



**HAL**  
open science

## Interagir avec ses données de spectrométrie de masse sous Galaxy

Julien Saint-Vanne, Lain Pavot, Ronan Cariou, Sébastien Hutinet, Farid  
Ichou, Yann Guitton, Franck Giacomoni

► **To cite this version:**

Julien Saint-Vanne, Lain Pavot, Ronan Cariou, Sébastien Hutinet, Farid Ichou, et al.. Interagir avec ses données de spectrométrie de masse sous Galaxy. 15èmes journées scientifiques du RFMF, May 2023, Perpignan, France. 2023. hal-04321401

**HAL Id: hal-04321401**

**<https://hal.science/hal-04321401>**

Submitted on 4 Dec 2023

**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.

# Interagir avec ses données de spectrométrie de masse sous Galaxy

Julien Saint-Vanne<sup>1</sup>, Lain Pavot<sup>2</sup>, Ronan Cariou<sup>1</sup>, Sébastien Hutinet<sup>1</sup>, Farid Ichou<sup>3</sup>, Yann Guitton<sup>1</sup>, Franck Giacomoni<sup>2</sup>



1 : ONIRIS, INRAE, LABERCA, 44300 Nantes

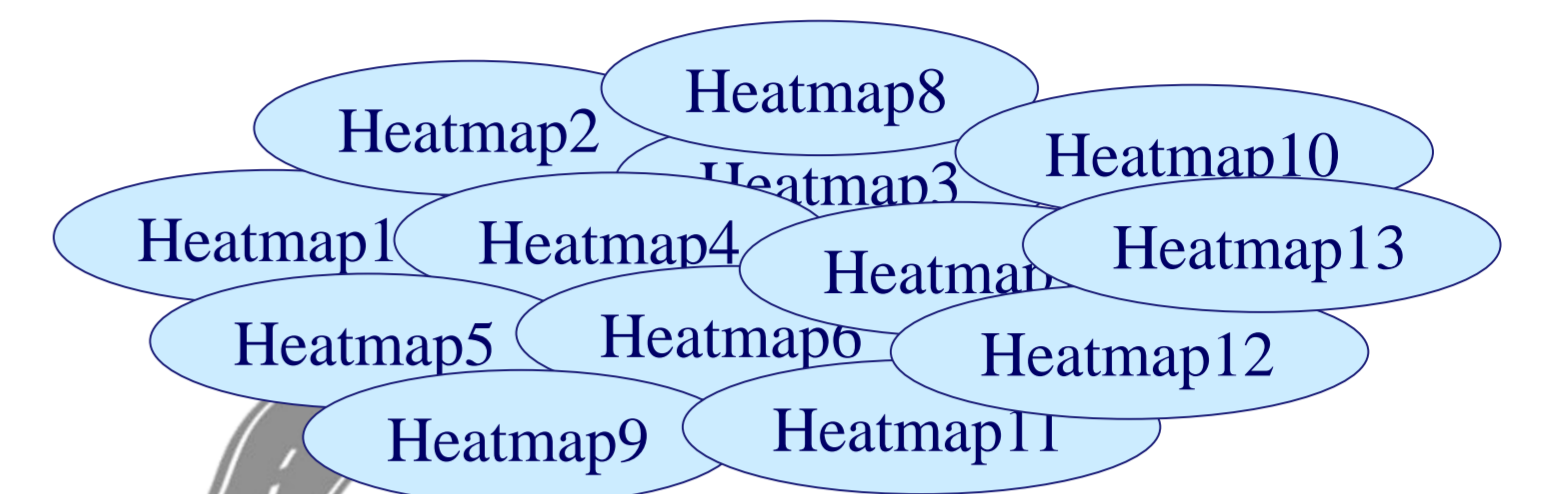
2 : Université Clermont Auvergne, INRAE, UNH, Plateforme d'Exploration du Métabolisme, MetaboHUB Clermont, Clermont-Ferrand, France

3 : ICAN Omics, Foundation for Innovation in Cardiométabolism and Nutrition (ICAN), Hôpital Pitié-Salpêtrière, F-75013 Paris, France

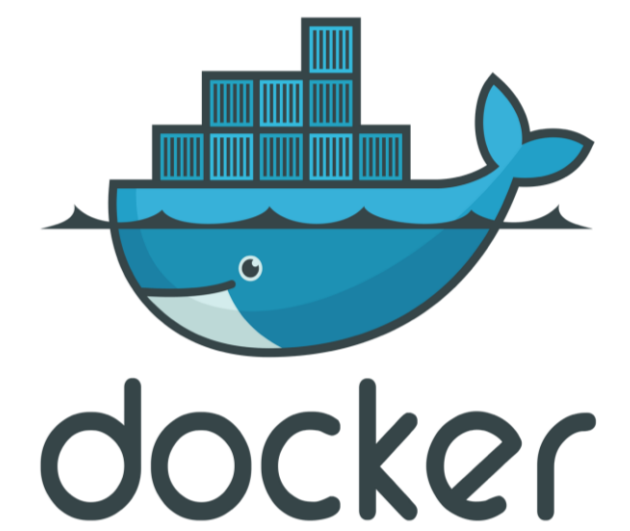
Galaxy est une plate forme open source pour l'analyse et le partage de données scientifiques. Elle fournit une approche sans frontières et efficace pour aider les utilisateurs dans leurs analyses. L'interface utilisateur Web conviviale de Galaxy fait le lien entre les outils en ligne de commande utilisés dans la recherche avec de nombreuses données et les chercheurs n'utilisant pas le code informatique. Il est possible d'appeler des outils conteneurisés (via « docker ») et des environnements interactifs au travers de l'interface Galaxy. Cela va ouvrir de nombreuses possibilités et facilités pour tous.



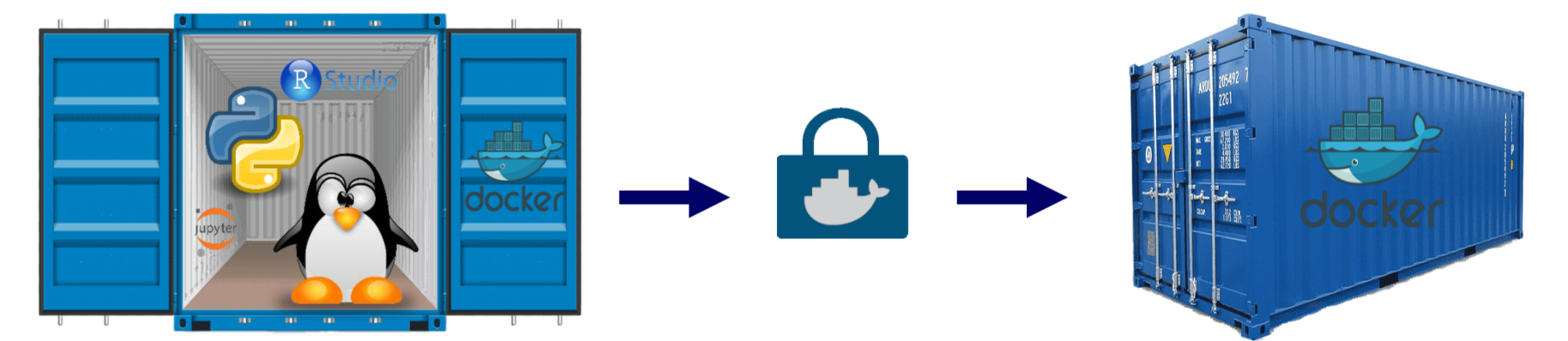
## Exemple de l'outil heatmap



- 1 paramètre => 1 job à lancer
- Chronophage
- Accumulation de données dans l'historique Galaxy
- Une image fixe en sortie, sans interaction possible

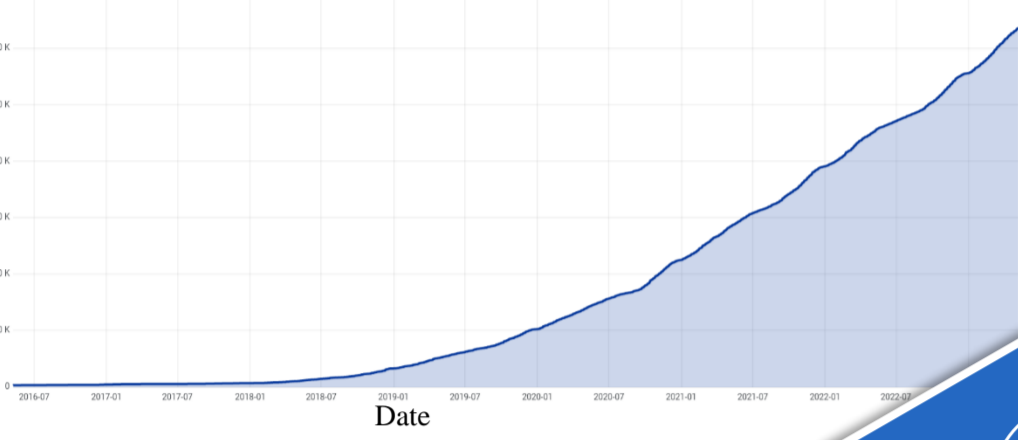


Un conteneur contient un environnement qui permet d'héberger les scripts, les outils ou encore les bibliothèques dont on a besoin. Ces conteneurs respectent les principes FAIR (Facile à trouver, Accessible, Interopérable et Réutilisable) et permettent une importante interactivité avec les données, grâce aux nombreux types de langages et d'applications pouvant être ajoutés.



## Quelques chiffres sur Galaxy Europe

### Evolution du nombre d'utilisateurs

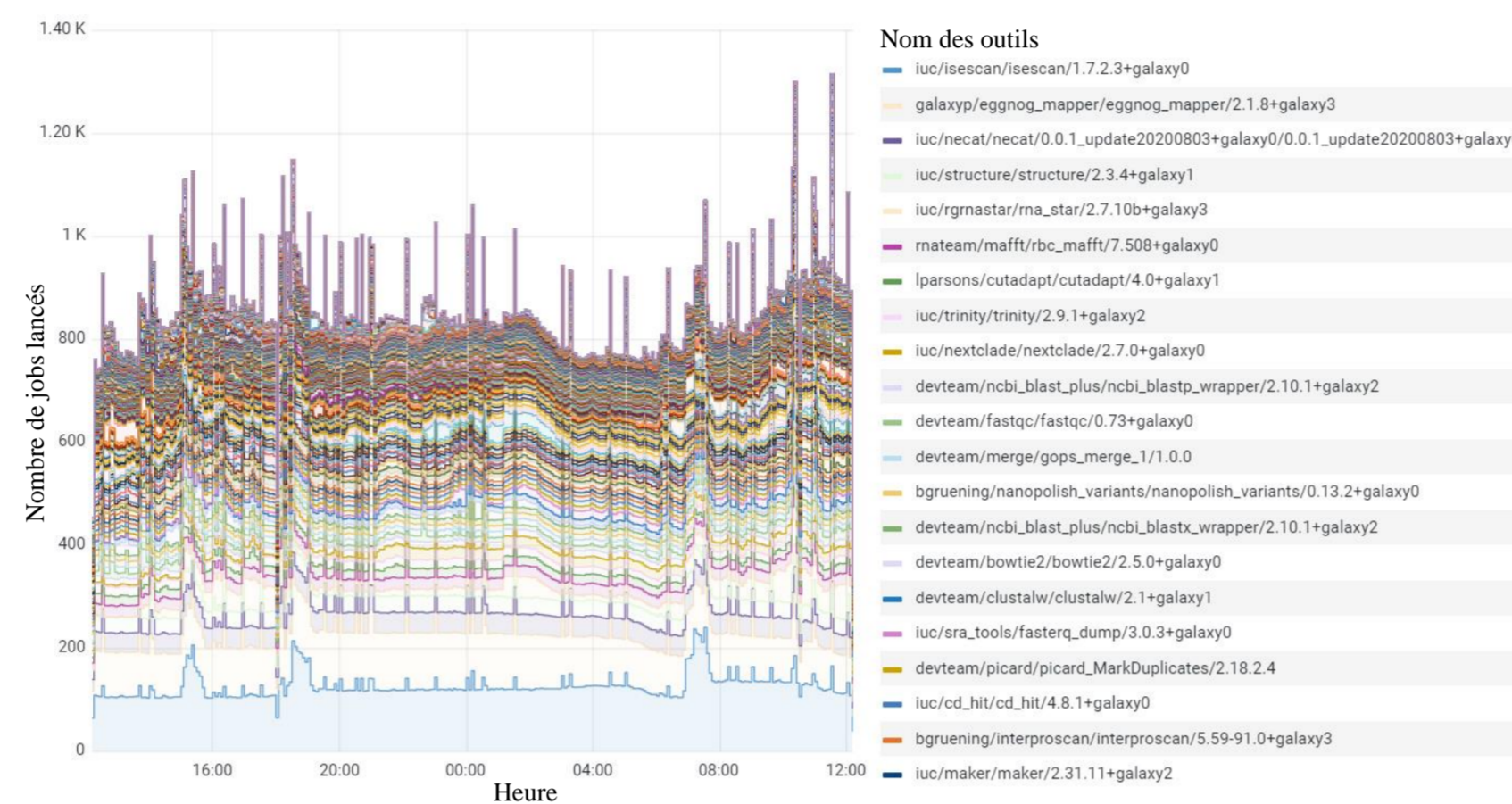


- + 100 pays
- + 180.000 workflows/mois
- ~ 70.000 utilisateurs
- + 1 million de tâches/mois
- + 14.000 stagiaires
- + 100 millions de jeux de données
- + 8.000 outils
- Thousands of maintainers

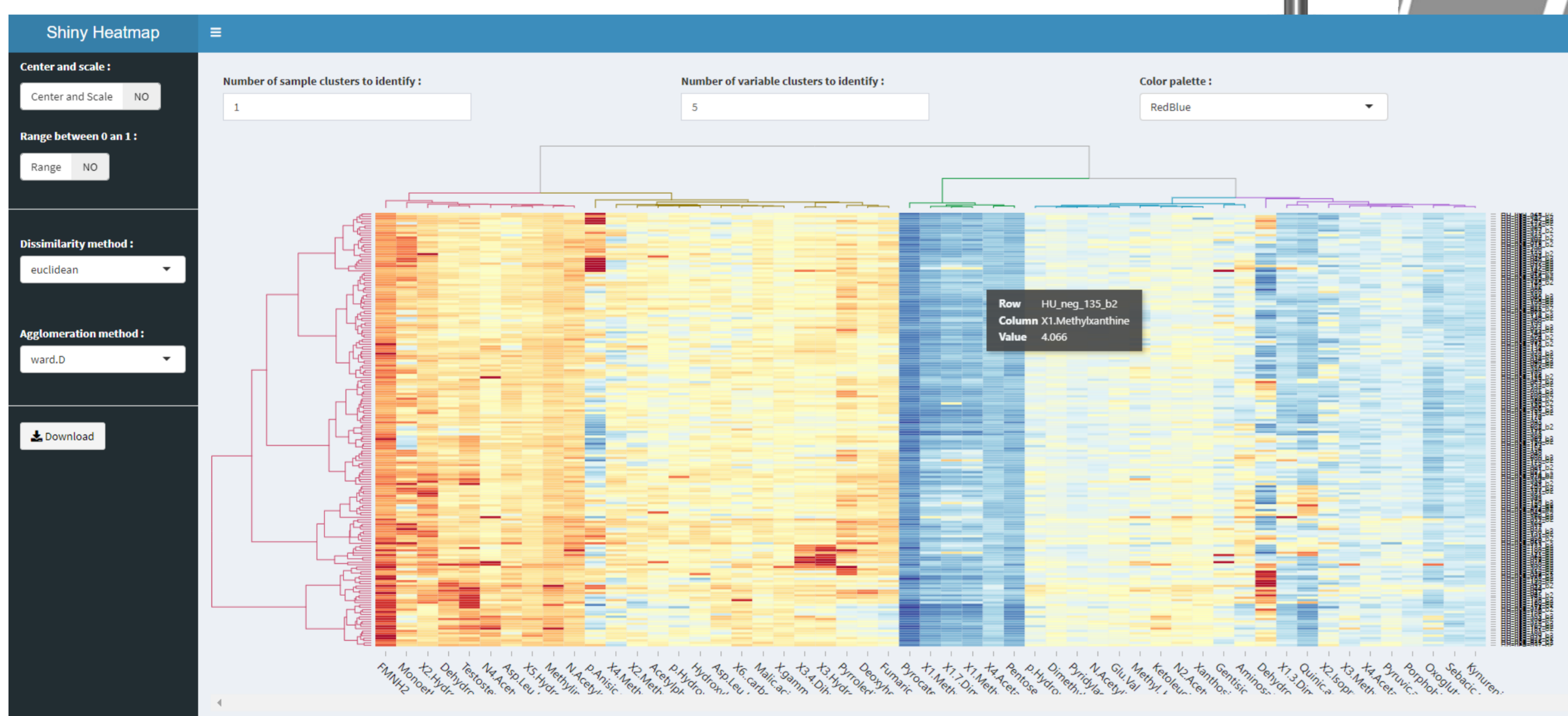
### Nombre de sessions d'entraînement à Galaxy



### Les jobs en cours sur Galaxy

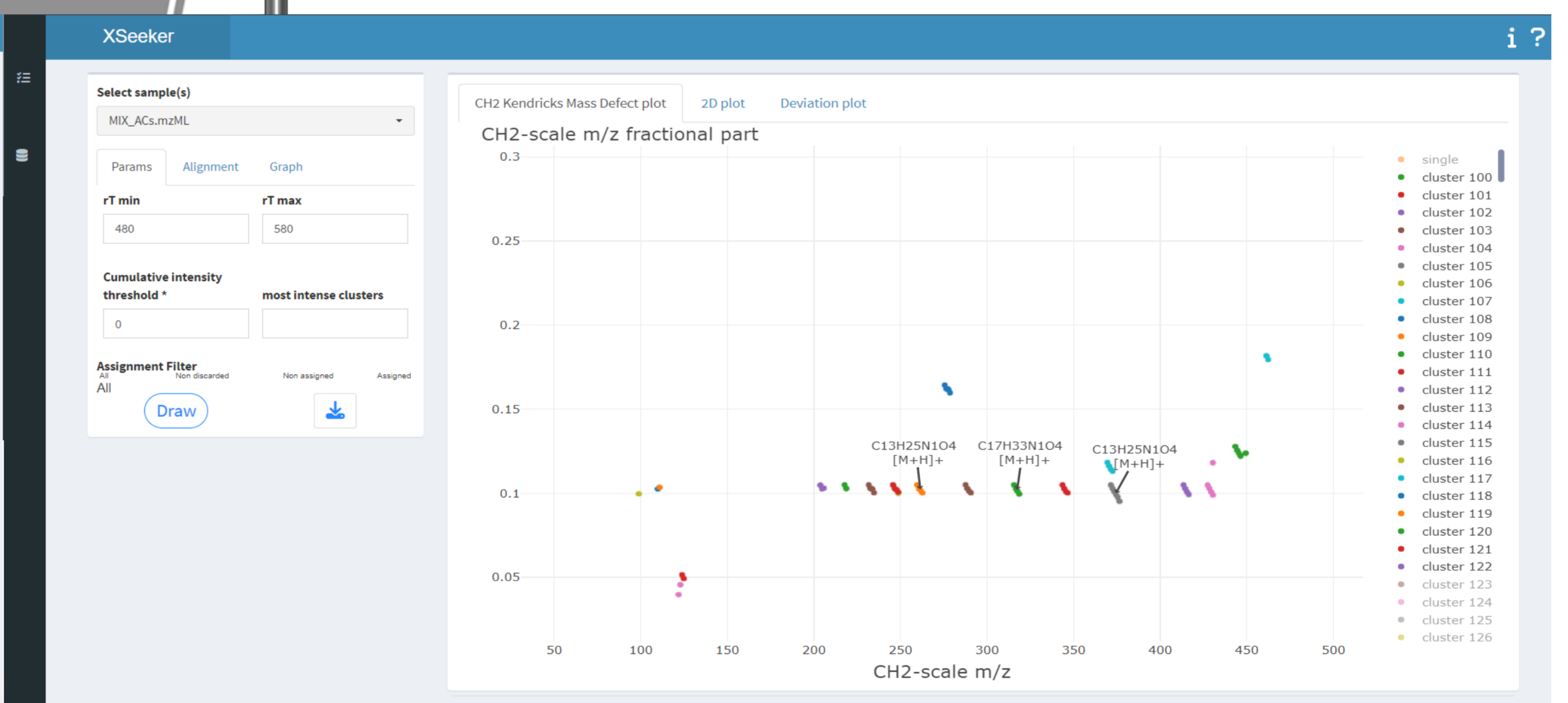


## Outil de visualisation interactive



- A partir du fichier « datamatrix » de l'historique Galaxy
- Changement des paramètres et des couleurs en temps réel (ou presque)
- Visualisation interactive (zoom, sélection de ligne, affichage des valeurs au survol)
- Export possible du graphique une fois satisfait (sur l'ordinateur directement en image)

## Outil d'annotation interactive



- A partir des fichiers de peak-picking de l'historique Galaxy
- Annotation facile par l'utilisateur et directement sur le graphique
- Interactivité très forte avec les données, modification des graphiques
- Export contenant les nouvelles annotations, directement dans l'historique Galaxy

**FAIR**

- FACILE A TROUVER**: Outil avec un identifiant global, unique et persistant pour chaque version (PIDs, DOIs). Outil richement documenté et décrit. Outil et documentation enregistrés ensemble.
- ACCESSIBLE**: Accès à l'outil défini et pour toujours. Accès facile à l'outil (web, doi, etc...). Accès protégé = description poussée. Accès même après indisponibilité de l'outil.
- INTEROPERABLE**: Outil en langage courant, accessible et partagé. Données utilisées ou produites qui suivent le principe FAIR. Liens vers d'autres sources documentés.
- REUTILISABLE**: Description précise et pertinente de l'outil. Licences indépendantes, claires et accessibles pour l'outil. Métadonnées et documentation conformes aux normes et applicables.

Lamprecht, Anna-Lena et al. 'Towards FAIR Principles for Research Software'. 1 Jan. 2020 : 37 - 59.

Galaxy met l'accent sur la transparence, la reproductibilité et la réutilisabilité, facilitant le partage (d'outils, de flux de travail, de données ou encore de visualisations). Afin d'améliorer la réutilisation des données de la recherche, les outils utilisés dans Galaxy s'inscrivent donc dans les principes FAIR. L'interactivité et la visualisation dynamique sont les deux principales actions demandées par les analystes et mises en avant dans les outils souvent payants. La mise en place d'outils hébergés dans des conteneurs permet à l'utilisateur d'avoir un maximum d'interactivité avec celles-ci, en conservant la facilité d'utilisation du service qu'est Galaxy.