



EMBRC
EUROPEAN
MARINE
BIOLOGICAL
RESOURCE
CENTRE

EMBRC - European Marine Biological Resource Center

EMBRC use case

K. Deneudt, I. Nardello
M. Obst, A. Hardisty, E. Corre, M. Hoebeke, G. Le Corguillé

EGI conference 2017

EGI-INDIGO workshop on community application support

May 10th, 2017

Catania

The **European Marine Biological Resource Centre** (EMBRC) is a distributed research infrastructure that aims to provide a strategic delivery mechanism for excellent and large-scale marine science in Europe. With its services, EMBRC will support both fundamental and applied research based on marine bioresources and marine ecosystems. In particular, EMBRC aims to drive forward the development of blue biotechnologies. EMBRC will provide the suitable research environment for users from academia, industry, technology and additional sectors.



BIOLOGICAL RESOURCES

ECOSYSTEM ACCESS

EXPERIMENTAL FACILITIES

TECHNOLOGY PLATFORMS

TRAINING AND EDUCATION

E-INFRASTRUCTURE, DATA AND SERVICES

SUPPORTING FACILITIES

EXPERT ADVICE

2 ESFRI PROJECTS

A world-class platform for fundamental and applied research on marine bioresources and marine ecosystems

TYPE: distributed
COORDINATING COUNTRY: FR

PARTICIPANTS: BE, EL, ES, FR, IL, IT, NO, PT, UK

TIMELINE

- ESFRI Roadmap entry: 2008
- Preparation phase: 2011-2014
- Construction phase: 2014-2016
- Operation start: 2016

ESTIMATED COSTS

- Capital value: 126 M€
- Preparation: 3,9 M€
- Construction: 4,5 M€
- Operation: 6 M€/year

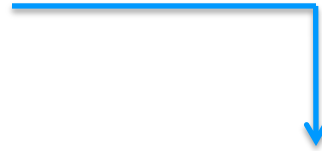
HEADQUARTERS

University Pierre and Marie Curie-UPMC
Paris
France

WEBSITE

<http://www.embrc.eu/>

User



To

- On-site access
- Remote access
- Virtual access

Service catalogue & request system

Biological resources



Ecosystem access



Experimental facilities



Technology platforms



Training and education



E-infrastructure, data and services

cdn12p0003r18	165.00
cdn07p0002m09	184.00
di012	226.00
cdn02p0003p14	230.00
cdn02p0003r22	243.00
isp000300	245.00
cdn07p0002e17	270.00
cdn07p0000a14	270.00
cdn12p0002h23	270.00
cdn08p0001p23	270.00
cdn01p0001a13	297.00
cdn03p0002h09	297.00
cdn04p0005g04	297.00
cdn08p0002b16	297.00
cdn13p0005g02	301.00

Supporting facilities



Expert advice



- **Listing initial requirements**

- Data storage and Computing capacity; Networking and connectivity; Repositories; Integrated thematic databases; Analysis tools; Registers and catalogues; Administrative tools; IT & Data management capacity; Training & standards

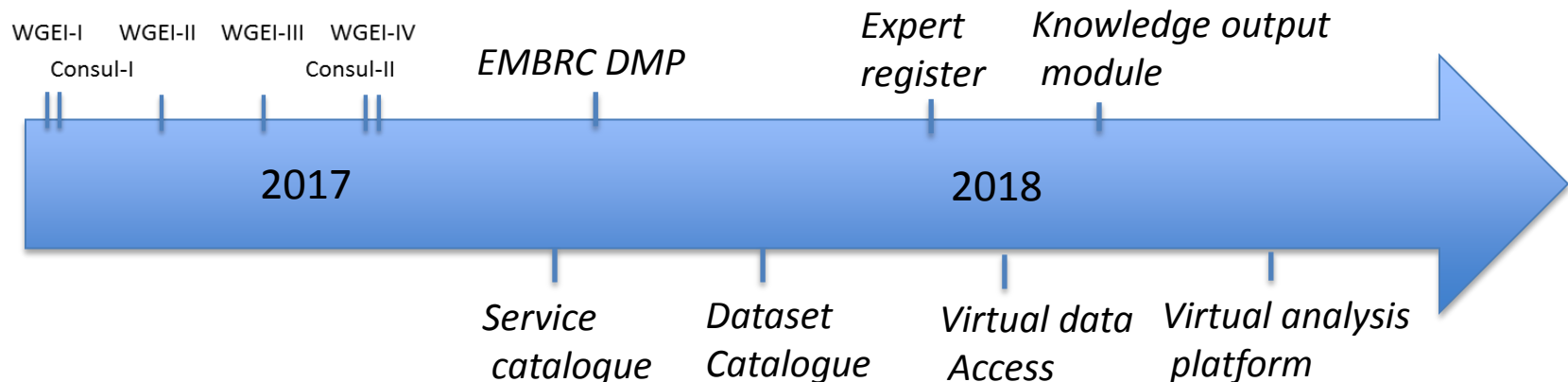
- **Preferred approach based on co-development and interoperability**



- **Defining architecture model**

- **Defining Priorities for implementation**

- **Data policy and DMP for EMBRC**

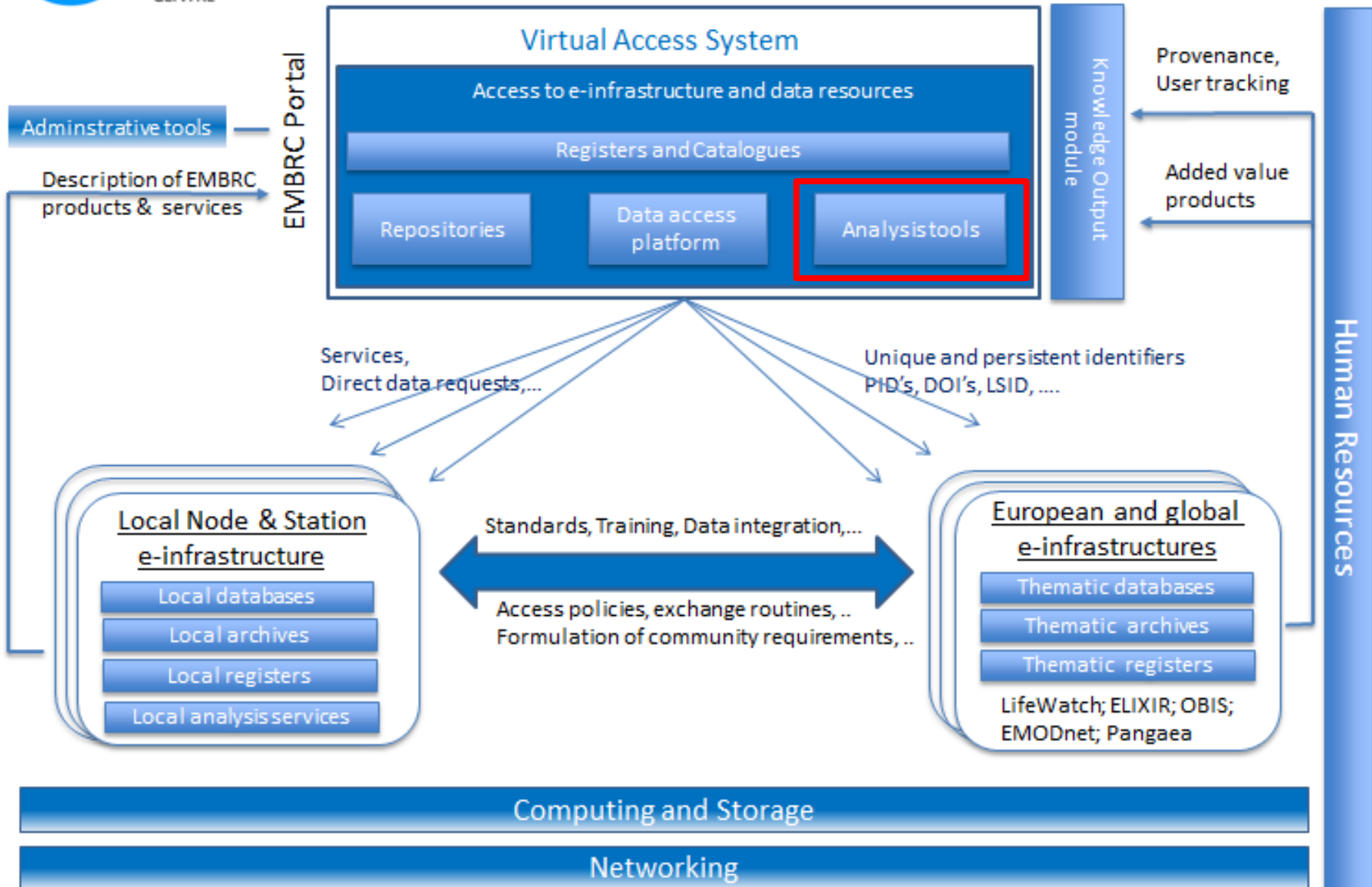




EMBRC
EUROPEAN
MARINE
BIOLOGICAL
RESOURCE
CENTRE



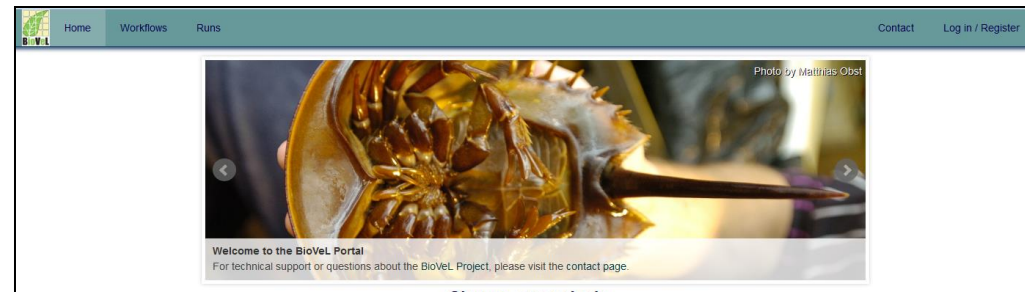
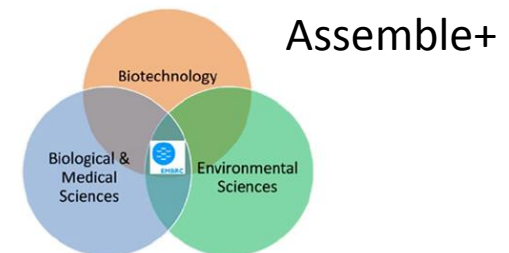
EMBRC e-infrastructure



- 1) Blue portal - on-line collaboration through sharing of e-workflows and workflow runs
- 2) ABIMS - Analysis and Bioinformatics for Marine Sciences (Galaxy instance)
- 3) EcoTaxa - Online plankton image annotation system
- 4) Genomic Observatories - Ocean Sampling Day pipelines

Community and use

- Existing tools developed as part of the BioVeL project (<https://portal.biovel.eu>)
- > 60 Web services and a large number of reusable workflows
- Biodiversity catalogue (<https://www.biodiversitycatalogue.org/>)
- User Community
 - > 540 registered users
 - > 50 papers

Requirements

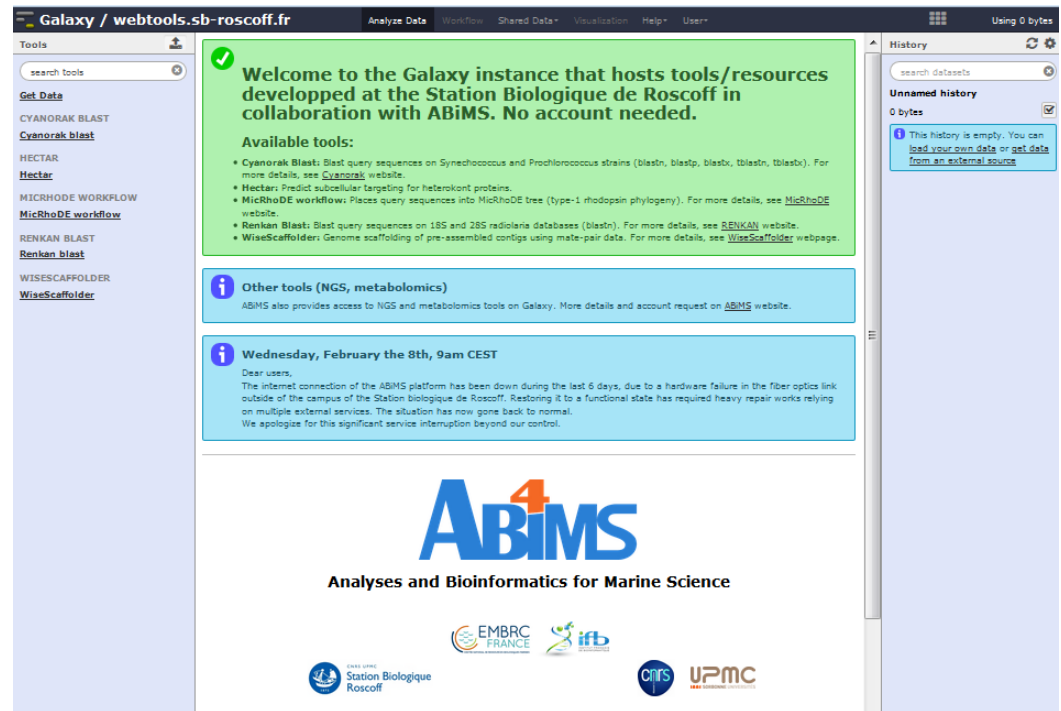
- The current biovel portal on Amazon AWS
 - Uses an EC2 instance of type "m3.xlarge".
 - Comfortable meets the present workload
 - When higher capacity is needed (e.g., for a workshop) additional instances are spinned up. (Max 4 or 5).
- Bleu Portal instance installed at VLIZ in the framework of Lifewatch (down)
- A Portal-package (Biovel-in-a-box) has already been made available at the EGI marketplace: <https://appdb.egi.eu/store/vappliance/biovel.portal>. (needs updating)
- Computational capacity:
 - Start implementing with moderate computational capacity, and only after scale up and stress the system after one year with large data processing tasks, e.g. e-workflows by calculating data products such as marine Essential Biodiversity Variables (EBV).
- The service requirement is 24Hx365Days, with 98% uptime being acceptable.
- 3h intervention would be ok.
- Daily backups of data associated with workflow? (never been asked to reinstate one. Most people are happy to re-run their work at present.)

Provisional list of items that might be useful include:

- Identify and Access Management
- OpenStack Cloud
- UDocker, or more likely VMs under control of the Infrastructure Manager
- Future Gateway
- PaaS Orchestrator

Community and use

- Analysis and Bioinformatics for Marine Sciences (SB Roscoff) - Galaxy instance
- User community: daily scientific use
 - National and international projects: Idealg, Oceanomics, EMBRIC, ELIXIR-Excelerate
 - EMBRC (Assemble+)



The screenshot shows the Galaxy web interface at webtools.sb-roscoff.fr. The main content area features a green welcome message: "Welcome to the Galaxy instance that hosts tools/resources developed at the Station Biologique de Roscoff in collaboration with ABIMS. No account needed." Below this, a list of available tools is provided, including Cyanorak Blast, Hectar, MicRhoDE workflow, Renkan Blast, and WiseScaffolder. A blue information box states: "Other tools (NGS, metabolomics) ABIMS also provides access to NGS and metabolomics tools on Galaxy. More details and account request on ABIMS website." Another blue information box provides a service update: "Wednesday, February the 8th, 9am CEST. Dear users, The internet connection of the ABIMS platform has been down during the last 6 days, due to a hardware failure in the fiber optics link outside of the campus of the Station biologique de Roscoff. Restoring it to a functional state has required heavy repair works relying on multiple external services. The situation has now gone back to normal. We apologize for this significant service interruption beyond our control." The footer contains the ABIMS logo and the text "Analyses and Bioinformatics for Marine Science", along with logos for EMBRC France, ifb, Station Biologique Roscoff, CNRS, and UPMC.

Requirements

- The Galaxy instance is currently running on subsection of a local cluster of 1600 CPU and scale to the analyse of small /medium size project (up to 100-150 million of reads)
- We would like to see if a replication could be installed at EGI/INDIGO CLOUD in order to provide additional computing power
- Data:
 - for Galaxy instance: 4-6G
 - for a project : raw data (up to 200 Go) and temporary and final data up to 1 - 1,5 To.

- Provisional list of items that might be useful include:
 - Identify and Access Management
 - OpenStack Cloud
 - Docker, or more likely VMs under control of the Infrastructure Manager and specific Galaxy docker (<https://github.com/bgruening/docker-galaxy-stable>)
 - Future Gateway
 - PaaS Orchestrator
 - In the case of Galaxy instance cloudification may be tools as Cloudman (<https://galaxyproject.org/cloudman/>) and galaxy ansible for installing and managing Galaxy servers (<https://github.com/galaxyproject/ansible-galaxy>).

3) EcoTaxa - Online plankton image annotation system

- 40 millions plankton images & their metadata and annotation
- Deployable instances under development – growing collections
- Long term storage capacity servers
- Specific requirements yet to be discussed

4) Genomic Observatories - Ocean Sampling Day pipelines

- Analysis pipelines creating Gene function matrices, Taxonomic matrices, Environmental data matrices,...
- Raw data storage
- Specific requirements yet to be discussed



EMBRC
EUROPEAN
MARINE
BIOLOGICAL
RESOURCE
CENTRE

Thank you