metagenopolis
mgps.eu

ACTIVEEON
SCALE BEYOND LIMITS

Microbiome Analytics

Machine Learning & ActiveEon Orchestration On-Prem and On-Clouds

Nicolas Pons, Sébastien Fromentin – INRAE-MetaGenoPolis
Denis Caromel, Amine Louati – ActiveEon

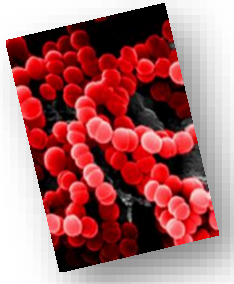
aristote **27 février 2020**

« Un check-up de l'IA pour la santé numérique, c'est grave docteur(s) ? »

by INRAE

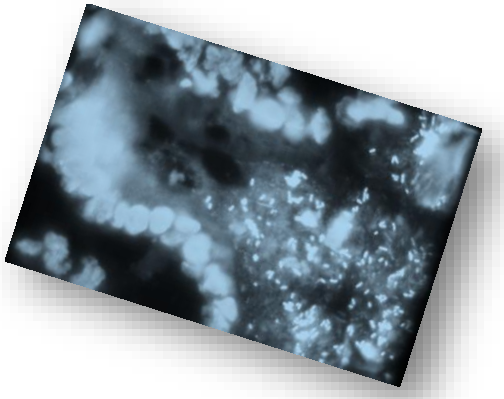
Our microbiota, a neglected organ...

100 000 000 000 microorganisms; 10 times more than cells in our own body, up to 2kg of our weight



A key actor between food and our body

In contact with the largest group of immune system cells, the second for neuronal cells (our second brain!)



...and a major actor in well-being, health and disease!

metagenopolis mgps.eu

MetaGenoPolis (MGP) is an **INRAE center** expert in gut microbiome research applied to health and nutrition.

In collaboration with industry, academia and clinics, MGP conceives and implements projects tailored to the partner's need.

PIA Winner: 19 M€ grant (2012-2019)
5.7 M€ grant (2020-2025)

INRAE

Pharmaceutical applications

Identification of bacterial biomarkers associated with dysbiosis: microbiome signatures used as a diagnostic or prognostic tools.

Stratification of individuals to a better personnalized medecine : microbiome responders vs non responders

Impact of drugs on the level of intestinal antibiotic resistance genes

Effects of drugs on the gut microbiome composition and functions

Identification of bioactive compounds (anti-inflammatory, antiproliferative, satiogenic)

Identification of microbial genes & metabolites able to interact with human cells (metabolic pathways, mechanisms of regulation...)


ACTIVEEON
SCALE BEYOND LIMITS

Activeeon
SCALE BEYOND LIMITS

Big Data + AI Solution

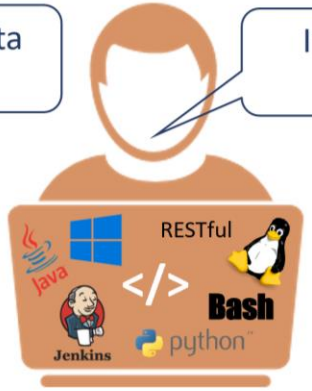
Software Vendor with worldwide references:

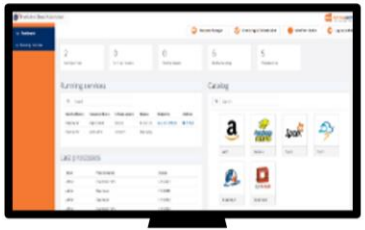
Home Office
SAFRAN Snecma
AXISPOINT HEALTH
McKESSON
KOMATSU
Legal & General
cnes
L'ORÉAL
INRA
cea
PUBLICIS GROUPE EPSILON
orange
RENAULT

Key Features:

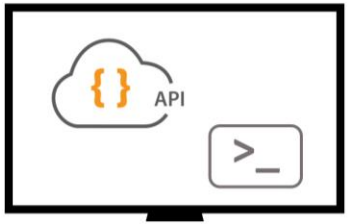
- No need to split your HPC Cluster
- Works with Slurm, PBS, LSF, & on various Premises, and Clouds
- Portals for all and New Kinds of Users
- Support Containers
- Secured On-Demand Services & GUI
- Ease of access to efficient execution, e.g. RDMA, Horovod, etc.
- An Open Platform!

Portals & Workflows for All

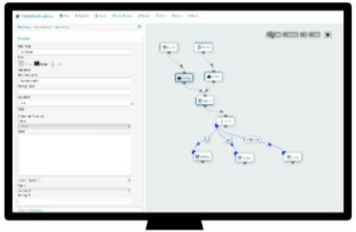




Automation Dashboard

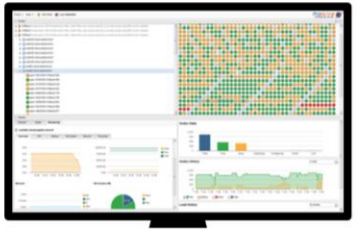


API / CLI



Workflow Studio

Scheduler



Resource Manager

Self-Service Catalog

Monitoring and Automation

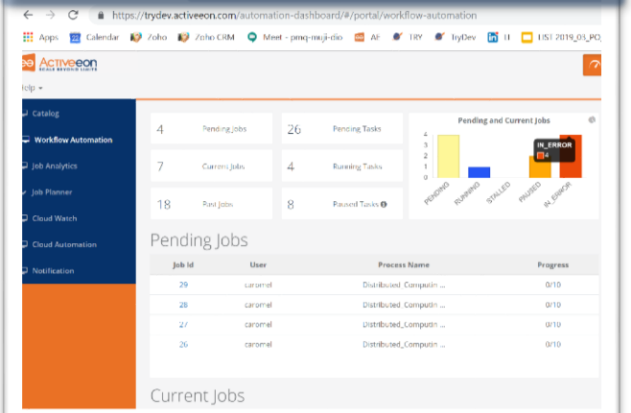
Workflows

Scheduler

Resource Manager

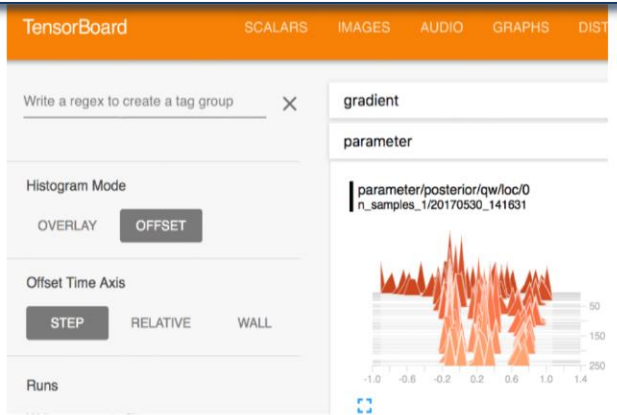
HPC + AI Unified Solution

Automation Dashboard



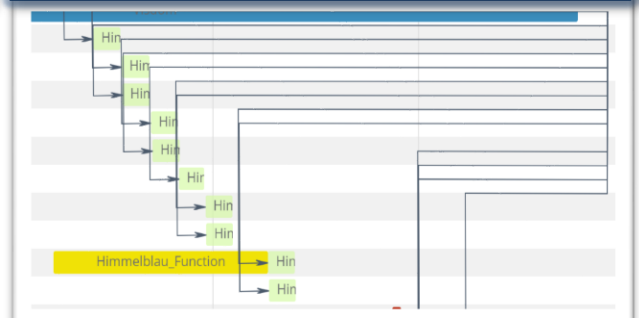
AI Studio & Data Connectors

User-defined GUI Services

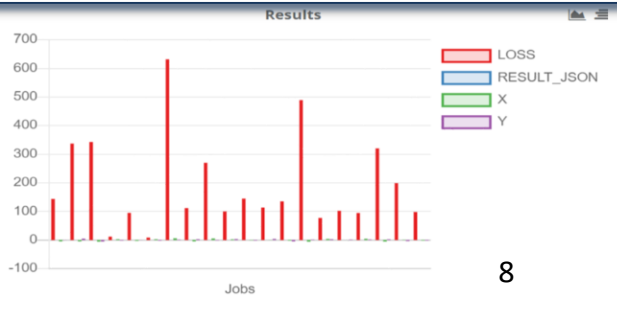


Python, Jupyter Integration

Job Dependencies - Gantt



Integrated Job Analytics



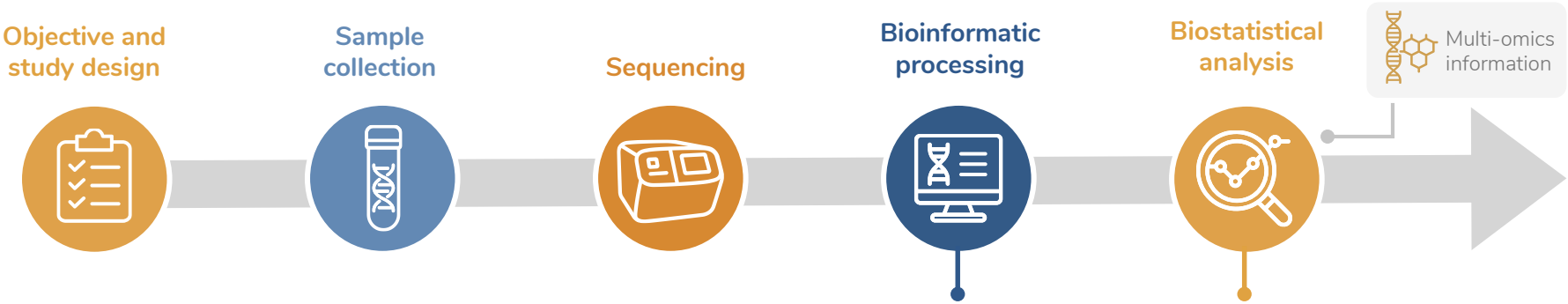
Microbiota as a clinical tool

Bringing Artificial Intelligence in microbiome

Acurate profiling with high resolution microbiome analysis

metagenopolis mgps.eu

Integrated metagenomic pipeline, exhaustive analysis at strain level



Some crucial steps conditionning all the analysis:

- Sample collection, conservation and processing
- Exhaustive and reproducible DNA extraction
- Sequencing (quality, depth, speed)
- Large genome reference ressources
- Big data access and security
- Big data analytics

10.4 M genes
Wen et al., Genome Biol 2017

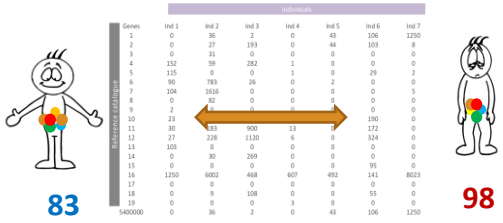
Nielsen, Almeida et al. Nature biotechnology 2014

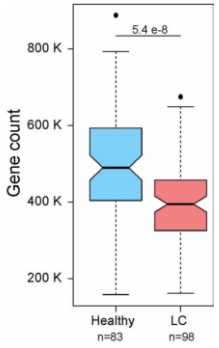
Plaza onate et al. Bioinformatics 2019

INRAE

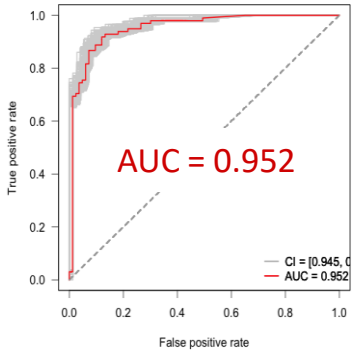
Diagnostic tool in liver cirrhosis

(Qin N. et al., Nature 2014)





discovery (83+98)



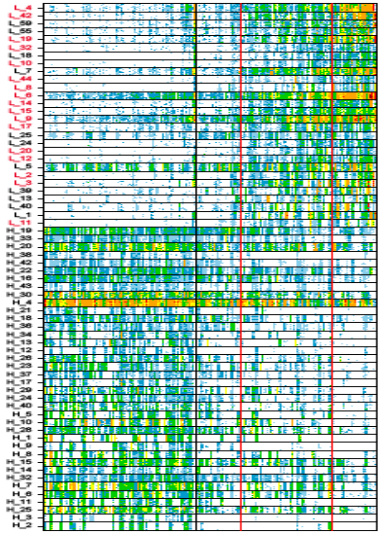
Accurate diagnostics
irrespective of

- disease origin: alcohol, hepatitis viruses
- disease status: compensated & de-compensated (ascites, encephalopathy...),
- medication: antivirals, beta blockers, proton pump inhibitors or none

66 marker species

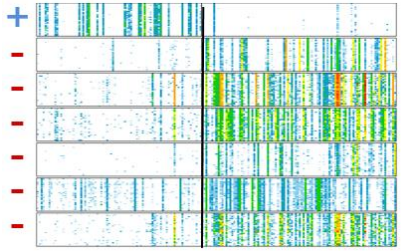
28 LC MGS

38 Healthy MGS



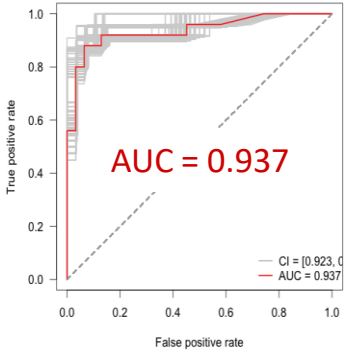
Healthy LC

7 model species



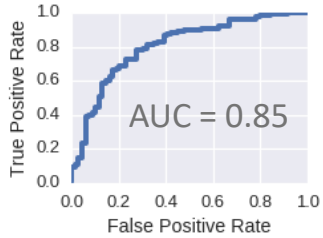
- Additive model
- Genetic algorithm for selection

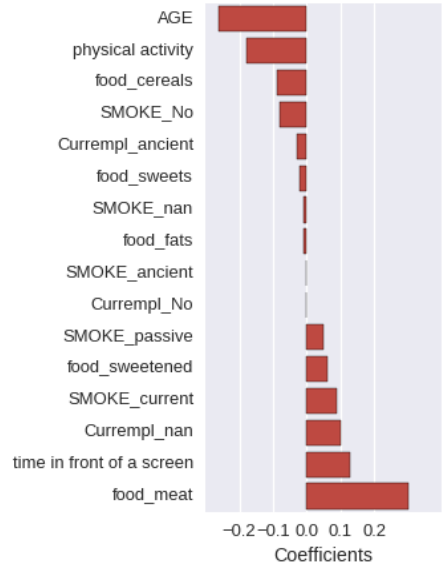
validation (31+25)



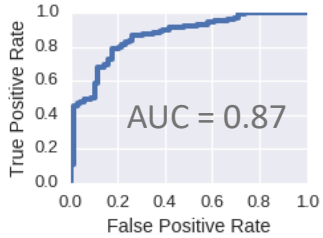
Microbiota-based obesity prediction modeling

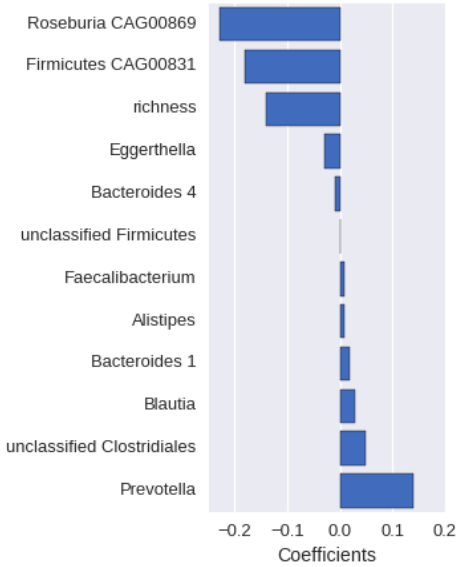
Epidemiological model



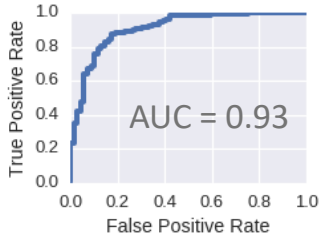


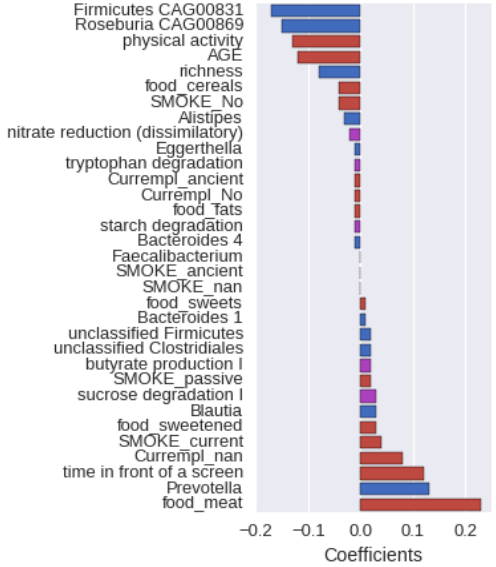
microbiota





microbiota+epidemiology





- LASSO penalized linear regression
- Cross-validation

Epidemiological variables
Microbiota variables
Functional variables

Courtesy of Sébastien Fromentin and Hanna Julienne, unpublished data

metagenopolis mgps.eu

Future

Expanding microbiome knowledge

French microbiomes project

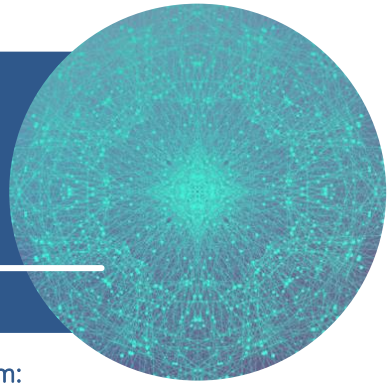
- French citizen microbiome science
- Sequence and analyze the gut microbiome of 100,000 French citizen, from newborns to the elderly, and from healthy individuals to individuals with various diseases
- Accelerate the French gut microbiome knowledge and innovation
- Access to high resolution standardized microbiome data
- Access to a network of world-renowned microbiome experts

Million Microbiome of Humans Project (MMHP)

Officially launched the October 26th, 2019 at the 14th International Conference on Genomics (ICG-14)

- **Analyze 1 million microbial samples** from intestines, mouth, skin, reproductive tract...
- **Draw a microbiome map** of the human body
- **Build the world's largest database** of human microbiome
- **Solid data foundation** for microbiome research
- **Explore the potential of the microbiome** to help people live better lives

MGP participates to this MMHP in bringing 100 000 french gut metagenomes



International microbiome research program:

- Karolinska Institute of Sweden
- Shanghai National Clinical Research Center for Metabolic Diseases in China
- University of Copenhagen, Denmark
- Technical University of Denmark
- MetaGenoPolis at the National Institute for Agricultural and Environment Research (INRAE), France
- Latvian Biomedical Research and Study Centre
- Shenzhen BGI Research

metagenopolis mgps.eu

MGP IT platform evolution

On prem and on cloud orchestration with ActiveEon-ProActive

@MgpsLab

INRAE

Architecture overview

Web Portal and
Integration with
Scientific tools

Quantitative Metagenomics Platform
for gene profiling and statistical analysis

REST / HTTPS

ProActive
Workflows & Scheduling

Secure data transfer

Pre, Post Processing of Data Analysis
Flexibility, Speed of Analysis
Granular execution
Distribution for fast execution
Secure data transfer

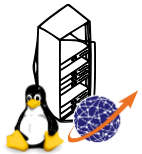
1TB / Sequence
Analysis

On-Prem

Total
DNA

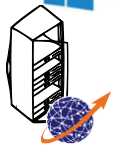
QC/Library preparation
Proton/Illumina
Sequencing

3.3PB
DataBase



Linux
Cluster 2
500 cores

Microsoft Azure



CPU cores on demand
Cloud Bursting

Cloud Bursting : PoC In The Box MS-Azure

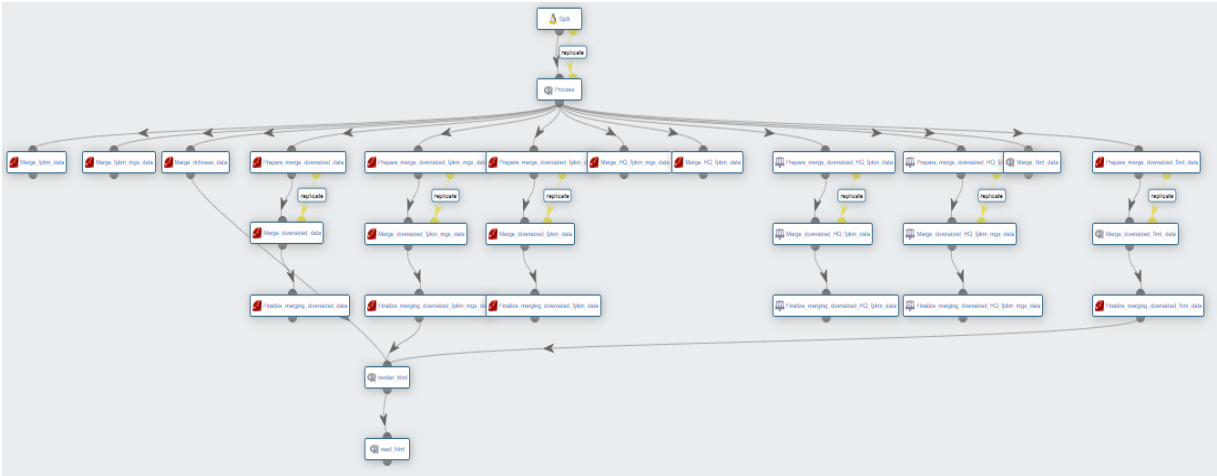
Quantitative Metagenomic workflow deployed throught ActiveEon Proactive

Workload distribution

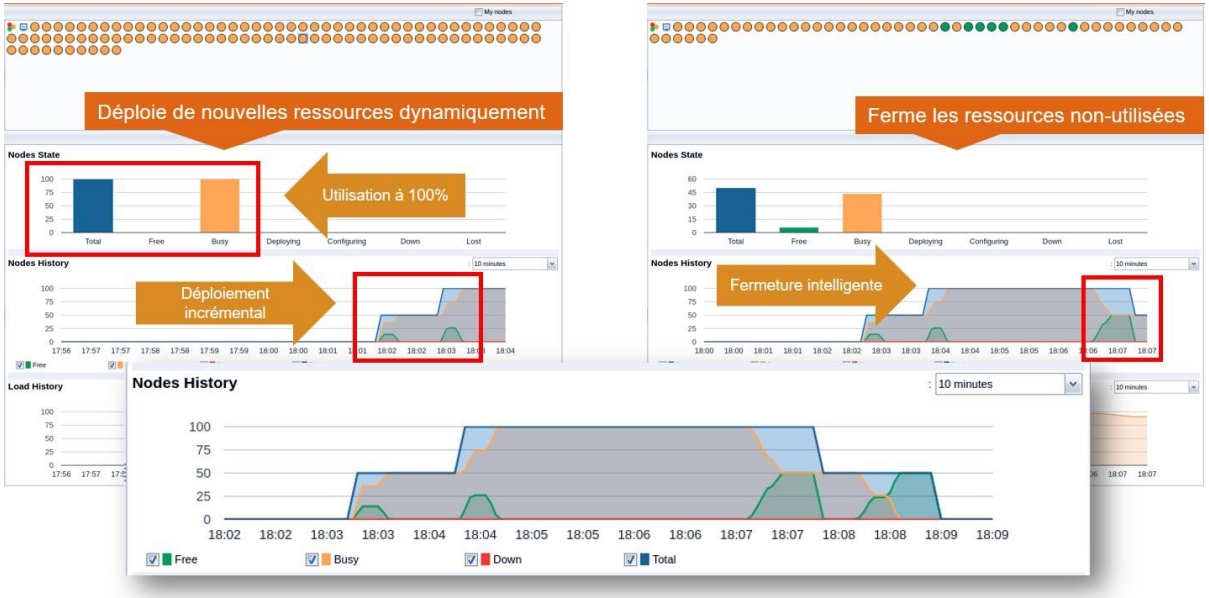
Dynamic and elastic resource pool

Multi-language workflows

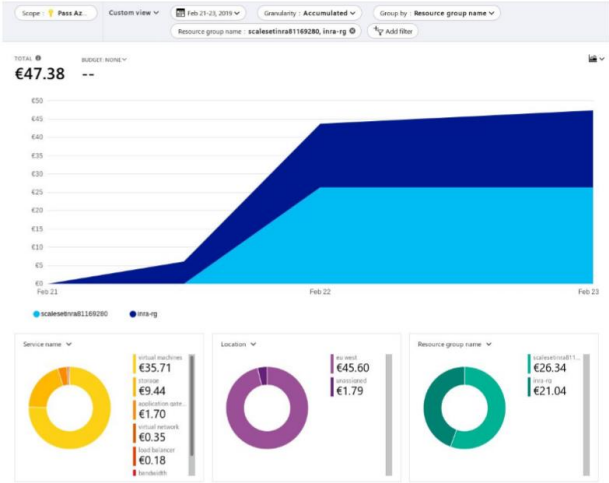
Visualization



Cloud Bursting : PoC In The Box MS-Azure



Dynamic and elastic resource pool



Cost control

Cloud Bursting : PoC In The Box MS-Azure

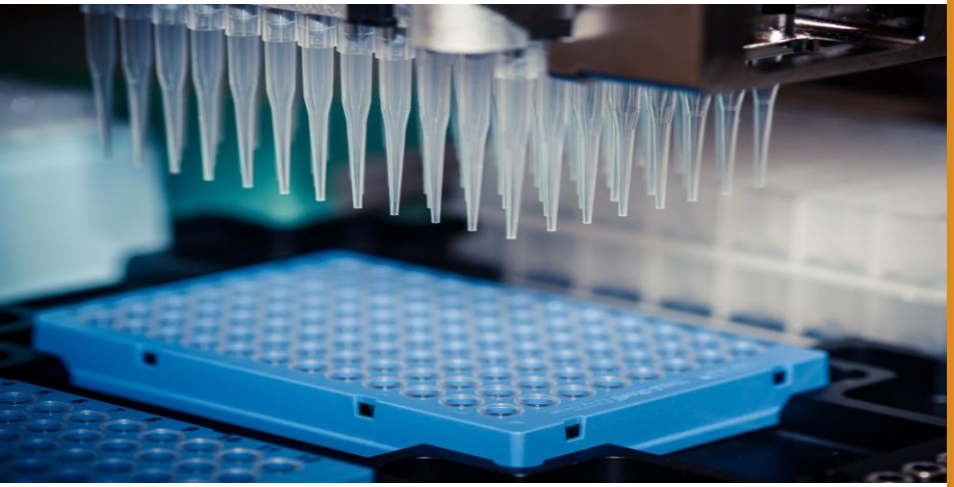
Résultats

- **Extension de capacité interne**
- Possibilité d'aller sur le cloud dans le cadre d'un **PRA** (plan de reprise d'activité)
- **Agilité du processus d'innovation**: itérations plus rapides, accès à une grande diversité de machines, machines récentes (GPU, CPU, …)
- **Définition des coûts** (calcul, stockage, …)

Opportunités

- Facilité le **débordement** sur le cloud en respectant les normes sécurité
- **Collaboration: mise à disposition d'une plateforme d'analyse du microbiote** à destination de cliniciens et industriels
- **Développement d'une offre machine learning** avec de nouvelles ressources de calcul disponibles à la demande

Next challenges



- What is a healthy, a normal microbiome ?
- Next Generation Big Data analytics
- Bringing Deep Learning algorithm in microbiome
- Integrating ethics for AI algorithm explanation
- Integrating ethics for data collecting, curation, sharing and dissemination
- **Data regulation : GDPR, HDS** *(vs Cloud Act)*
  Exploring European provider (OVHCloud)

Thanks

Business developer
& strategic partnership manager:
karine.valeille@inrae.fr

Communication manager:
anne-sophie.alvarez@inrae.fr