



HAL
open science

FAIRing Research Data to Live @AuBi Platform

Francis Ogereau, Mateo Hiriart, Pierre Marin, Bérénice Batut, Cléa Siguret, Philippe Ruiz, Sophie Desset, Pierre Pouchin, Faustine Souc, Nils Paulhe, et al.

► **To cite this version:**

Francis Ogereau, Mateo Hiriart, Pierre Marin, Bérénice Batut, Cléa Siguret, et al.. FAIRing Research Data to Live @AuBi Platform. JOBIM, Jun 2024, Toulouse (FR), France. hal-04643973

HAL Id: hal-04643973

<https://uca.hal.science/hal-04643973>

Submitted on 10 Jul 2024

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 Attribution - NonCommercial - ShareAlike 4.0 International License

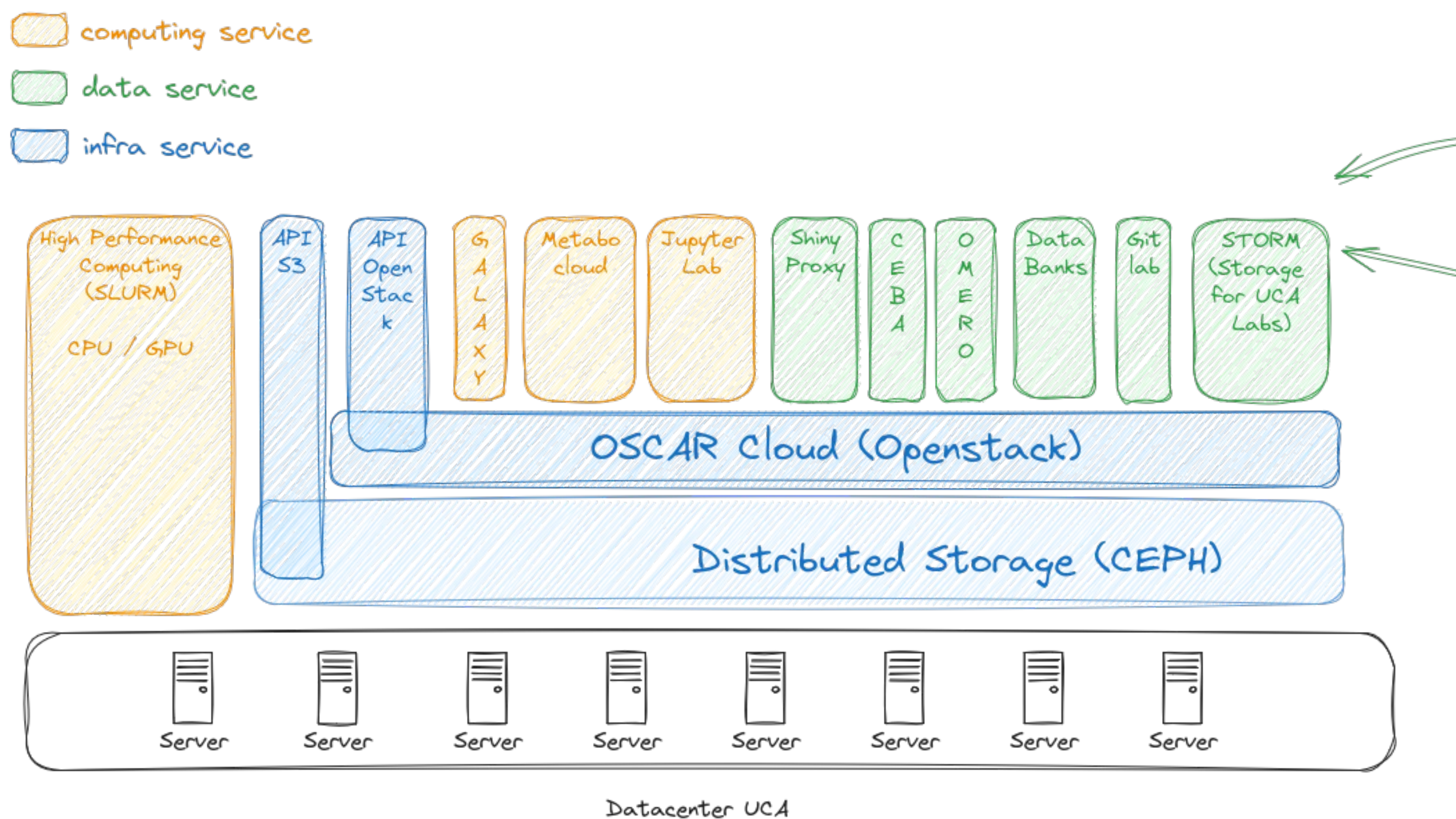
FAIRing Research Data to Live @AuBi Platform

Francis OGEREAU¹, Mateo HIRIART¹, Pierre MARIN^{1,2}, Bérénice BATUT^{1,2}, Cléa SIGURET^{1,2}, Philippe RUIZ³, Sophie DESSET⁴, Pierre POUCHIN⁴, Faustine SOUC⁵, Nils PAULHE⁵, Franck GIACOMONI⁵, David GRIMBICHLER¹, Thomas BELLEMBOS¹, Valérie LEGUÉ⁶, Pierre PEYRET³, Antoine MAHUL¹ and Nadia GOUÉ^{1*}

1 Université Clermont Auvergne, Plateforme AuBi and Mésocentre Clermont-Auvergne, F-63000 Clermont-Ferrand, France ; 2 CNRS, Institut Français de Bioinformatique, IFB-Core, UAR 3601, 91000, Evry, France ; 3 Université Clermont Auvergne, INRAE, MEDIS, F-63000 Clermont-Ferrand, France ; 4 Université Clermont Auvergne, Inserm, iGrED, F-63000 Clermont-Ferrand, France ; 5 Université Clermont Auvergne, INRAE, UNH, Plateforme d'Exploration du Métabolisme, MetaboHUB Clermont, F-63000 Clermont-Ferrand, France ; 6 Université Clermont Auvergne, INRAE, PIAF, F-63000 Clermont-Ferrand, France * corresponding author : nadia.goue@uca.fr

Hosted by the Pôle d'Appui Recherche Mésocentre Clermont-Auvergne (PARM), the Auvergne Bioinformatique (AuBi) platform began structuring its community in 2015. AuBi is member of the Institut Français de Bioinformatique (IFB) since 2018, and France Génomique since 2022. AuBi contributes to Open Science through joint initiatives with the recently labeled Atelier de la Donnée : Gérer Ouvrir Trouver Obtenir les Données de l'Université Clermont Auvergne (GOTODO).

Datacenter & Mesocenter Infrastructures



- PARM current services
- ✓ HPC
 - 2 500 CPU, 14 GPU
 - Access : CLI, X2Go, JupyterLab
 - Gitlab server
 - Centralized authentication (IAM)
 - ✓ Cloud OSCAR
 - 1 900 vCPU, ~ 10 Tb RAM
 - ✓ Storage STORM
 - 3 PB via Samba (CEPH)
 - Storage (CEPH, S3)
 - 1 800 Tb

In the coming year, and through a FEDER funding, we will update to a new HPC with 24 nodes (128 CPU, 1.5 TB RAM and 2 x 8 cards H100 GPU). In 2025, cloud service will be increased with new hypervisors (550 K€) and STORM storage will reach 2 Pb.

Community of AuBi users



Around 15 labs and sequencing and imaging platforms contribute to AuBi activities with scientific questions related to health, microbes, agronomy, and environment disciplines.

Assistance to complete project DMP

dsw.france-bioinformatique.fr

- ✓ Working with diversity of communities
- ✓ Offers entity DMPs:
 - Imaging (OMERO as AuBi service)
 - Imaging with CLIC platform
 - Proteomic & Metabolomic with Metahub

dmp.opidor.fr

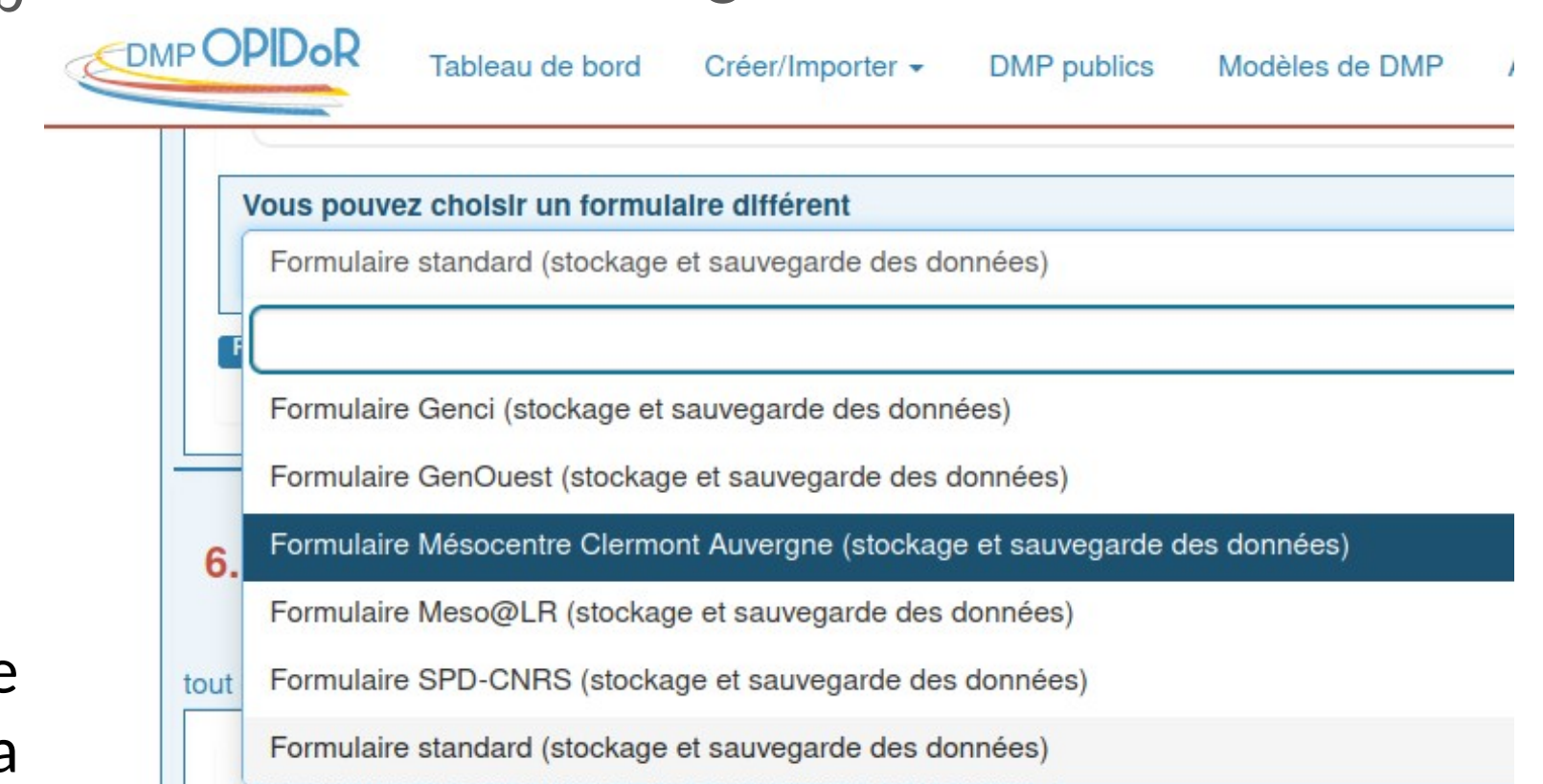
- ✓ AuBi and PARM indexed on cat.opidor.fr
- ✓ Collaboration with INIST team to :
 - Analysis cost on cluster
 - Storage cost

Bioimaging

Metabolomic



As part of the MuDiS4LS project, IFB-core and IFB members collaborate with Data Stewart Wizard to produce entity DMP.



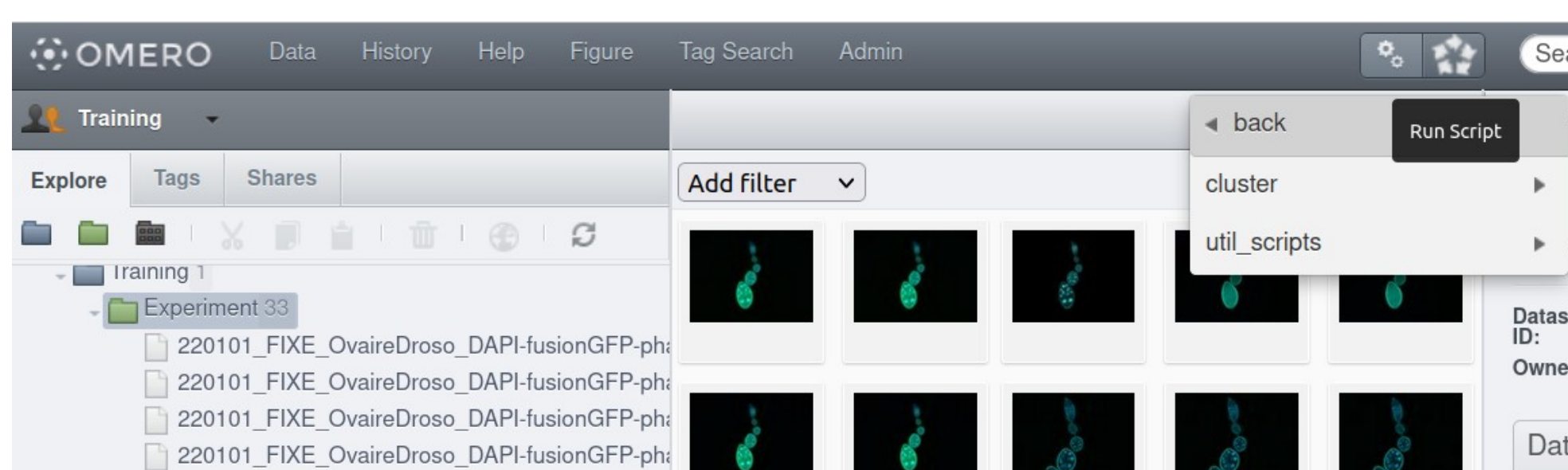
Origin & Diversity of Data stored and analyzed at AuBi

Category	Sequences	Images	Quantitatives Qualitatives	Digital models
Sequences	GDEC, MEDIS, PIAF, M2iSH, UNH, Laboratoire Microorganismes Genome et Environnement			
Images	GReD, IMOST, GDEC, NEURO-DOL, PIAF			
Quantitatives Qualitatives	PIAF, UNH, INRA, MEDIS, UMR1213			
Digital models	PIAF, GReD, NEURO-DOL, MEDIS			

FAIRisation of Image Data by Metadata Storage Management

omero.mesocentre.uca.fr

- ✓ Adding a "Public" folder with the possibility to get an DOI
- ✓ Guidance for metadata



50 active users
12 projects
3 Tb

FAIRisation with Reproducible Workflows in Galaxy

galaxy.mesocentre.uca.fr

- ✓ Open access tools
- ✓ Active communities : Hackathons [bio.tools]

usegalaxy.fr

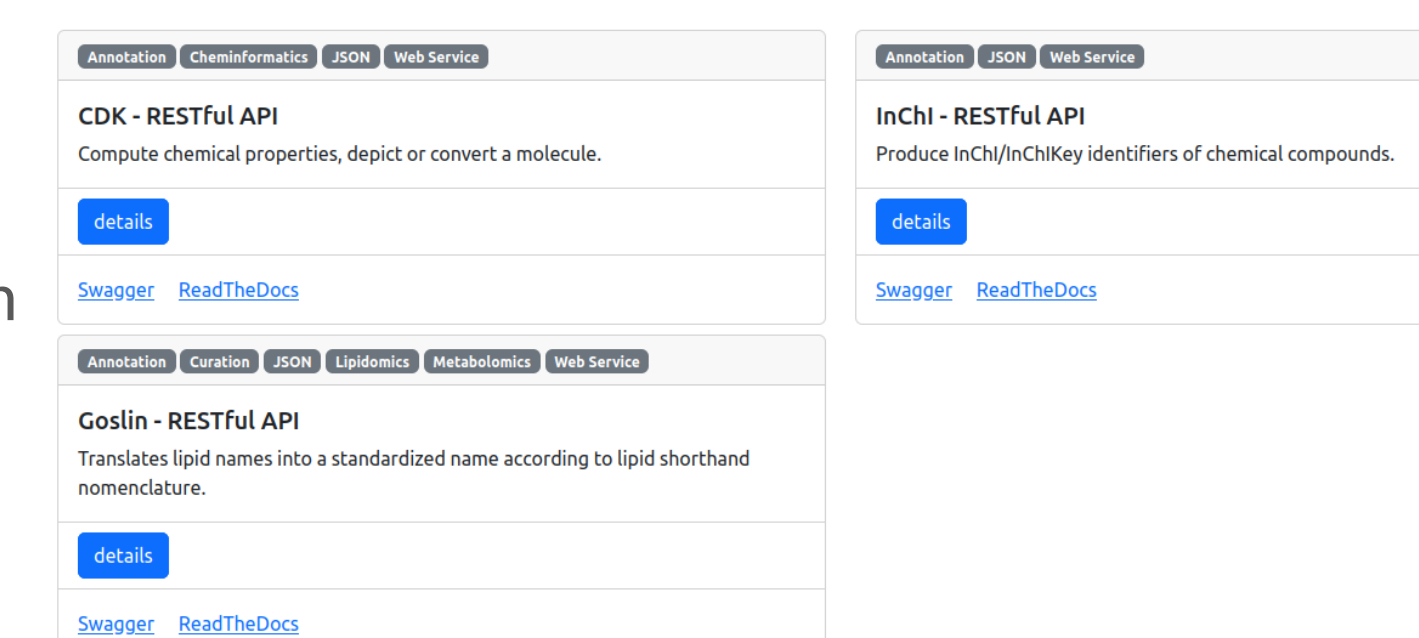
- ✓ ABRomics project
- ✓ Specific workflows
- ✓ Specific Galaxy Training Network

- ✓ Galaxy Working Group & Eurobioimaging
- ✓ Enable reuse of workflows [DOI]
- ✓ Active with Galaxy Training Network

97 active users
30 Tb

FAIRisation with API microservices

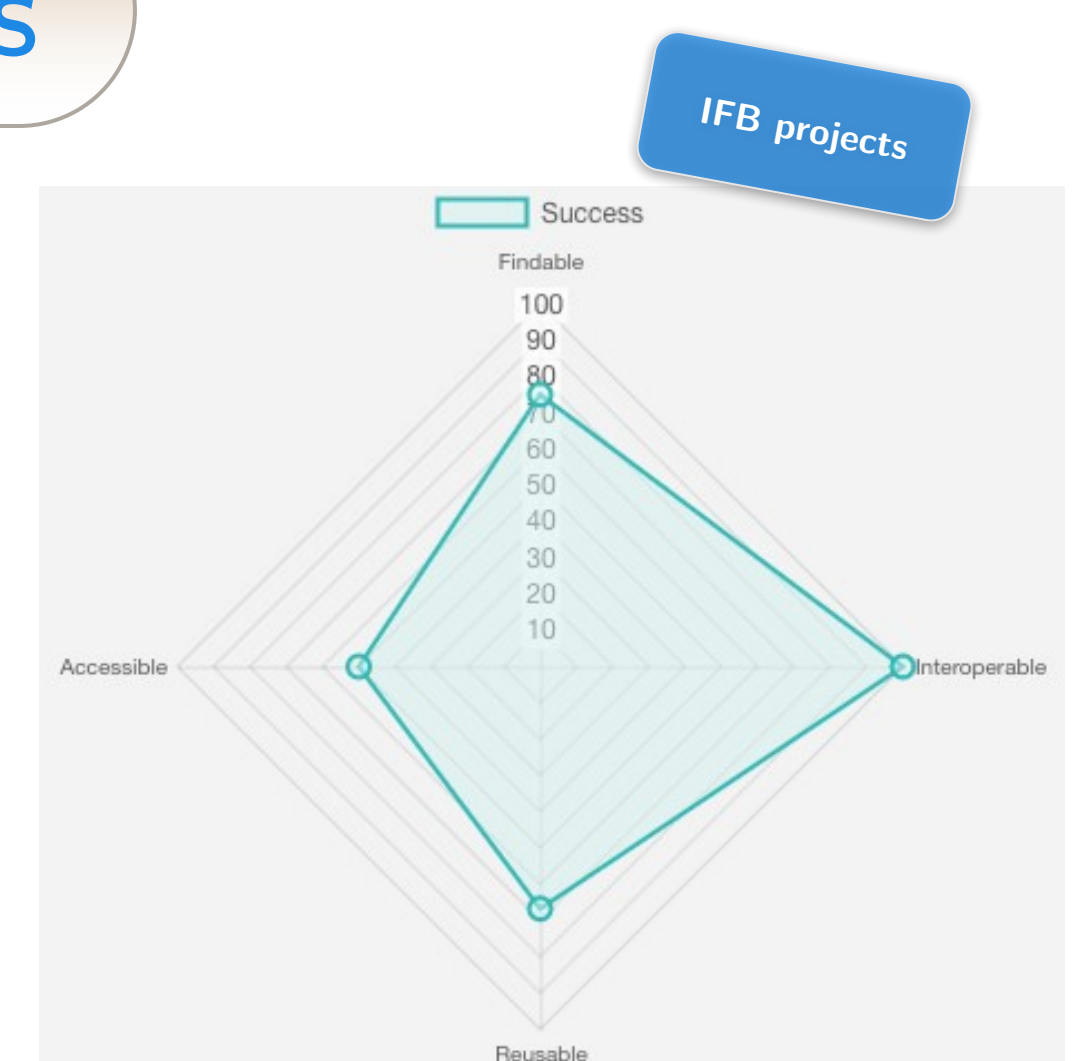
Welcome on the MetaboCloud catalog !



metabocloud.mesocentre.uca.fr

- ✓ Open access tools through API
- ✓ Including metadata and documentation
- ✓ bio.tools identification
- ✓ Improving website with FAIRChecker

GO AND SEE POSTER : MetaboCloud : A catalog of microservices hosted on a Cloud infrastructure and addressing issues linked to FAIR principles and open science



Trainings for FAIRisation and Open Science

Coordinating with PARM, Atelier de la donnée (G.O.T.O.D.O) and IFB actions

- ✓ Training for metadata management with OMERO
- ✓ Training "Atelier de la donnée" on data management (PhD students)
- ✓ Training for project management with DSW (Master degree)
- ✓ Training for compute best practices, Snakemake workflow management
- ✓ Working session with Galaxy Training Network
- ✓ Using IFB FAIR-Bioinfo training support for AuBi community