand, PhD
Manager**



2017-now



Laura Leroi, engineer



2018-now



Cyril Noël, PhD



2019-now



Alexandre Cormier, PhD



2020-now



Pauline Auffret, engineer



2022-now



David Goudenege, PhD



2024-now



**Master students
in season**



**Marie Jossé, engineer
(ISI/SeBiMER)**



2023-2025

About FAIR data management for bioinformatics projects

- **Data handling of DNA sequencing information... mostly: huge text files (*fastq.gz format*)**

ID	@M00801:988:000000000-KYPK9:1:1101:9531:14493 1:N:0:320
DNA	CGTCAGATGTGTATAAGAGACAGCATGGGAGG
	+
Score	BCCCCCFFFFFFGGGGGGGGGHHHHHHHHGG

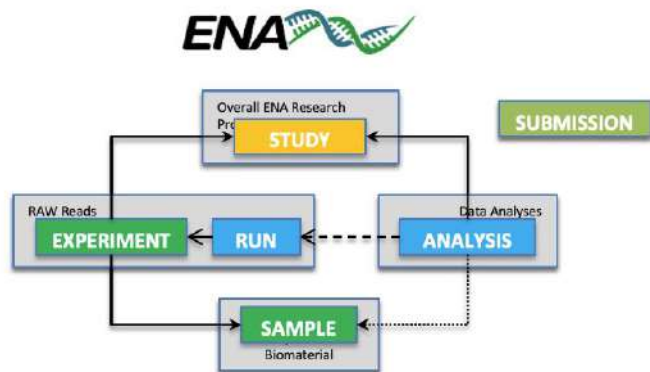
x several millions

Standard fastq file

- **Currently: 280+ data projets; 60Tb of raw sequencing data on DATARMOR**
(mostly: eDNA, RNASeq, WGS)

About FAIR data management for bioinformatics projects

- Metadata management describing experiment and sequence files; must follow international standard such as those from ENA (*European Nucleotide Archive*)



« the metadata model »



1. Collect metadata
2. Validate everything
3. Upload to ENA

rule	mandatory	mandatory	mandatory	mandatory
1	mandatory	mandatory	mandatory	mandatory
2	format	free text	free text	free text
help	Temporary accession number, format: YYYYMMDD_LABNAME_SUBMITTER_INITIALS	Short name for the study (e.g. DATAREF project name)	Title of the study as would be used in a publication. Must contain the following elements: study type, genus, species, project name, year. Example: Whole genome sequencing of Atlantic bluefish tuna for THON project, 2022.	More extensive free description of the st
3	tag	alias	name	title
4	value			study_description
5				
6				
7				
8				
9				
10				
11				

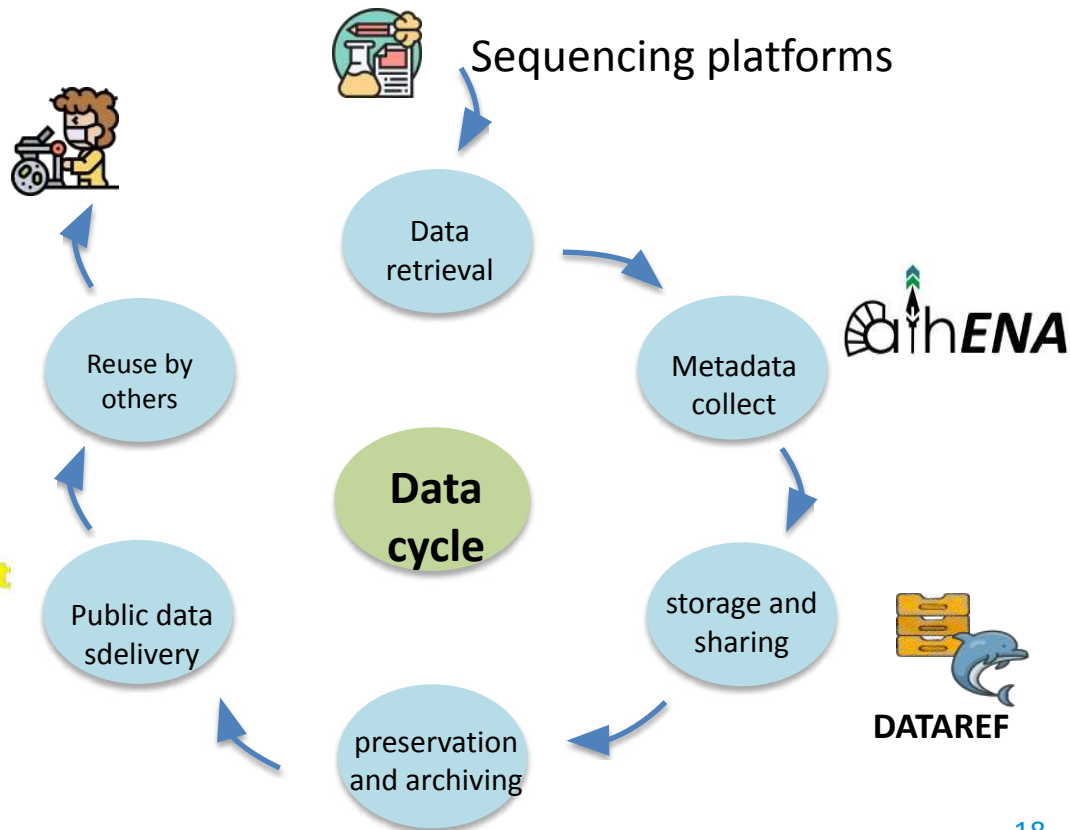
« the metadata model implementation for data validation and brokering »

<https://gitlab.ifremer.fr/bioinfo/workflows/athena>

FAIR data management cycle for bioinformatics projects at Ifremer

Data management

From data retrieval to public delivery to scientific community





ABiMS - SeBiMER synergies



SeBiMER
Bioinformatique marine

ABiMS



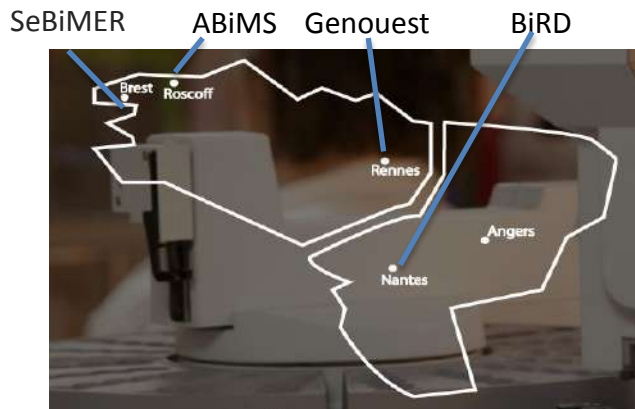
Biogenouest

GIS Bretagne / Pays de la Loire



- 22 institutions - 34 technological platforms – 428 people – 11 national research infrastructures
- 7 areas : genomics, proteomics, functional exploration, bio-imagery, structural analysis and metabolomics, **bioinformatics**

Bioinformatics area



- Biogenouest bioinformatics area focus on:
 - Pooling resources
 - Pooling knowledge and know-how
 - Enhancing team skills
 - Joint training
 - Joint projects (FP management)
 - Links with IFB (Institut Français de Bioinformatique)

Training and teaching

- Since 2014, schedule several sessions on phylogeny analysis, metabarcoding data analysis, genome assembly and annotation;
- Since 2024, teaching on FAIR principles applied to genome assembly, Master of Bioinformatics, University of Rennes;
- Some trainings and teachings are organized with Genouest, sometimes in the context of EBAll and IFB

R&D projects

- **MyGOD**: a web-based platform to analyse, visualize and share eDNA-based diversity data (project funded by Biogenouest);
- **Phaeoexplorer**: a web-based platform to analyse, visualize and share RNASeq data of viruses;
- **BEAURIS**: a Gitlab CI/CD pipeline to compile and push on production web portals to share and visualize genomes data (project headed by Genouest);
- **ORSON**: a Nextflow pipeline to annotate genomics and transcriptomics data
 - *BEAURIS and ORSON expected to be applied on 4500+ genomes produced by the ATLASea project (2023-2028)*

Co-organisation : C. Noël (Ifremer), E. Corre (ABiMS), E. Morin (INRAe)

EventsR





SeBiMER
Bioinformatique marine



ABIMS