SIWAA, le site web Galaxy pour la simulation & l'analyse mathématique des agroécosystèmes Séminaire Interne MIAT, équipe RECORD













Présentation

de la présentation de SIWAA

- L'objet de la présentation
- Motivations passées et actuelles
- Le service au travers de l'interface
- L'architecture système et son administration
- Outils et développements
- Animation
- Perspectives











SIWAA

Qu'est ce que c'est?

- Un site web https://siwaa.toulouse.inrae.fr/
- Administré et développé par l'équipe RECORD
- Reposant sur Galaxy
- Mobilisant des ressources informatiques
 - INRAE
 - Muse
- Pour faciliter le déploiement et l'utilisation d'expérimentation numériques pouvant nécessiter des ressources de calcul importantes.



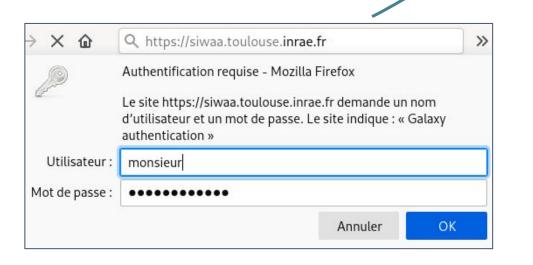


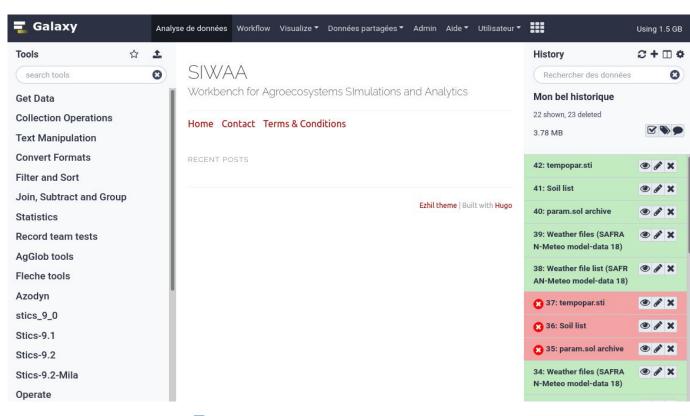






Comment ça marche









Un retour d'expérience sur le déploiement de l'instance Galaxy SIWAA séminaire MIAT du vendredi 7 Mai 2021 / RECORD / Patrick Chabrier



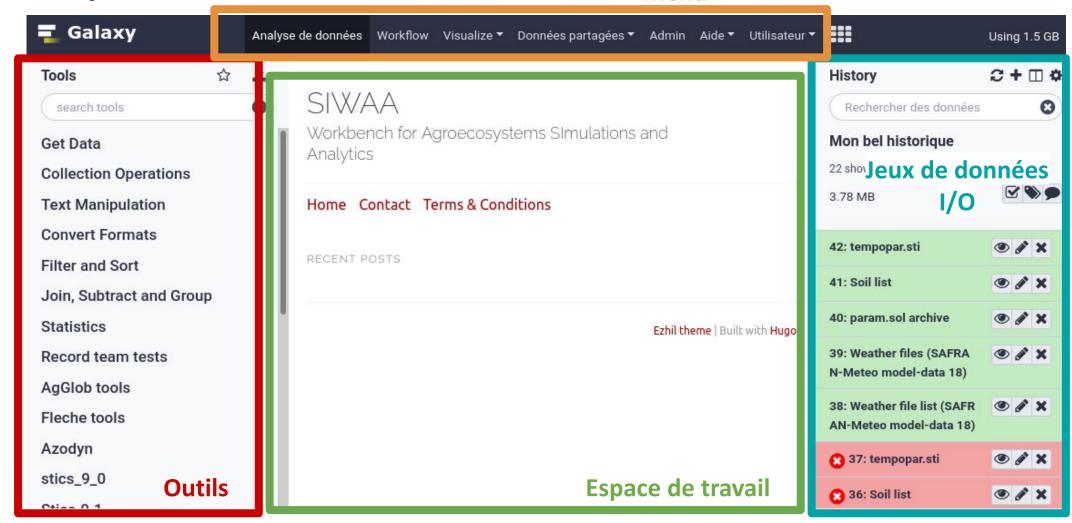






Comment ça marche

Menu



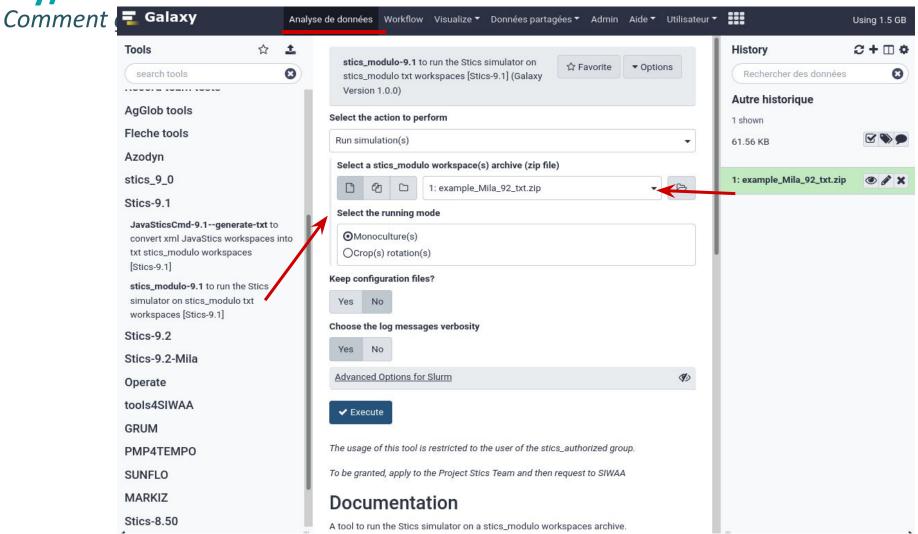












département MathNum



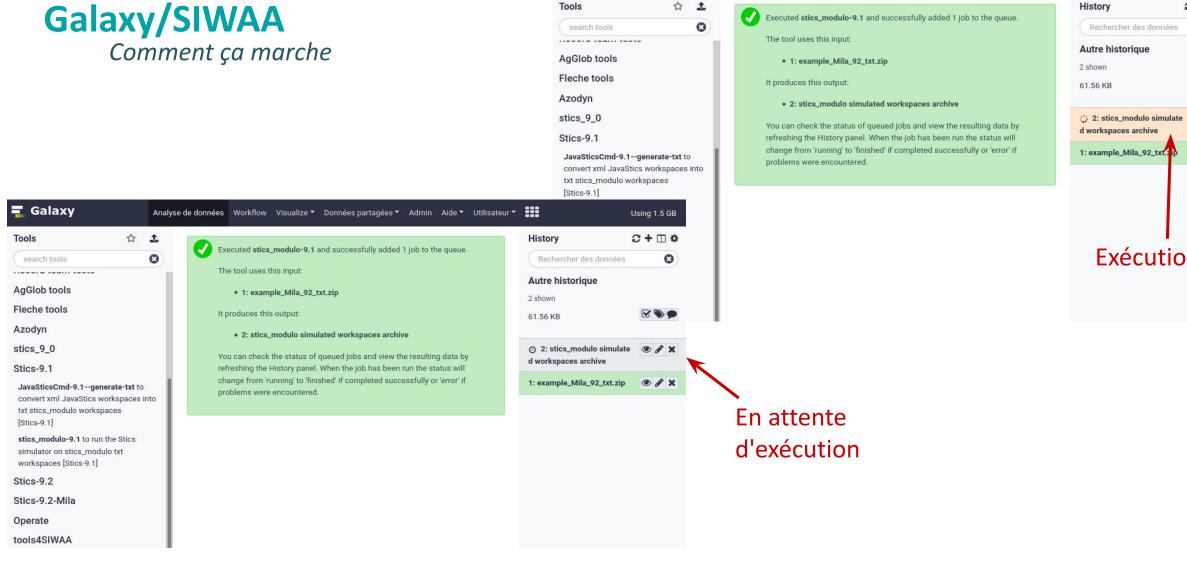












■ Galaxy













Analyse de données Workflow Visualize ▼ Données partagées ▼ Admin Aide ▼ Utilisateur ▼



Using 1.5 GB

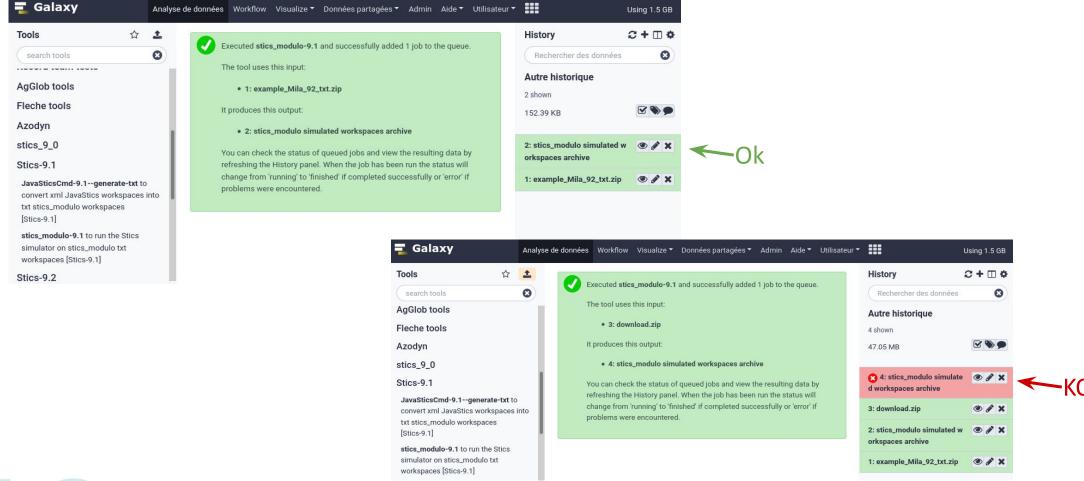
2+ 11 4

Rechercher des données

Exécution

0

Comment ça marche







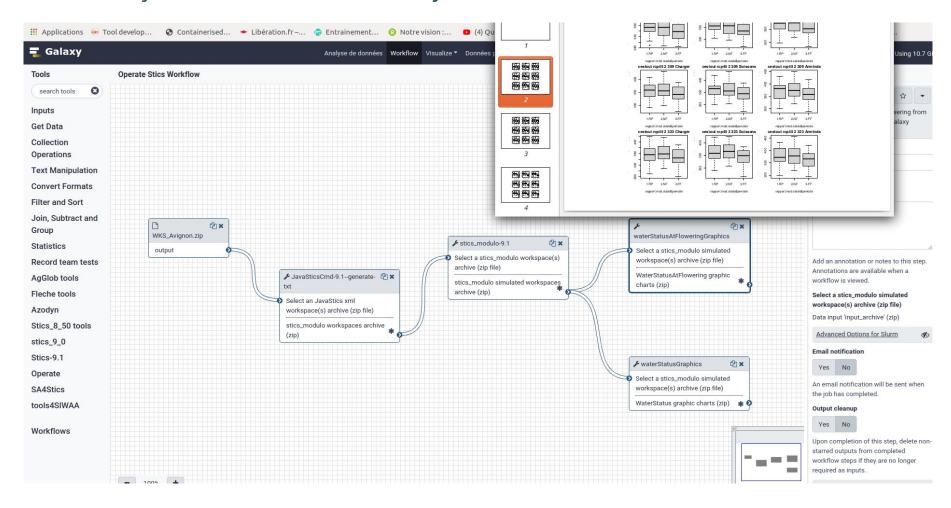








Comment ça marche...notion de workflow





INRAe











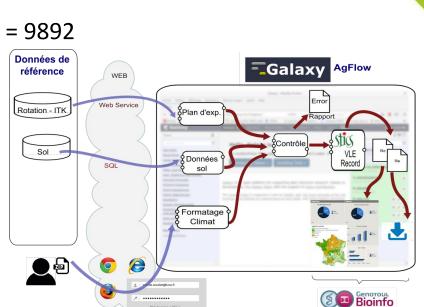
SIWAA

Pourquoi? ... Le projet Agglob, InfoSol, AgroClim, ODR

- Demandes récurrentes pour des plan de simulation massifs
- 2 projets emblématiques EFESE et 4p1000
- Plans de simulation sur des périodes longues 30 ans, pas de temps journaliers
- Echelle France, maille Safran. 8x8km = 9892
- Beaucoup de partenaires
- Source de données distribuées
- n x 10Kh de calcul

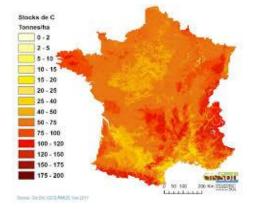
Vers un changement de paradigme :

Reproductibilité - Open Science - FAIR



















SIWAA

Pour qui? Pas pour tout le monde.

- Travailler sans coder / sans rien installer
- Ne pas mobiliser sa machine pour les calculs
- Interface frustre
- Projet Scientifiques
- Agroécosystèmes....mais pas de façon exclusive

Pour l'équipe:

- Mobiliser toutes les compétences, projet structurant (admin, web, dev, math)
- Nous rendre service aussi











Galaxy *Pourquoi?*

- Accessibilité
 - o interface web, pas de code pour l'utilisateur
 - capacité d'intégration des logiciels
- Flexibilité dans l'interaction aux ressources de calcul
- Partage / Reproductibilité / Transparence
 - Tout est tracé et peut être rejoué
 - Les outils et leurs dépendances sont versionnés
 - Les utilisateurs peuvent échanger leurs workflows / leur données
- Communauté très active (tchat, formation)
- Communauté locale & INRAE Toulouse, Sigenae, ChemFlow, W4M
- Gestion fine des droits d'accès
- Open source, gratuit











Galaxy C'est quoi?

- Un framework python
- Un gestionnaire de contenu orienté expérimentation scientifique
- Une solution relativement clé en main
- Un domaine d'application d'origine la bioinformatique.
- D'autres domaines d'application :
 - Ecologie
 - Métabolomique
 - Chimiométrie
 - Climatologie





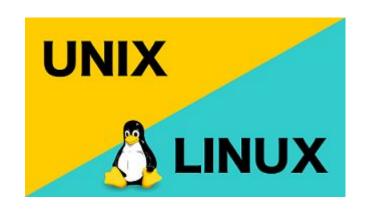






Galaxy/SIWAA *Qu'est ce qu'un outil ?*

Un logiciel utilisable en ligne de commande sur un OS Unix.











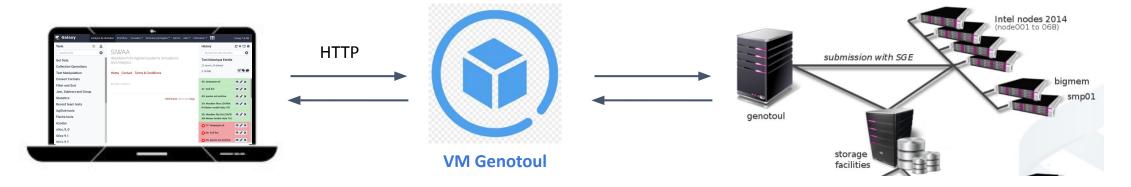




SIWAA

C'est quoi? Hier, Aujourd'hui, Demain (1)

- 2017/2018
- 1ère instance "clés en main" grâce à la plateforme Genotoul bioinfo
- Tester Galaxy sans se plonger dans son administration
 - (au passage, merci à l'équipe)
- Tests concluants













SIWAA

C'est quoi? Hier, Aujourd'hui, Demain (2)

- 2018
- Recherche d'un hébergement pour notre instance Galaxy (VM + Calcul)
- Conseil d'Ingenum
- Plusieurs pistes























- Mésocentre de calcul de Montpellier
- Avantages :
 - Ressource de calcul académique
 - Indépendante du domaine d'application
 - Domaine de compétence de l'équipe et de l'unité (CentOS, slurm)
 - Ressources nécessaires et scalables
 - Factorisations des utilisations
 - Flexibilité assurée par Singularity
 - Accès permanent (pas uniquement sur projet) et facile moyennant finance

• Inconvénients:

- Service VM non assuré mais montage SSHFS possible
- Données non sauvegardées (à revisiter)

















SIWAA

C'est quoi? Hier, Aujourd'hui, Demain (3)

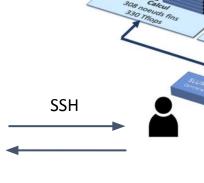
- 2019 -> aujourd'hui
- Hébergement réparti entre la DSI INRAE et meso@LR
- Subtilité : montage SSHFS entre VM et meso-centre

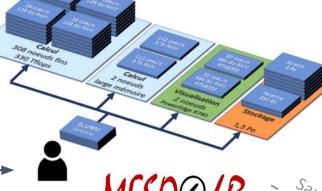
Utilisation d'un plugin Galaxy spécifique SSH

• Aide de Damien et Mikaël











Serveur unité

















- 308 nœuds de calcul Dell PowerEdgeC6320
 - bi processeurs Intel Xeon E5-2680 v4 2,4 Ghz(broadwell)
 - 8624 cœurs
 - 128 Go RAM par nœuds
 - 280 TflopsLinpack
- 1 Po de stockage rapide
- 326 To de stockage pérenne
- Réseau d'interconnexion Intel OmniPath100 Gb/s
- Pas d'accélérateur
- 2 nœuds épais : 80 cœurs, 1To de RAM

https://meso-lr.umontpellier.fr/wp-content/uploads/2018/11/1-Presentation_cluster_Muse.pdf

















Tarifs calcul	Laboratoires de l'Université de Montpellier & Partenaires académiques MUSE	Académiques Autres	Privé
Heure cœur	0,01 €HT		0,02 €HT
Nœud standard à usage exclusif par an 3 To de stockage associés inclus	1 000 € HT		2500 € HT

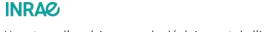
Tarifs stockage / To / an (2)	Laboratoires de l'Université de Montpellier & Partenaires académiques MUSE	Académiques Autres	Privé
To Associé au calcul	40,00 €HT	45,00 €HT	120,00 €HT
To par an (sans minimum)	50,00 €HT	55,00 €HT	
To par an > 100 To	45,00 €HT	50,00 €HT	135,00 €HT
To par an > 1000 To (soit 1 Po) (1)	30,00 €HT	35,00 €HT	90,00 €HT

Académique : 1 000€ pour 100 000 h de calcul

Les ressources sont attribuées par projet. Un projet contient autant de comptes utilisateurs qu'on le souhaite.

https://meso-lr.umontpellier.fr





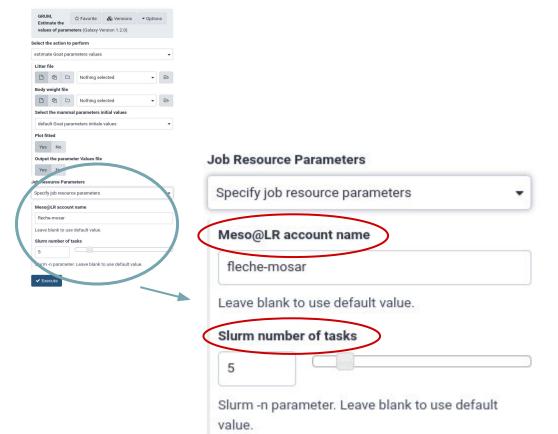








SIWAA, la gestion de la ressource calcul un focus sur l'accessibilité fournie aux utilisateurs



But:

- Décompter les ressources utilisées du bon projet meso@LR
- Choisir le nombre de processus pour la parallélisation (impossible sur useGalaxy.eu)
- Personnalisation du fonctionnement des "destinations" Galaxy: avec un peu de code python













SIWAA, les droits d'utilisation sur les outils un focus sur l'accessibilité fournie aux utilisateurs

- Pas vraiment dans l'esprit de Galaxy
- Mais important pour nos fournisseurs de modèles
- Fonctionnement:
 - Tous les outils sont visibles par tout le monde
 - Mais l'exécution sera bloquée ou avec peu de ressources pour les utilisateurs non autorisés
- Techniquement:
 - Utilisation des rôles et des destinations de Galaxy
 - Code Python spécifique s'appuyant sur des mécanismes de Galaxy











SIWAA, administration Aujourd'hui et demain

Actuellement

- VM "classique" DSI (VMWare)
- Installation manuelle & Wiki de doc
- Identification Idap Inra + comptes extérieurs















SIWAA, administration Aujourd'hui et demain

Actuellement

- VM "classique" DSI (VMWare)
- Installation manuelle & Wiki de doc
- Identification Idap Inra + comptes extérieurs

En préparation

- VM sur le cloud de la DSI (OpenStack)
 - Plus d'autonomie dans la mise en place de VM
- Automatisation de l'installation et des mises à jour (Ansible)
 - Installation décrite donc documentée
 - Versionnement de l'état des machines
 - (ré)installation facilitée
 - (mise à jour facilitée à éprouver)
 - Système de sauvegarde à revoir
- Identification Flixing
 - Annuaire des membres d'Elixir + orcid + linkedin + google



















SIWAA, panorama des outils Disponibles

Aujourd'hui

Les modèles de cultures

- Stics natif 8.50, 9.0, 9.1, 9.2, 9.2-mila, multi-simulation, SA, Calibration, R | Stics Docker | Stics R Docker]
- Stics encapsulé dans VLE pour [VLE MPI Docker]
 - AgFlow, plan d'expérience et rotation
 - Markiz4Vaccarm
- Azodyn [VLE Docker]

Un modèle de phénologie

PMP4TEMPO, simulateur Vle Docker [VLE Docker]

• Les modèles d'intérêt zootechniques

- GRUM modèles de poids vif de mammifère, R Docker
- SighmaZ modèle de troupeau de chèvre, Python Docker

Petits Outils

- Générateur d'hypercube latin pour le projet Flèche, R, + paquet conda
- Visualisation de sorties pour Operate, R

Interface d'accès à des données

- BD Sol, query sql pure python
- DRIAS, données climatiques simulées, convertisseur, scénarios climatiques











SIWAA, Développement

Ajouter un outils, en quoi ça consiste, fondamentalement?

- Ecrire un fichier XML (tool config file, wrapper)
 - Définir le script qui doit être exécuté
 - Définir les entrées = Définir un formulaire
 - fichiers (qui se trouvent dans l'environnement utilisateurs)
 - paramètres des formulaire
 - Définir les sorties = fichiers qui vont être disponibles dans l'environnement utilisateurs



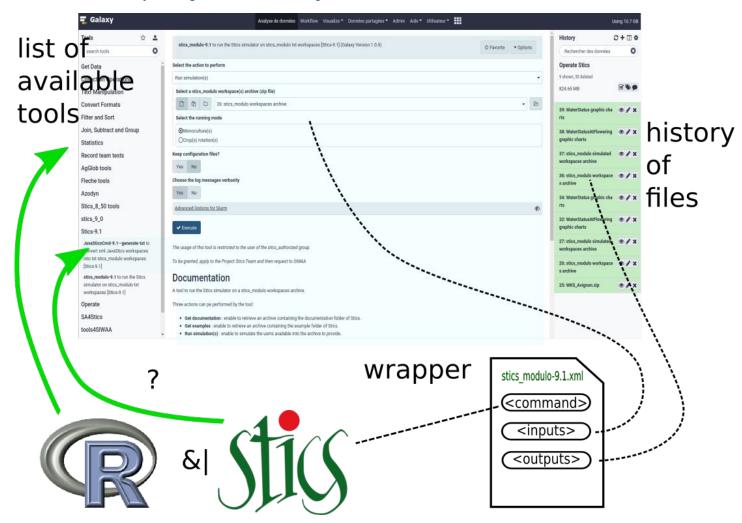






SIWAA, Développement

Ajouter un outils, en quoi ça consiste, fondamentalement?













SIWAA, Développement Ajouter un outils, scripts et dépendances?

- Ecrire un fichier XML (tool config file, wrapper)
 - Définir le script qui doit être exécuté
 - Définir les entrées = Définir un formulaire
 - fichiers (qui se trouvent dans l'environnement utilisateurs)
 - paramètres des formulaire
 - Définir les sorties = fichiers qui vont être disponibles dans l'environnement utilisateurs
 - Définir les dépendances
 - L'approche Conda = les commandes, paquets sont des paquets
 - A L'installation de l'outils il est confié à Conda de résoudre les dépendances
 - L'approche Container = la ligne de commande sera exécutée dans le container
 - Avant l'exécution le container est téléchargé si besoin
- Ajouter des fichiers accompagnants l'outils
 - scripts R, Python, ...
 - fichiers de données











SIWAA, Développement Ajouter un outils, tests et documentation?

- Ajouter des balises pour les tests.
 - Tests fonctionnels
 - Jeux de données de tests
- Ajouter des balises pour de la doc.
 - Disponible directement sur le formulaire











SIWAA, Développement Galaxy toolshed



- Un site web companion qui héberge les outils Galaxy Packagés.
- L'interface d'administration permet d'installer les outils à partir d'un tool shed
- Gestion de version des Packages.
- Permet de partager les outils sur toutes les instances.
- Augmenter le niveau de reproductibilité.
- Plus d'indépendance entre les outils et les instances.

Plus une démarche, qu'une obligation.



Utiliser son propre toolshed, pour avoir une démarche cadrée, mais en s'autorisant de la flexibilité. Exemple: utilisation de logiciels (|images) proriétaires.











SIWAA, Développement Les grandes lignes

- Utilisation des forges
- Automatisation des processus CI/CD
- Utilisation du toolshed, versionning et transparence
- Pas d'a priori sur ce qui est disponible sur le serveur de calcul
 - Conda
 - Docker
- Utiliser planemo, suite logicielle pour faciliter la validation et le déploiement d'un outils galaxy





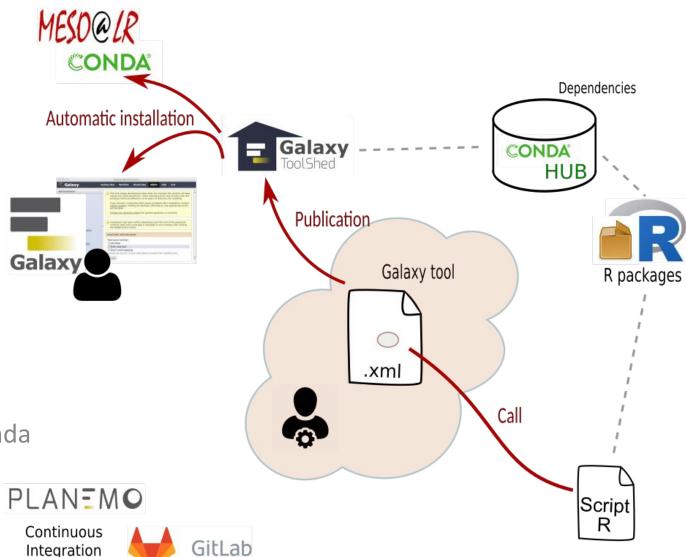






SIWAA, Développement Le cas d'un tool classique

- Tous les Codes hébergés sur forgemia
 - Recette Conda
 - Scripts python, R, ...
 - wrapper tool Galaxy
- CI/CD de gitlab
 - tester les tools galaxy
 - publier les tools sur le toolshed
 - construire et publier les paquets conda
 - Hub Conda
- Installation des outils manuelle
- Résolution des dépendance par galaxy/conda à l'installation









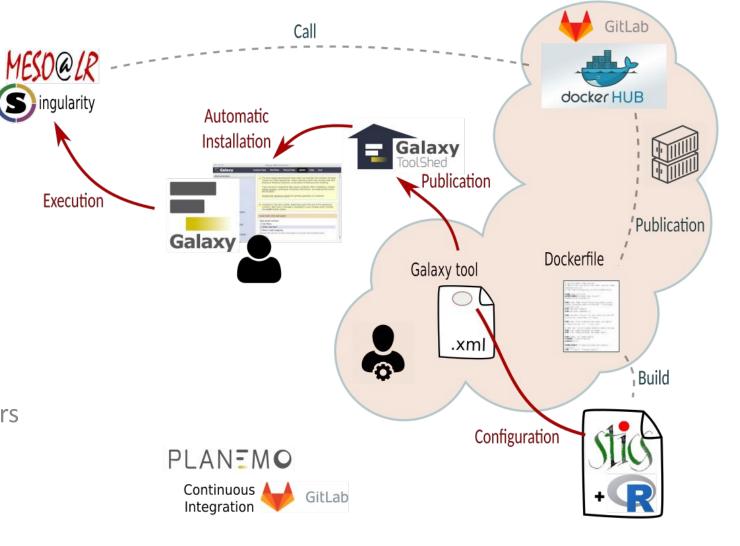




SIWAA, Développement Le cás d'un tool containerisé

Tous les Codes hébergés sur forgemia

- Recettes docker
- Scripts python, R, ...
- wrapper tool Galaxy
- CI/CD de gitlab
 - tester les tools galaxy
 - o publier les tools sur le toolshed
 - construire et publier les containers
 - Docker Hub
 - Registry miaforge
- Installation des outils manuelle
- Résolution de la dépendance par Singularity (cache)













Le réseau GEOC

Depuis quand, Qui sommes nous, Pour quoi faire?

- depuis 2017
- Galaxy Extra Omic Community
- 3 Instances de Galaxy & 3 Domaines d'applications
 - <u>ChemFlow</u>, Chimiométrie
 - <u>W4m</u>, Métabolomique
 - SIWAA, Simulation Agro Ecosystèmes
- 3 CATI = réseaux métiers (structurer la production)
 - Empreinte, Molecular PhEnotyping and blochemical daTa Engineering
 - o IUMAN, Informatisation et Utilisation des Modèles pour les Agroécosystèmes Numérique
 - Codex, COnnaissance et Données EXpérimentales

geoc@groupes.renater.fr

- eric.latrille@inrae.fr
- estelle.ancelet@inrae.fr
- franck.giacomoni@inrae.fr
- jean-claude.boulet@inrae.fr
- lain.pavot@inrae.fr
- patrick.chabrier@inrae.fr
- virginie.rossard@inrae.fr











SIWAA, GEOC, HIT Le hackathon sur les IT















lundi 7 décembre	14:00 - 17:00	Yvan & Patrick	Lancement & Introduction
Mardi 8 décembre	9:00 - 17:30	Anthony & Estelle	Admin - Configurer Galaxy
Mercredi 9 décembre & jeudi 10 décembre	9:00 - 13:00 & 9:00 - 13:00	Franck, Yvan , Lain, Romain, Coline, Elie	Développement d'IT pour Galaxy

- Ancelet Estelle INRAE
- Arnaud Elie MNHN
- Boulet Jean-Claude INRAE
- Bretaudeau Anthony INRAE IFB
- Chabrier Patrick INRAE
- Dallet Romain CNRS IFB
- Giacomoni Franck INRAE
- Latrille Éric INRAE
- Le Bras Yvan MNHN
- Pavot Lain INRAE
- Rossard Virginie INRAE
- Royaux Coline MNHN

Une vingtaine de participants











SIWAA, et la communauté Galaxy Participation à des évènements



Janvier 2021, Estelle Ancelet, contribution à la formation



Juin-Juillet 2021, GEOC -> rédaction d'un nouveau tutorial pour développer des IT (outils interactifs), avec IFB & Muséum

& Correction de bugs sur https://github.com/galaxyproject









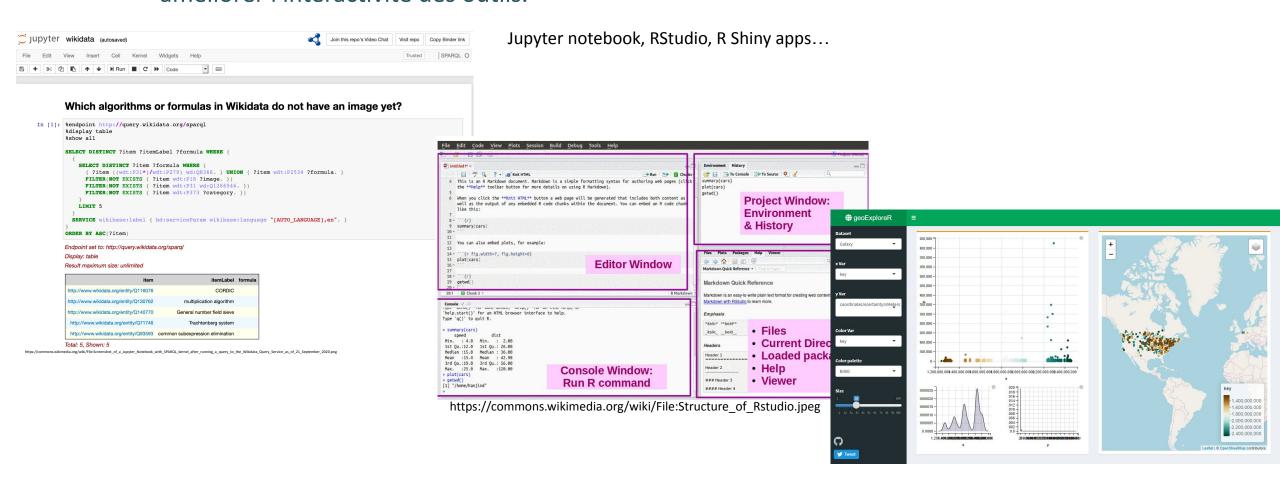




SIWAA et les Perspectives liées aux Interactive tools

Les outils Galaxy, au top pour la reproductibilité, mais.... comment aller plus loin ?

→ améliorer l'interactivité des outils.















SIWAA, Les perspectives liées à des projets rien de mieux pour progresser



- CarSolEl: Dans la continuité de 4p1000, Construction d'une méthodologie et d'un référentiel sur les flux de carbone dans les sols agricoles des territoires d'élevage bovin, plan d'expérience massif, Stics encapsulé dans VLE...
- CarboSeq: EJP Soil, Simulations à différentes échelles de l'estimation du potentiel de séquestration du carbone dans le sol.(2022), ToolBox R pour interfacer les bases de données avec le modèle RothC (package soilR)
- GRUM: Simulation du poids vifs de chèvres, estimation des paramètres & identifiabilité, valorisation de données prélevées à l'étable, R Docker
- Operate: parallélisation de simulations avec le simulateurs Stics natif. Stics in Docker.











SIWAA, autres perspectives

- Interopérabilité avec des bases de données.
- Utilisation de l'API Rest de Galaxy en complément de l'API erecord
 - Permet de lancer des tâches sans passer par l'interface web
- Mieux et plus de monitoring pour:
 - Les administrateurs
 - Les utilisateurs
- Amélioration de la description de l'offre
- Communication









