



HAL
open science

Déploiement de la plateforme de traitement des données phénotypage haut débit 4P sur l'infrastructure France Grilles.

Vincent Nègre, Eric David, Philippe Burger, Romain Chapuis, Boris Adam, Anne Tireau, Patrick Moreau, Antony Tong, Gallian Colombeau, Samuel Thomas, et al.

► To cite this version:

Vincent Nègre, Eric David, Philippe Burger, Romain Chapuis, Boris Adam, et al.. Déploiement de la plateforme de traitement des données phénotypage haut débit 4P sur l'infrastructure France Grilles.. JRES (Journées réseaux de l'enseignement et de la recherche) 2019, Dec 2019, Dijon, France. 2019. hal-04807042

HAL Id: hal-04807042

<https://hal.science/hal-04807042v1>

Submitted on 27 Nov 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 4.0 International License

Déploiement de la plateforme de traitement des données phénotypage haut débit 4P sur l'infrastructure France Grilles



Vincent Nègre¹, Eric David², Philippe Burger¹, Romain Chapuis¹, Boris Adam¹, Anne Tireau¹, Patrick Moreau¹, Anthony Tong², Gallian Colombeau³, Samuel Thomas⁴, Pascal Neveu¹, Jérôme Pansanel⁵, Frédéric Baret¹, Marie Weiss¹.

1: INRA; 2: EPHESIA CONSULT; 3: HIPHEN; 4: ARVALIS; 5: CNRS

I - LE PHÉNOTYPAGE DES PLANTES À HAUT DÉBIT

- Le phénotypage consiste à observer les caractères ou **traits** qui caractérisent la structure ou le fonctionnement d'un individu (les plantes dans notre cas).
- Le suivi de ces traits permet aux agronomes d'améliorer la sélection variétale en identifiant par exemple des variétés plus adaptées à un environnement donné (climat, sol, pathogènes, ...).
- Afin d'établir des corrélations robustes, il est nécessaire de considérer le phénotypage haut-débit, c'est-à-dire de suivre un grand nombre d'individus (plusieurs centaines voire plusieurs milliers), parfois dans différentes modalités (introduction de pathogènes ou non, différentes conditions d'irrigation ou d'apport azoté) et de mesurer un grand nombre de traits (jusqu'à plusieurs dizaines).
- Au champ, la dynamique de développement de milliers de micro parcelles (variétés/modalités) est caractérisée tout au cours du cycle de la culture.



Vue aérienne d'une expérimentation de phénotypage haut-débit.

II - QUE MESURE-T-ON ET AVEC QUELS CAPTEURS?

Capteur		Utilisations
Camera haute résolution/RGB	Fraction verte 	<ul style="list-style-type: none"> Taux de couverture verte Indice foliaire Efficacité d'interception PAR Comptage de plantes Enherbement Symptômes de maladies Risque de verse
LIDAR Mesure de distance et profondeur (LIDAR-stéréo)		<ul style="list-style-type: none"> Hauteur du couvert Port foliaire Structure du couvert Expérimentation
Caméra IR thermique		<ul style="list-style-type: none"> Température de surface Stress hydrique Pilotage de l'irrigation
Caméra multi spectrale		<ul style="list-style-type: none"> Indice foliaire Contenu en chlorophylle Symptômes maladies Pilotage de l'azote
Spectromètre		<ul style="list-style-type: none"> Chlorophylle/statut azoté Stress (hydrique, ...) Indice foliaire Composition biochimique

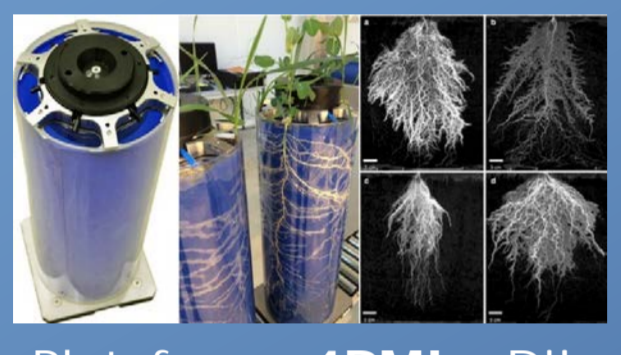
III - DISPOSITIFS DE PHÉNOTYPAGE À HAUT DÉBIT



Plateforme M3P – Montpellier



Phenomobile et Drone – Toulouse, Montpellier



Plateforme 4PMI - Dijon



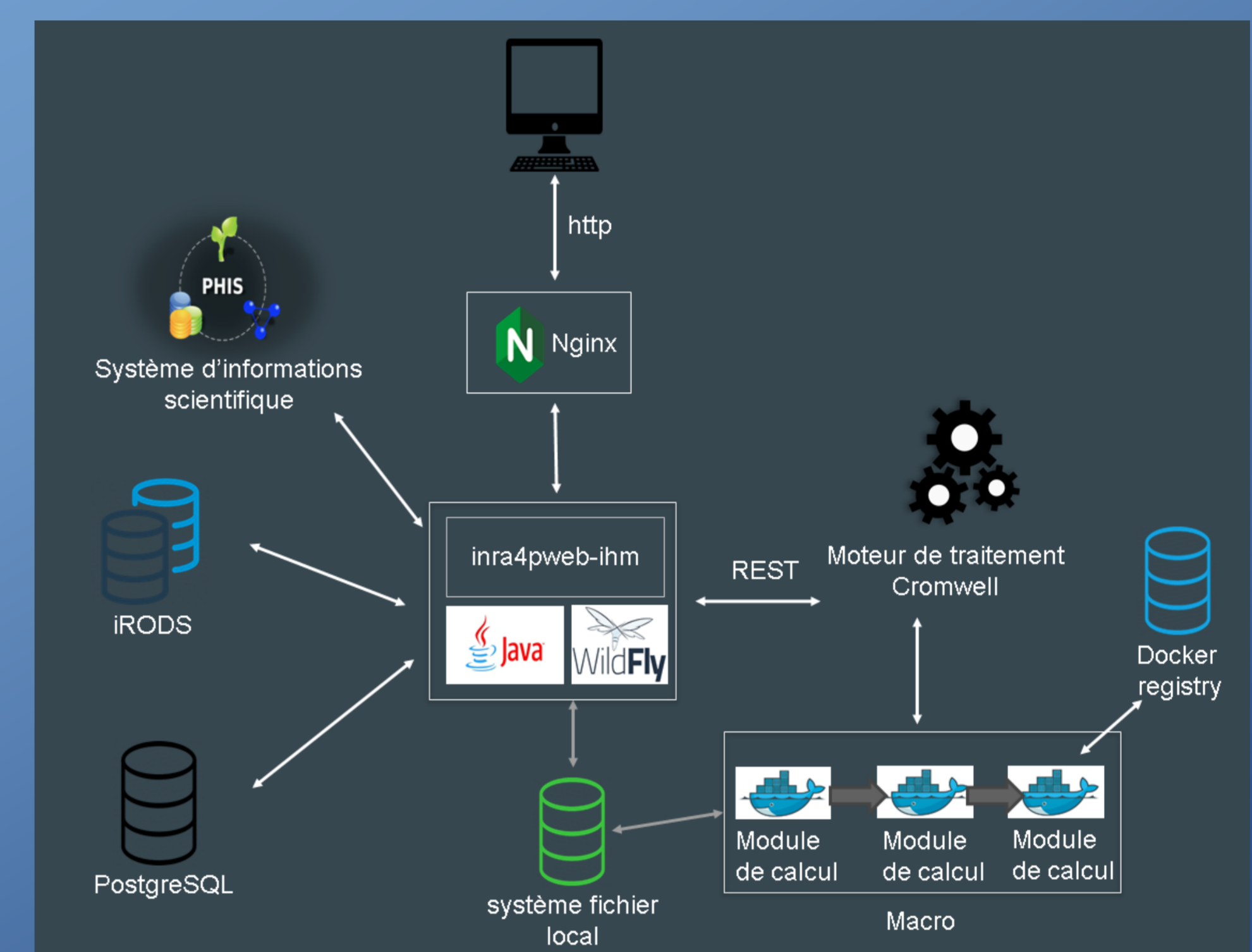
Plateforme Pheno3C – Clermont-Ferrand

Conditions contrôlées

Conditions semi-contrôlées ou au champ



IV - ARCHITECTURE LOGICIELLE



V- CRÉATION D'UN MODULE

- Un module correspond à une étape d'un traitement.
- Un module de calcul peut être écrit dans n'importe quel langage (Java, C, C++, Python, etc.) mais doit suivre les conditions suivantes:
 - être embarqué dans une image Docker incluant tout l'environnement d'exécution ;
 - exécutable en ligne de commande sans intervention de l'utilisateur ;
 - les paramètres (déclarés dans 4P) doivent être fournis en arguments.

```
CLI Example: docker run -v <host>:<mount_point> -w <mount_point> <module_name>:<tag> [arguments]
```

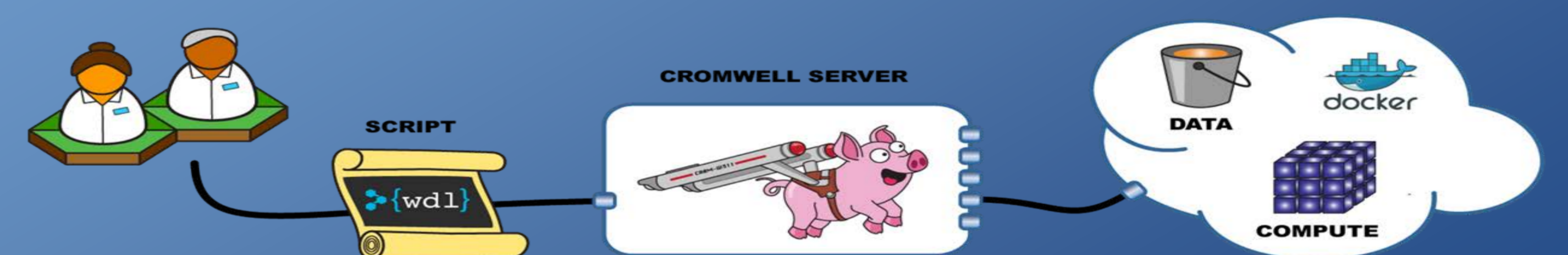
- Un module doit être publié sur un dépôt git (Gitlab, Bitbucket, Github, etc.) et accessible en lecture depuis l'application 4P.



- Intégration continue avec Jenkins et Apache Groovy

VI - MOTEUR DE TRAITEMENT CROMWELL

- Un traitement (ou macro) correspond à l'enchaînement d'un ou plusieurs modules.
- Le moteur d'exécution **CromWell** développé par le Broad Institute assure la gestion des traitements.
- Lorsque l'utilisateur crée une macro, l'application 4P génère automatiquement un fichier WDL (Workflow Description Language) décrivant les modules et leur enchaînement ainsi qu'un fichier JSON décrivant les paramètres d'entrées.
- Cromwell lance l'exécution des traitements sur le cluster Docker Swarm.



<https://software.broadinstitute.org/wdl/>

VII - DÉPLOIEMENT SUR FRANCE GRILLES

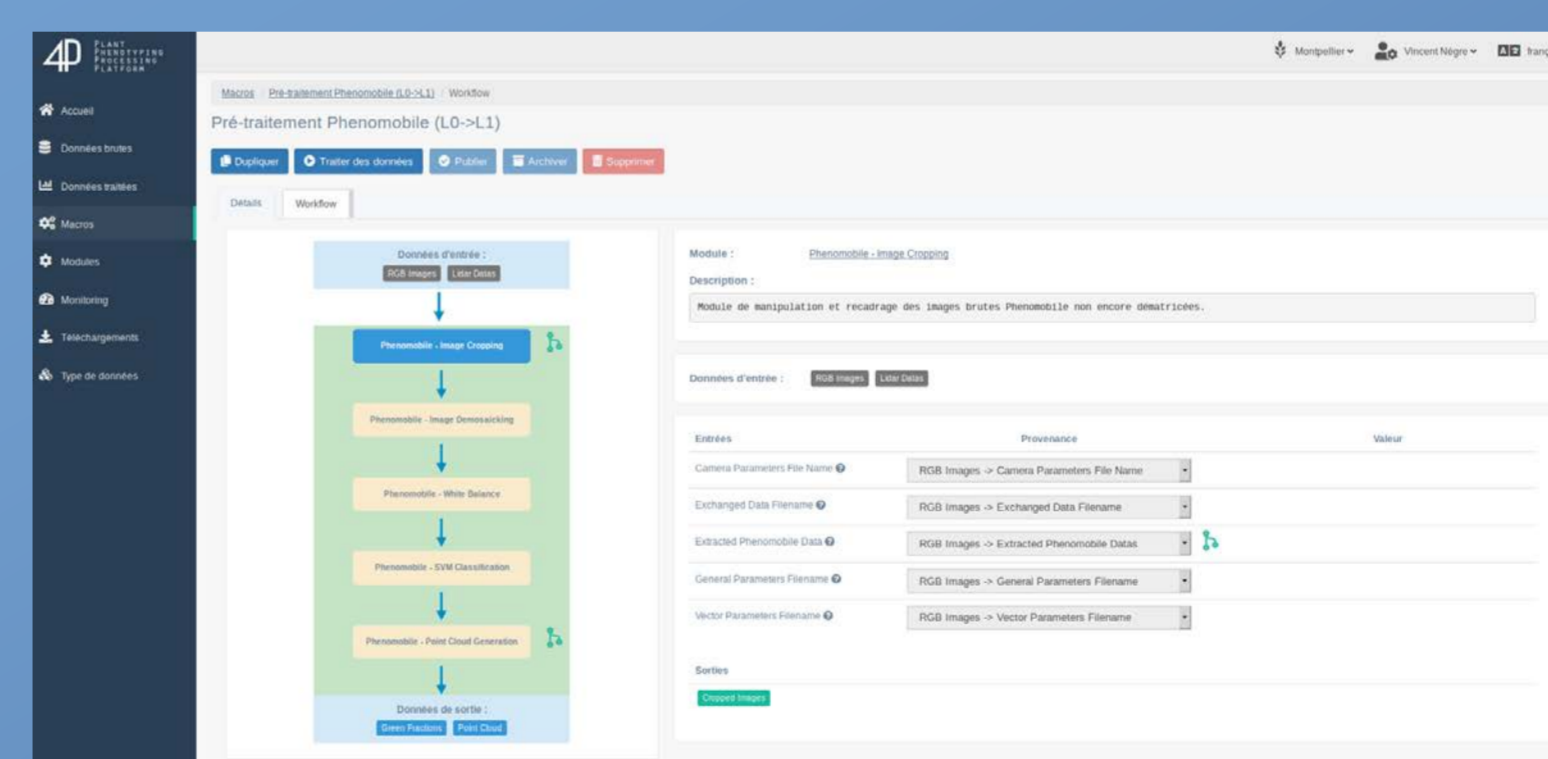
CALCUL

- La plateforme 4P est déployée sur l'infrastructure Cloud opérée par France Grilles à travers son service **FG-Cloud**.
- Le service FG-Cloud est basé sur la technologie **OpenStack**.
- Nous utilisons six machines virtuelles (CentOS 7) afin de répartir la charge et les services.

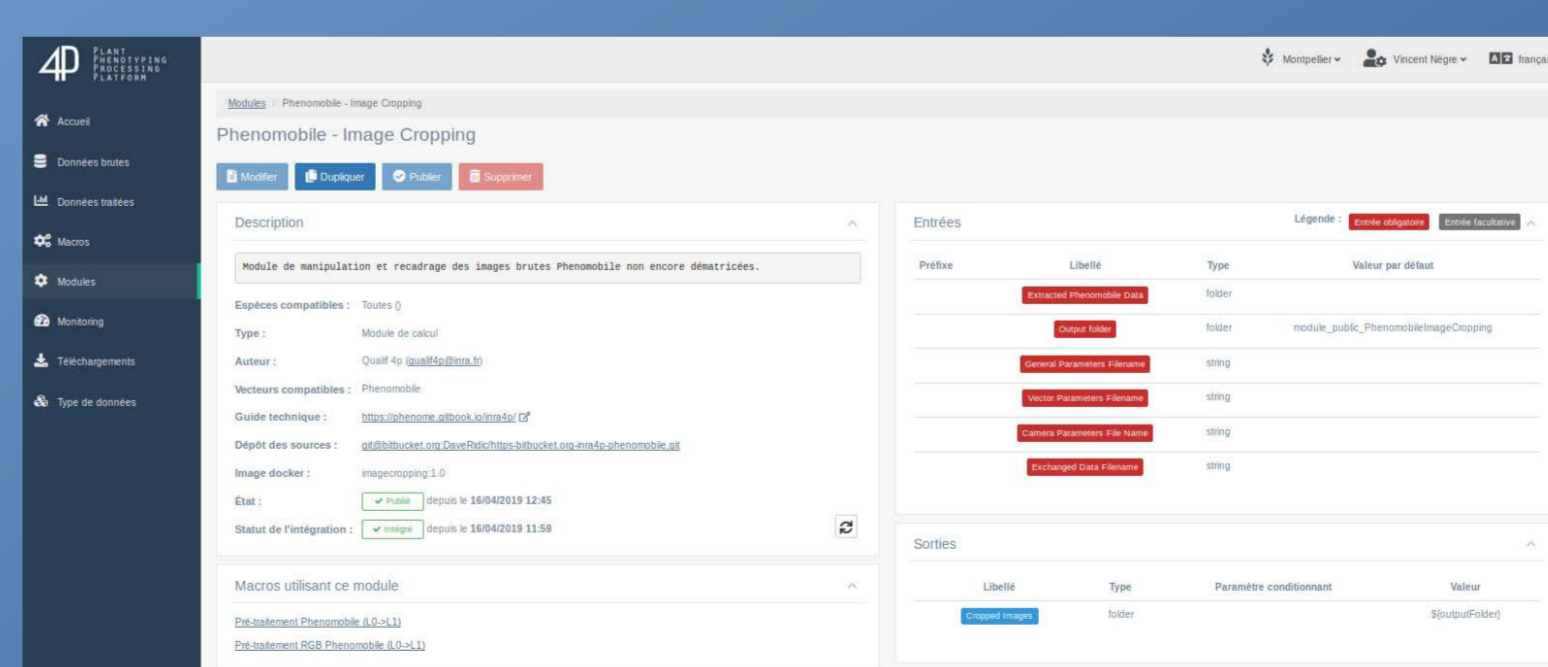
STOCKAGE

- Les données brutes sont stockées sur le système **iRODS** opéré par l'infrastructure France Grilles à travers son service de gestion de données **FG-IRODS**.
- L'utilisateur peut transférer des fichiers allant jusqu'à plusieurs dizaines de Go via le protocole http et l'API W3C file.

VIII - INTERFACES UTILISATEUR



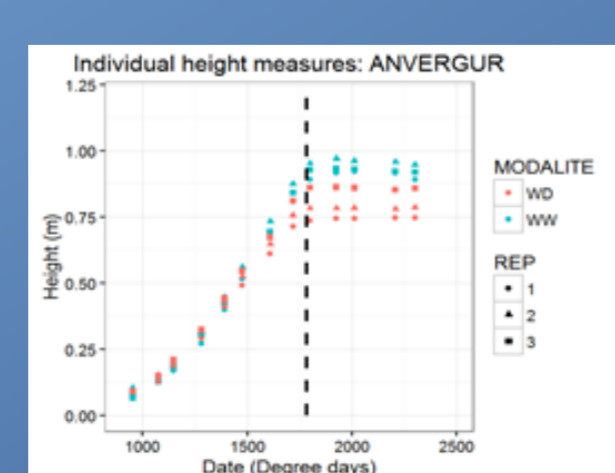
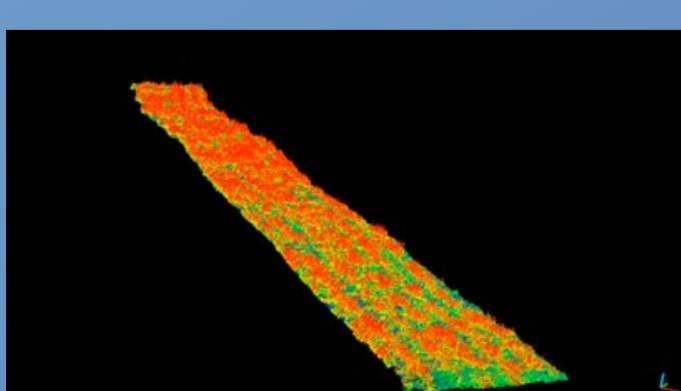
Interface de gestion des macros



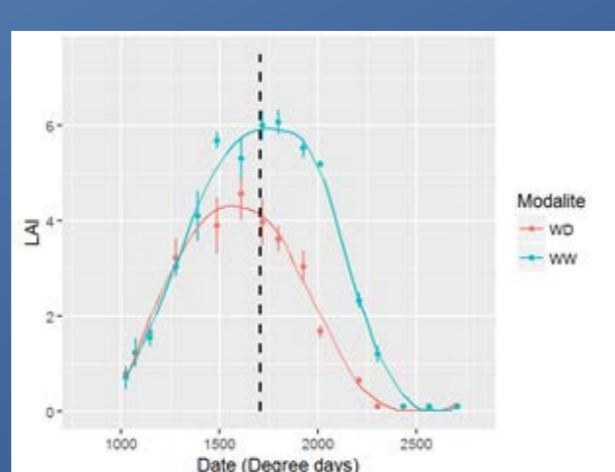
Interface de gestion des modules

IX- EXEMPLES D'UTILISATIONS

- Analyse de données LIDAR pour le suivi de la hauteur de couvert (S. Madec, S. Thomas).



- Analyse de données issues d'une caméra multi spectrale pour le suivi spatial et temporel de l'indice foliaire (surface de feuilles sur 1m² de sol).



X - PARTENAIRES

