

INRAE

➤ AG du CATI Bioinformatics for Omics and metaOmics of Microbes (BOOM)

7 et 8 janvier 2021

Points à discuter

News CATI (projets en cours, formation ThinkR, AG Massy fin 2021 ?, nouveaux arrivants, site web, Intercati : nouvelles dates hackathon) 3-4 dias

News Migale & Bioinformatics (Valentin) Infra stockage, rstudio, shiny, formations, tutos

-

8-10 dias

News IFB PIA3, PF Antibiorésistance (Hélène) 3-4 dias

...



Nouvelles du CATI

Événements de l'année 2020-2021

Hackathon inter-cati 2020 :

- bascule à distance sur deux jours
- 2 ateliers :
 - workflows
 - Shiny

Reprogrammation du hackathon du 04 au 07/10/2021 à Sète (*à confirmer*)

Formation interne 2021 :

- Développeur R “avancé”-
Création de packages versionnés
- Société ThinkR, à distance
- 8 participants qui feront un retour aux intéressés (contenu *et* pédagogie à distance)

- AG BOOM à reprogrammer fin 2021

- Projet MICA Open16S



Nouvelles du CATI

Organisation des WP et principales actions de l'année 2020

WP DATA et WP ANNOT : projet de fusion avec un binôme d'animateurs AL Abraham et H. Devillers

- Fiches d'évaluations de ressources d'annotations fonctionnelles
- Tutos de soumission à l'ENA

WP STATS : animation par M. Berland et J. Aubert

- Evaluation d'applications Shiny
- Veille

WP TEXT : animation par R. Bossy

- Montage d'une formation au text mining avec le CATI GREP (inter CATI)

TRANSVERSAL, INTER-CATI

- site web, tutos (Gitlab pages), retours d'expériences MOOC, hackathons



Nouvelles du CATI (suite et fin)

Bienvenue à Véronique Jamilloux qui nous rejoindra en 2021

Merci à Valérie pour le nouveau site web :
<https://boom.cati.inrae.fr>

Interactions avec DipSO et département MathNum prévues en janvier-février 2021



Migale : faits marquants

Infrastructure & outils

Nouveaux serveurs de virtualisation et nouveaux espaces de stockage centralisés (netApp)

Réorganisation des espaces utilisateurs

- Changement des chemins
 - `/projet/groupe/save/login -> /home/login/save`
 - `/projet/groupe/work/login -> /home/login/work`
- 1050 espaces utilisateurs migrés

Quotas par utilisateur

- `/home/login` – 1 Go
- `/home/login/save` – 50 Go
- `/home/login/work` – 150 Go
- **Quota doublé** : 100 Go save / 300 Go work pour les **utilisateurs du cluster** (sur demande)

Save : snapshot journalier, rétention 1 semaine

- Accès sur `/save_home/snapshot/LOGIN`

Facturation des demandes d'espace supplémentaires

- 330 € / To / an

Pour les projets avec **utilisateurs multiples** :

- Espace /projet
- Quotas (work / save)
- Gestion des droits

Espaces **Galaxy** toujours gérés indépendamment

Généralisation de Conda

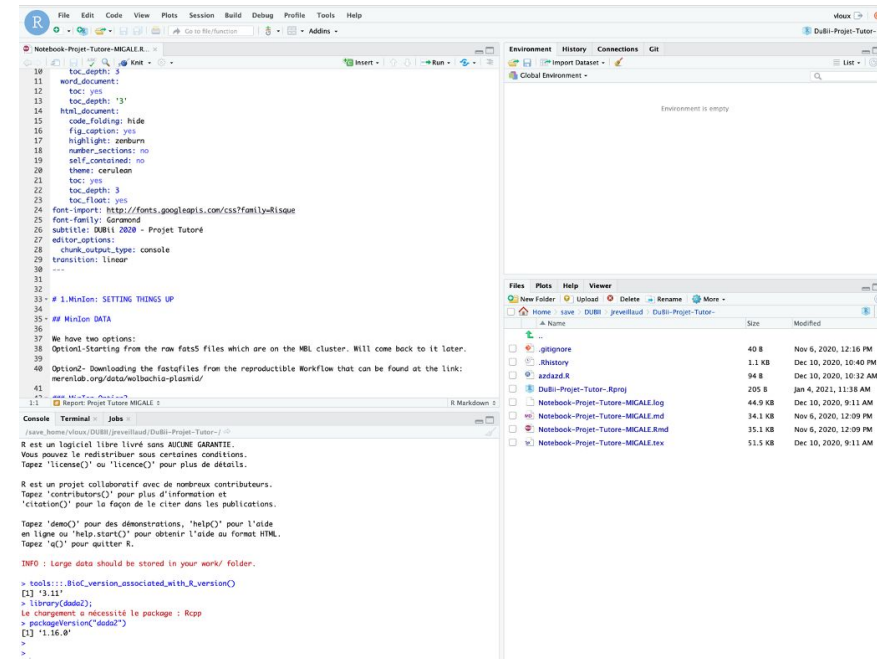
Perspectives 2021 :

- Noeud bigmem (2 To)
- Renouvellement d'une tranche du cluster (1536 coeurs, financement demandé).
- Chantier de **gestion des comptes** utilisateurs (charte, données,...)

Migale : faits marquants

Rstudio

- Depuis un navigateur web
- Espaces personnels (work/save) accessibles
- Mêmes paquets que le R en ligne de commande
- Pour le prototypage / les analyses rapides
- Les calculs longs doivent être lancés sur le cluster



<https://rstudio.migale.inrae.fr>



INRAE

AG du CATI BOOM

7 janvier 2021 / Introduction / H. Chiapello & V. Loux

Migale : faits marquants

Cycle de formation à la bioinfo par la pratique

Cycle 2020

- 14 modules uniques
- 15 sessions de formations.
- 20 jours cumulés.
- 23 formateurs différents
- 85 personnes formées (-30%)
- **Formations maintenues** malgré le contexte sanitaire
- Tarifs harmonisés au sein de BioinfOmics

3 nouveaux modules :

- **metagénomique shotgun** (2 jours, lc)
- **génomique comparée** (1 jour, Galaxy)
- R graphics with **ggplot2** (1 jour, Rstudio)

Cycle 2021 :

- Deux nouveaux modules :
 - Shiny
 - Bonnes pratiques

Modalités pédagogiques :

- formations majoritairement en présence, avec report si nécessaire
- formation “NGS sous Galaxy” à distance

Migale : faits marquants

Tutoriels Migale

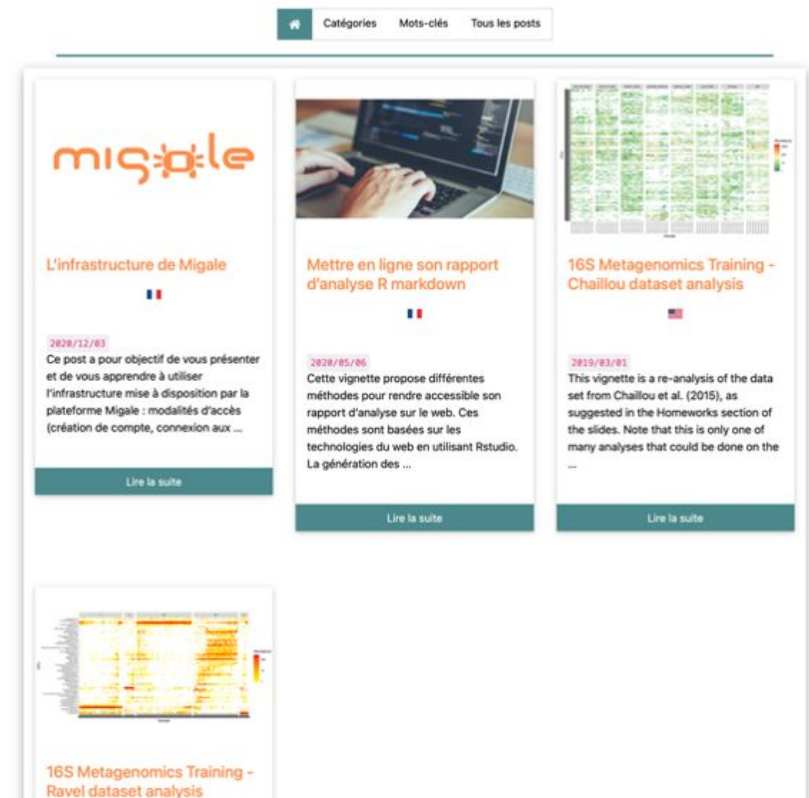
Cas pratiques, explications et accompagnement pas à pas

- Vidéos
- Exemples pratiques
- Quizz final

A venir :

- Snakemake
- Bonnes pratiques (git/github/md)

Documentation & Tutoriels



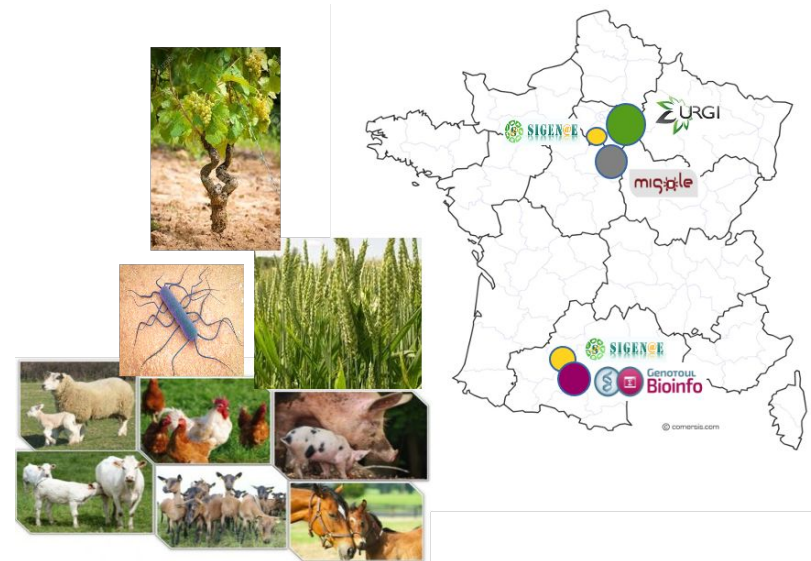
The screenshot shows the 'Documentation & Tutoriels' page of the Migale website. At the top, there are navigation links for 'Catégories', 'Mots-clés', and 'Tous les posts'. Below this, there is a grid of tutorial cards. Each card features the Migale logo, a title, a date, a brief description, and a 'Lire la suite' button. The visible cards are:

- L'infrastructure de Migale** (2020/12/03): Ce post a pour objectif de vous présenter et de vous apprendre à utiliser l'infrastructure mise à disposition par la plateforme Migale : modalités d'accès (création de compte, connexion aux ...)
- Mettre en ligne son rapport d'analyse R markdown** (2020/05/06): Cette vignette propose différentes méthodes pour rendre accessible son rapport d'analyse sur le web. Ces méthodes sont basées sur les technologies du web en utilisant Rstudio. La génération des ...
- 16S Metagenomics Training - Chaillou dataset analysis** (2019/03/01): This vignette is a re-analysis of the data set from Chaillou et al. (2015), as suggested in the Homeworks section of the slides. Note that this is only one of many analyses that could be done on the ...
- 16S Metagenomics Training - Ravel dataset analysis** (partially visible at the bottom):

BioinfOmics

Infrastructure de Recherche INRAE en bioinformatique

- 4 collectifs, > 30 ETP
- PlantBioinfoPF (BAP/SPE/ECODIV)
- MIGALE (MATHNUM/MICA)
- GenoToul Bioinfo (MATHNUM)
- SIGENAE (GA/PHASE/SA/MATHNUM)
-



Pourquoi BioinfOmics ?

Réponse à une demande institutionnelle :

organiser et « typer » les **infrastructures scientifiques collectives** d'INRAE pour les inscrire dans une **feuille de route** d'établissement, puis nationale

Pour les ISC's participantes :

Moyens : Accéder aux moyens dédiés aux outils collectifs labellisés (postes, équipements, fonctionnement, appui,...)

Coordination : actions de formation, réponses aux appels à projets, veille technologique et scientifique,...

Ecoute : interactions régulières avec les communautés, les Cds, le DISC et autres directions/délégations

Capacité plus grande de réponse aux appels à projets : projets de plus grande envergure, complémentarité des compétences

Visibilité : bénéficier d'un portail d'entrée unique avec ré-orientation sur nos PFs

•



BioinfOmics : comment

- Une AG de lancement « virtuelle » le 10/11
- Des groupes de travail qui se lancent sur des sujets identifiés par les membres du projet :
 - Formation
 - Données
- Réunions bi-mensuelles du comité de direction
- Une réponse fédérée et coordonnée au projets PIA3 Equipex+
- Une ambition de projet scientifique commun atour de compétences complémentaires : intégration de données, holobionte

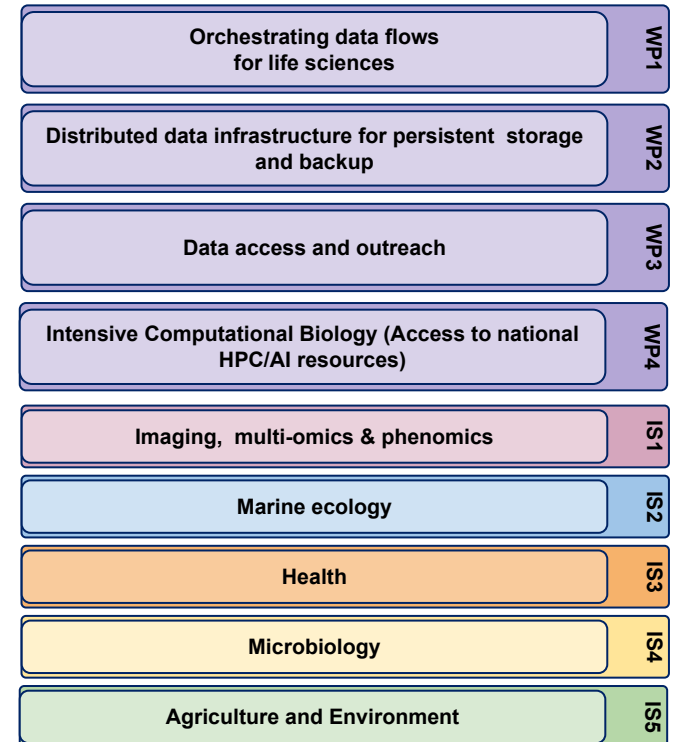


News de l'Institut Français de Bioinformatique

Projet IFB - PIA3 Espaces numériques mutualisés pour des données FAIR en biologie-santé (MuDiS4LS) accepté !

Porteurs : **J. van Helden** (co-direction IFB), **J. Seiler & G. Le Corguillé** (responsables Core Cluster IFB)

- 39 partner teams
- 14 partner organisms
- 4 national + 7 regional data centers
- 4 partner mesocenters
- 170 persons, 39 FTE;
- Requested funding: 20M€ (dont 13.7 M€ équipement)
- 4 technological work packages (WP)
- 5 thematic implementation studies (IS)
- Engagement des tutelles : 7 postes permanents



News de l'Institut Français de Bioinformatique

Data brokering et formations

> **Nouveau data broker à l'IFB depuis octobre 2020** Thomas Denecker

=> Plus de détails dans l'exposé à venir

Formations :

> **Nouvelle formation IFB "FAIRdata : Science ouverte et Plan de Gestion de données"**

Comment gérer des jeux de données haut-débit en sciences de la vie et de la santé ?

<https://ifb-elixirfr.github.io/IFB-FAIR-data-training/>

Première édition en mars 2021

> **EBAii (Ecole de Bioinformatique AVIESAN - IFB - INSERM) niveau 2** du 25 au 28 mai 2021, <https://ressources.france-bioinformatique.fr/fr/evenements/ebaii2021n2>

> **2ème édition de "FAIRbionfo : Comment rendre un projet plus reproductible "** les 28-29-30 juin 2021, voir <https://ifb-elixirfr.github.io/IFB-FAIR-bioinfo-training/>

> **3ème édition du DUBii (DU de Bioinformatique Intégrative) en préparation**
<https://www.france-bioinformatique.fr/dubii/>



News de l'Institut Français de Bioinformatique

Développement d'une plateforme nationale multi-omique intégrée pour la recherche en antibiorésistance

Projet en cours de montage

Coordination : J. van Helden (IFB), C. Médigue (IFB et Genoscope/Microscope), P. Glazer (I. Pasteur), R. Bonnet (CNR Résistance aux Antibiotiques, CHU Clermont)

Partenaires principaux : CNR, ANSES, CHU, I. Pasteur, MGP, IFB

Objectifs

Obj 1. Assurer l'hébergement d'une plateforme multi-omiques nationale (stockage, sécurité) et animer le réseau qui en dépendra

Obj 2. Créer et implémenter des banques d'échantillons, ainsi que des **bases de données microbiologiques multi-omiques intégrées** pour les secteurs Homme-Animal-Environnement en s'assurant de l'interopérabilité de ces bases entre elles et avec les bases de données Épidémiologiques

Obj 3. Développer, en parallèle, des **banques d'outils mathématiques et (bio)informatiques pour modéliser** l'évolution de la résistance aux antibiotiques, les phénomènes de transmission et de dissémination des clones et des gènes



Programme de la matinée

- 9h30-10h00 : **Accueil et news** (Hélène Chiapello & Valentin Loux)
- 10h00-10h30 : **Présentation du projet MICA Open 16S** (Marion Leclerc & Hélène Chiapello)
- 10h30-11h : **Présentation de la ressource et de l'API Florilège** (Robert Bossy)

11h00-11h15 : pause

- 11h15-11h45 : **Présentation du projet de data brokering IFB** (Thomas Denecker)
- 11h45-12h30 : **Retours d'expérience sur les outils d'annotation et bases de données de gènes d'antibiorésistance** (Sébastien Leclercq et Olivier Zemb)

