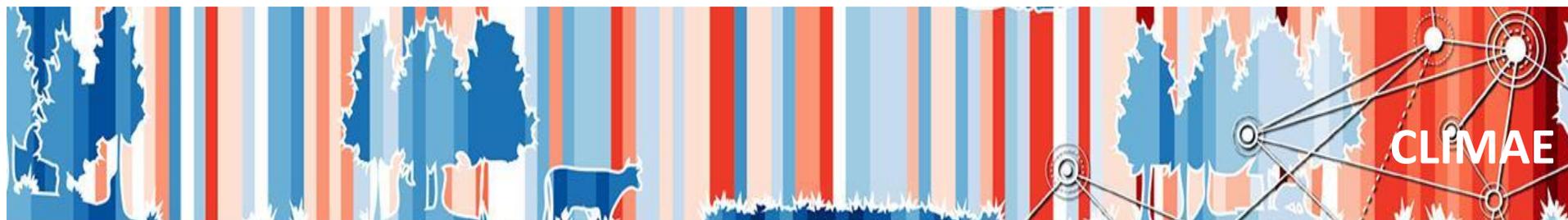




➤ **Development, deployment and execution of simulation workflows to study the impact of climate change on dairy farms**

Patrick Chabrier, Graux Anne-Isabelle, **Eric Casellas**, Klervi Le-Floch, Patrice LeCharpentier, Renan Le-Roux, Fabien Ferchaud

XIII STICS Seminar STICS - 13-16 November 2023



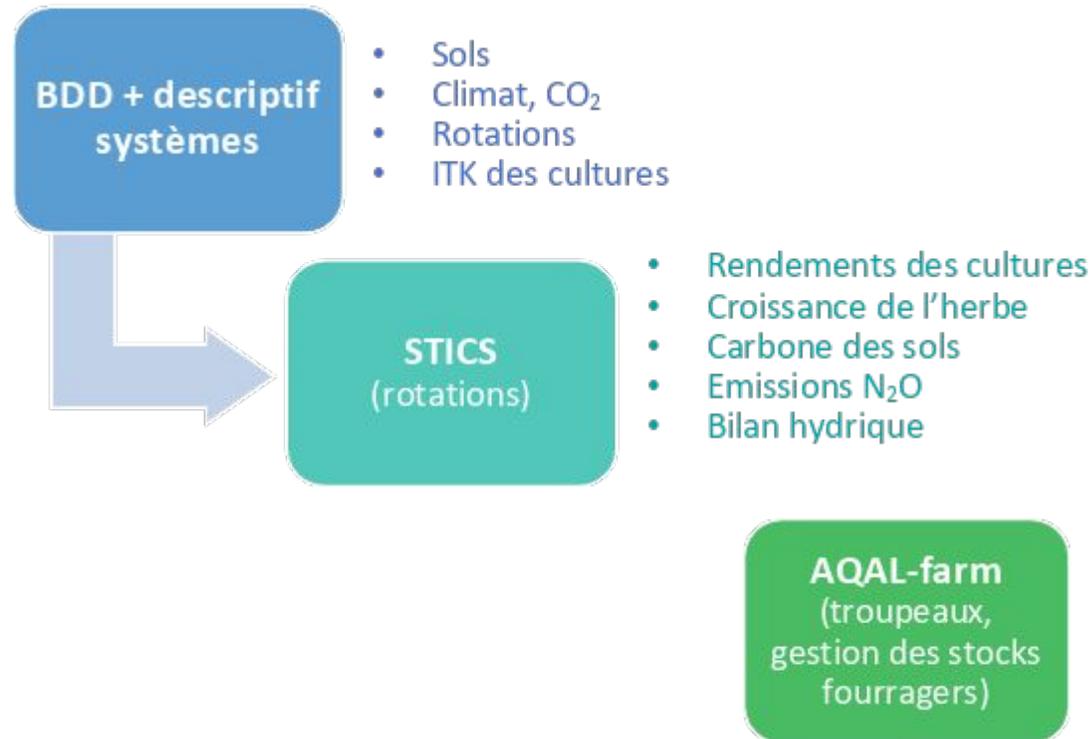
REDELAC

métaprogramme CLIMAE

REDELAC (2023 -2024) - Resilience And Sustainability Of Lowland Dairy Farms To Climate Hazards



Pour le territoire du
Pays de Fougères



- Typologies de fermes
- Systèmes de cultures
- plan de rotations
- UPC
- Scénarios climatiques
- ?~> 10 millions de simulations Stics



INRAE

Contraintes et besoins

pour le passage à l'échelle du plan d'expérience

- En référence un workflow R SticsRpacks Windows opérationnel
 - Rester en conformité
 - En entrée des dossiers excel spécifiant des plans de rotations pour une UPC
- Solution pour l'extension du plan
- Utilisation de Stics sur une branche de recherche "Prairie"
 - Evolution rapide
- Utilisation des paquets SticsR Packs pour rester dans un cadre partagé entre ingénieurs et chercheurs.
 - Évolutions probables aussi
- Accessibilité à des ressources de calcul de type HPC
- In Fine, travail collaboratif entre plusieurs groupes. équipe scientifiques, équipe Stics, équipe simulations



SIWAA

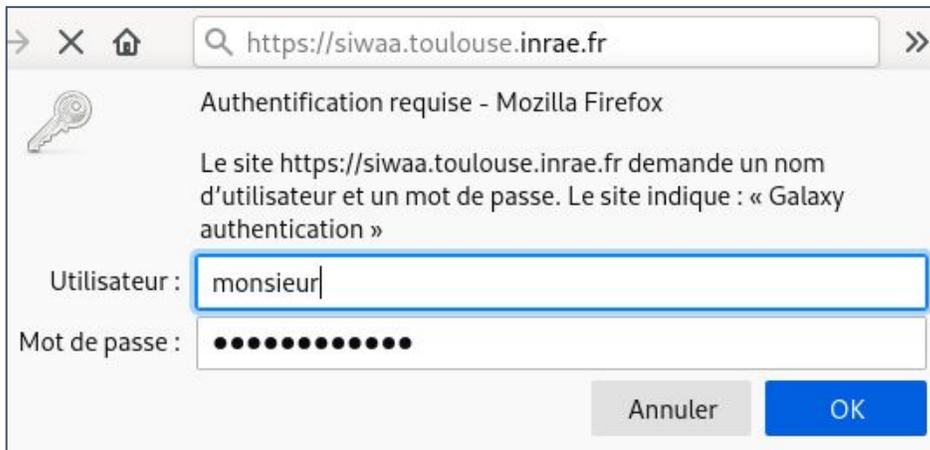
pourquoi?

- 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.



Galaxy/SIWAA

Comment ça marche



https://siwaa.toulouse.inrae.fr

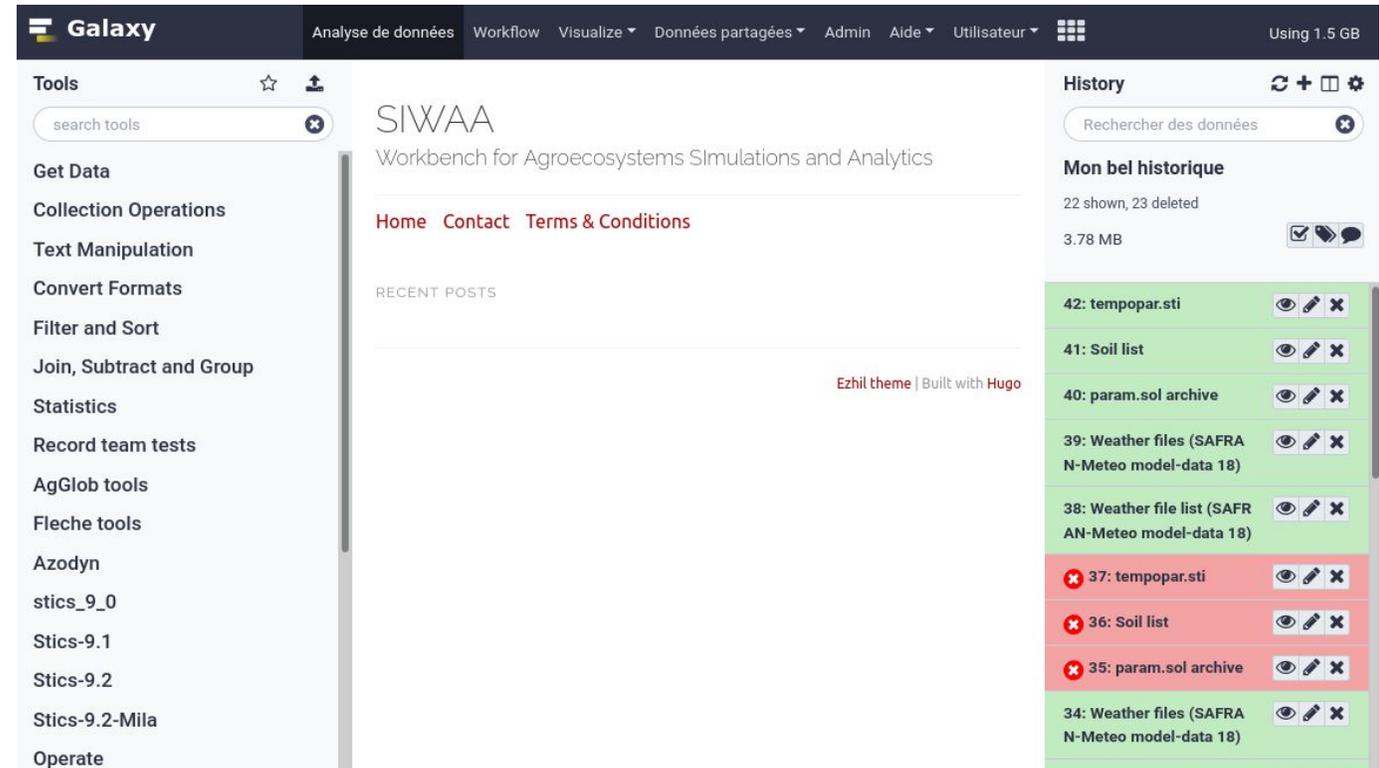
Authentification requise - Mozilla Firefox

Le site https://siwaa.toulouse.inrae.fr demande un nom d'utilisateur et un mot de passe. Le site indique : « Galaxy authentication »

Utilisateur : monsieur

Mot de passe : ●●●●●●●●●●

Annuler OK



Galaxy

Analyse de données Workflow Visualize Données partagées Admin Aide Utilisateur Using 1.5 GB

Tools

search tools

SIWAA

Workbench for Agroecosystems Simulations and Analytics

Home Contact Terms & Conditions

RECENT POSTS

Ezhil theme | Built with Hugo

History

Rechercher des données

Mon bel historique

22 shown, 23 deleted

3.78 MB

42: tempopar.sti	👁️ ✎️ ✕️
41: Soil list	👁️ ✎️ ✕️
40: param.sol archive	👁️ ✎️ ✕️
39: Weather files (SAFRA N-Meteo model-data 18)	👁️ ✎️ ✕️
38: Weather file list (SAFRA AN-Meteo model-data 18)	👁️ ✎️ ✕️
✕️ 37: tempopar.sti	👁️ ✎️ ✕️
✕️ 36: Soil list	👁️ ✎️ ✕️
✕️ 35: param.sol archive	👁️ ✎️ ✕️
34: Weather files (SAFRA N-Meteo model-data 18)	👁️ ✎️ ✕️



INRAE

Galaxy/SIWAA

Comment ça marche

Menu

The screenshot shows the Galaxy/SIWAA web interface. At the top, a dark navigation bar contains the 'Galaxy' logo and a menu with items: 'Analyse de données', 'Workflow', 'Visualize', 'Données partagées', 'Admin', 'Aide', and 'Utilisateur'. The main content area is divided into three sections:

- Tools (left, red border):** A sidebar menu with a search bar and categories: 'Get Data', 'Collection Operations', 'Text Manipulation', 'Convert Formats', 'Filter and Sort', 'Join, Subtract and Group', 'Statistics', 'Record team tests', 'AgGlob tools', 'Fleche tools', 'Azodyn', 'stics_9_0', and 'Stics_0_1'. A red box highlights this entire sidebar, with the label 'Outils' at the bottom.
- SIWAA (center, green border):** The main workspace area. It features the title 'SIWAA' and subtitle 'Workbench for Agroecosystems Simulations and Analytics'. Below this are navigation links for 'Home', 'Contact', and 'Terms & Conditions'. A section for 'RECENT POSTS' is visible. At the bottom right, it says 'Eznil theme | Built with Hugo'. A green box highlights this central area, with the label 'Espace de travail' at the bottom.
- History (right, cyan border):** A sidebar showing a list of data sets. It includes a search bar 'Rechercher des données' and a section 'Mon bel historique'. The first entry is 'Jeux de données' (3.78 MB, I/O). Below it are several entries with IDs and names, each with icons for view, edit, and delete. A cyan box highlights this sidebar, with the label 'Jeux de données' at the top.



INRAE

Galaxy/SIWAA

Comment

The screenshot displays the Galaxy web interface for configuring the `stics_modulo-9.1` tool. The interface is divided into three main sections:

- Left Sidebar (Tools):** Lists various tool categories such as AgGlob tools, Fleche tools, Azodyn, stics_9_0, Stics-9.1, Stics-9.2, Stics-9.2-Mila, Operate, tools4SIWAA, GRUM, PMP4TEMPO, SUNFLO, MARKIZ, and Stics-8.50. A red arrow points to the `stics_modulo-9.1` tool entry.
- Main Configuration Area:** Shows the tool name `stics_modulo-9.1` and its version (1.0.0). It includes a "Select the action to perform" dropdown set to "Run simulation(s)", a "Select a stics_modulo workspace(s) archive (zip file)" dropdown set to "1: example_Mila_92_txt.zip", and a "Select the running mode" section with radio buttons for "Monoculture(s)" (selected) and "Crop(s) rotation(s)". There are also sections for "Keep configuration files?" (Yes/No) and "Choose the log messages verbosity" (Yes/No). A blue "Execute" button is at the bottom.
- Right Sidebar (History):** Shows a search bar and a list of historical runs. One run is highlighted in green: "1: example_Mila_92_txt.zip". A red arrow points to this entry.



INRAE

Galaxy/SIWAA

Comment ça marche

The screenshot shows the Galaxy web interface with a dark header. The main content area is divided into three panels. On the left is the 'Tools' panel with a search bar and a list of tools including 'stics_9_0' and 'Stics-9.1'. The central panel displays a green notification box with a checkmark, stating: 'Executed stics_modulo-9.1 and successfully added 1 job to the queue. The tool uses this input: 1: example_Mila_92_txt.zip. It produces this output: 2: stics_modulo simulated workspaces archive. You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.' On the right is the 'History' panel, which shows a search bar and a list of jobs. The top job is '2: stics_modulo simulated workspaces archive' with a status icon of a gear and a checkmark. Below it is '1: example_Mila_92_txt.zip' with a status icon of a checkmark. A red arrow points from the text 'Exécution' below to the gear icon of the top job in the history panel.

This screenshot shows the Galaxy web interface with a dark header. The 'Tools' panel on the left is visible. The central panel shows a green notification box with a checkmark, identical to the one in the top screenshot. The 'History' panel on the right shows a search bar and a list of jobs. The top job is '2: stics_modulo simulated workspaces archive' with a status icon of a gear and a checkmark. Below it is '1: example_Mila_92_txt.zip' with a status icon of a checkmark. A red arrow points from the text 'En attente d'exécution' below to the gear icon of the top job in the history panel.

Exécution

En attente d'exécution



Galaxy/SIWAA

Comment ça marche

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'Analyse de données', 'Workflow', 'Visualize', 'Données partagées', 'Admin', 'Aide', 'Utilisateur', and 'Using 1.5 GB'. The left sidebar lists tools under categories: 'Tools', 'AgGlob tools', 'Fleche tools', 'Azodyn', 'stics_9_0', 'Stics-9.1', and 'Stics-9.2'. The main content area displays a green notification box with a checkmark: 'Executed stics_modulo-9.1 and successfully added 1 job to the queue.' Below this, it lists the input: '1: example_Mila_92_txt.zip' and the output: '2: stics_modulo simulated workspaces archive'. A 'History' panel on the right shows 'Autre historique' with '2 shown' and '152.39 KB'. The job entry '2: stics_modulo simulated workspaces archive' is highlighted in green, with a green arrow pointing to it and the text 'Ok'.

The screenshot shows the Galaxy web interface with a failed job execution. The top navigation bar is identical to the previous screenshot. The left sidebar is also identical. The main content area displays a green notification box with a checkmark: 'Executed stics_modulo-9.1 and successfully added 1 job to the queue.' Below this, it lists the input: '3: download.zip' and the output: '4: stics_modulo simulated workspaces archive'. The 'History' panel on the right shows 'Autre historique' with '4 shown' and '47.05 MB'. The job entry '4: stics_modulo simulated workspaces archive' is highlighted in red, with a red arrow pointing to it and the text 'KO'.



Galaxy/SIWAA

Comment ça marche...notion de workflow

The screenshot displays the Galaxy web interface for editing a workflow named "Operate Stics Workflow". The workflow consists of the following steps:

- WKS_Avignon.zip**: Input file.
- output**: Intermediate output.
- JavaSticsCmd-9.1--generate-txt**: Select an JavaStics xml workspace(s) archive (zip file).
 - Input: stics_modulo workspaces archive (zip)
- stics_modulo-9.1**: Select a stics_modulo workspace(s) archive (zip file).
 - Input: stics_modulo simulated workspaces archive (zip)
- waterStatusAtFloweringGraphics**: Select a stics_modulo simulated workspace(s) archive (zip file).
 - Input: WaterStatusAtFlowering graphic charts (zip)
- waterStatusGraphics**: Select a stics_modulo simulated workspace(s) archive (zip file).
 - Input: WaterStatus graphic charts (zip)

On the right, a preview window shows a grid of box plots for various scenarios: "cestout rp452 309 Charger", "cestout rp452 309 Soissons", "cestout rp452 309 Arminda", "cestout rp452 323 Charger", "cestout rp452 323 Soissons", and "cestout rp452 323 Arminda". Each plot compares three scenarios: 1.RP, 2.NF, and 3.FF. The y-axis represents a numerical value, and the x-axis represents the scenario.

Below the workflow, configuration options are visible:

- Advanced Options for Slurm**: A link to expand options.
- Email notification**: Yes No. An email notification will be sent when the job has completed.
- Output cleanup**: Yes No. Upon completion of this step, delete non-starred outputs from completed workflow steps if they are no longer required as inputs.



INRAE

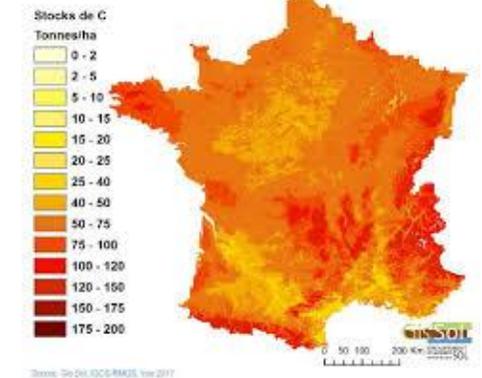
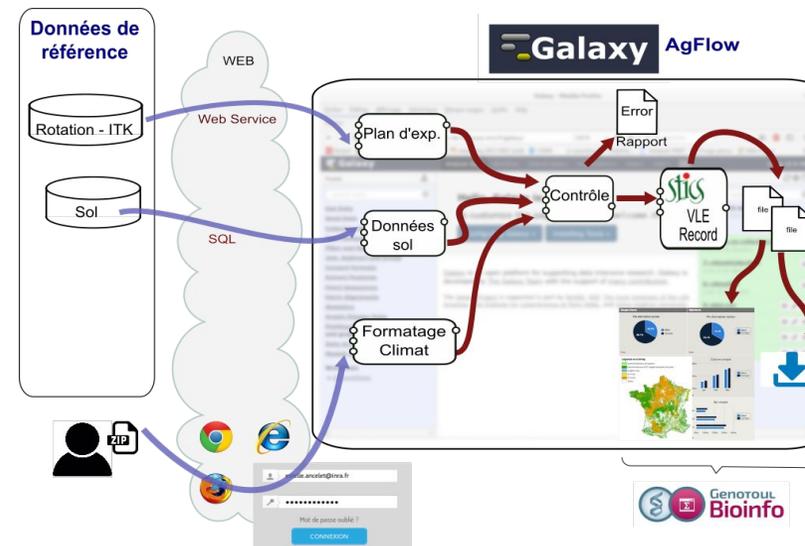
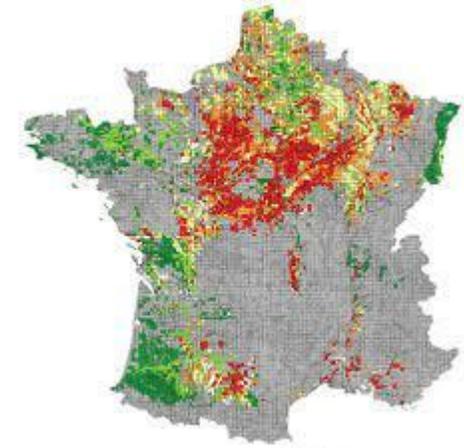
SIWAA

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



INRAE

Galaxy

Pourquoi?

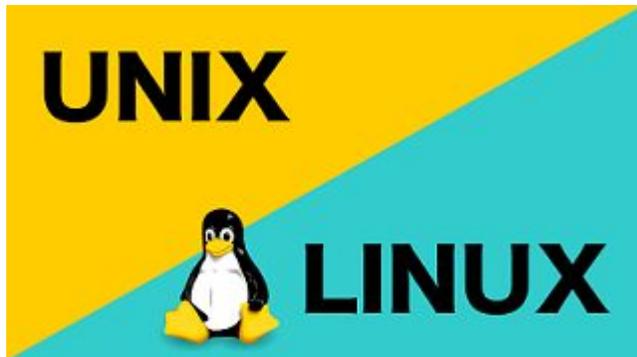
- **Accessibilité**
 - interface web, pas de code pour l'utilisateur
 - capacité d'intégration des logiciels
- **Flexibilité dans l'interaction aux ressources de calcul**
- **Partage / Reproductibilité / Transparence**
 - Tout est tracé et peut être rejoué
 - Les outils et leurs dépendances sont versionnés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



Galaxy/SIWAA

Qu'est ce qu'un outil ?

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



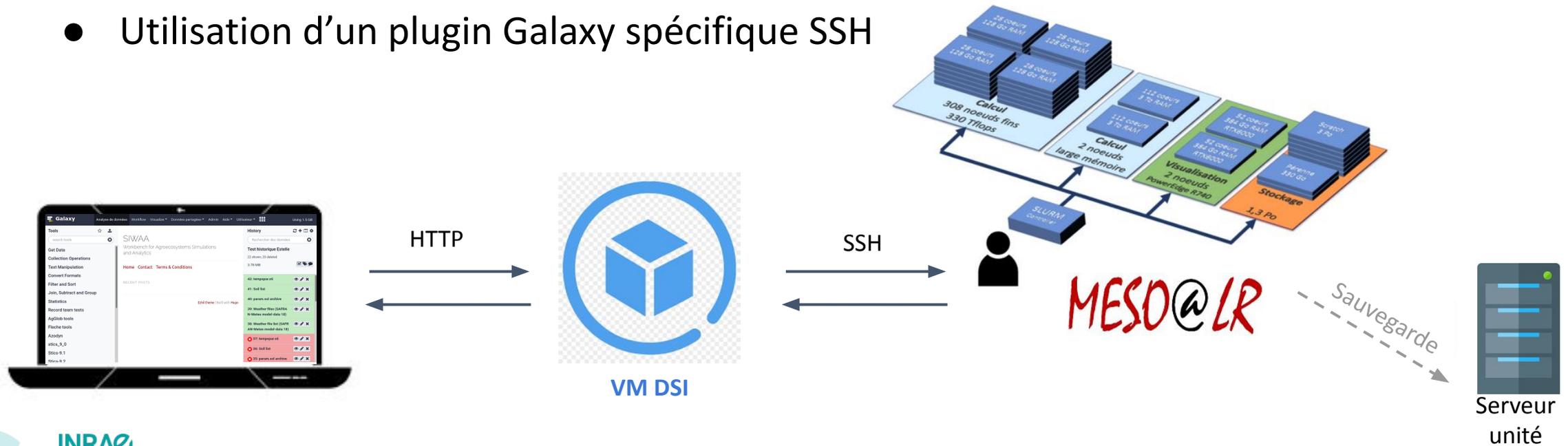
- Mésocentre de calcul de Montpellier
- Avantages :
 - Ressource de calcul académique
 - Indépendante du domaine d'application
 - Domaine de compétence de l'équipe et de l'unité (CentOS, slurm)
 - Ressources nécessaires et scalables
 - Factorisations des utilisations
 - Flexibilité assurée par Singularity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



INRAE

- 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

Académique : 1 000€ pour 100 000 h de calcul

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	150,00 €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

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

The image shows a screenshot of the SIWAA Galaxy tool interface. On the left, there is a sidebar with various tool options, including 'Job Resource Parameters', which is circled in blue. An arrow points from this sidebar to a larger, detailed view of the 'Job Resource Parameters' tool on the right. In this detailed view, two fields are circled in red: 'Meso@LR account name' with the value 'fleche-mosar' and 'Slurm number of tasks' with the value '5'. The interface includes a dropdown menu for 'Specify job resource parameters', a text input field for the account name, and a slider for the number of tasks. Below the slider, there is a note: 'Slurm -n parameter. Leave blank to use default value.'

- 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, Développement

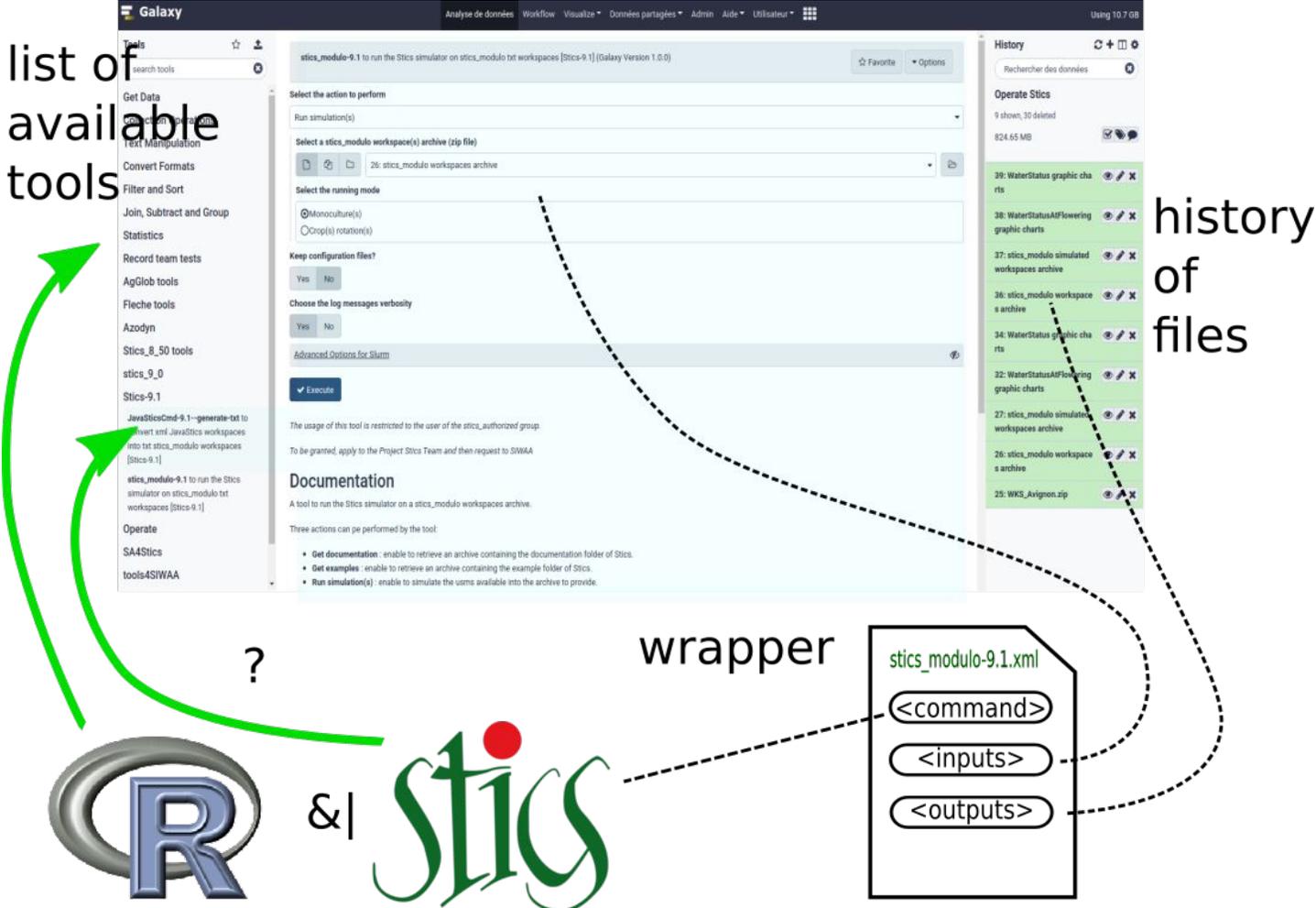
Ajouter un outils, en quoi ça consiste, fondamentalement?

- 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?





- 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) propriétaires.

SIWAA, Développement

Les grandes lignes

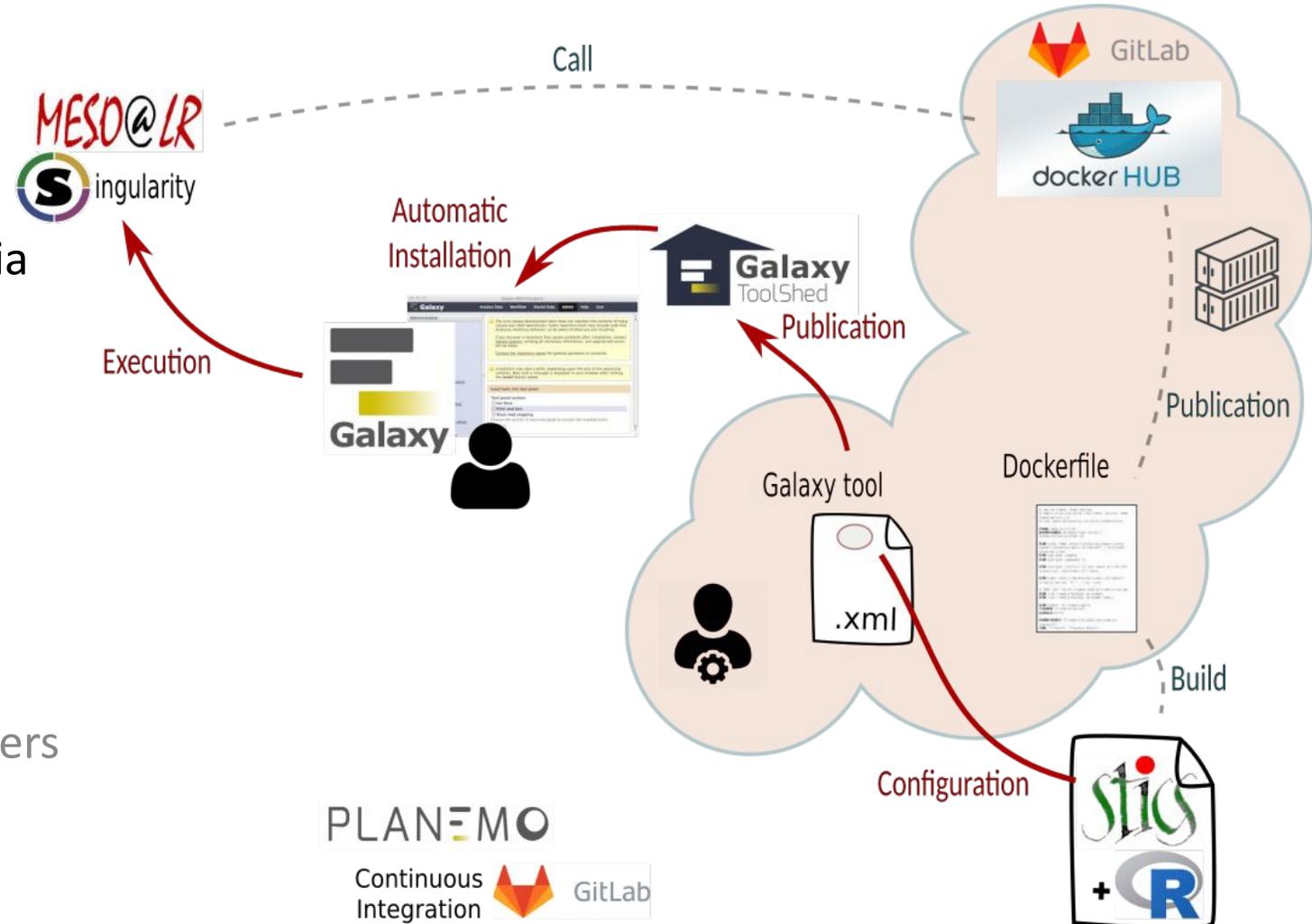
- Utilisation des forges
- Automatisation des processus CI/CD
- Utilisation du toolshed, versionning et transparence
- Pas d'à priori sur ce qui est disponible sur le serveur de calcul
 - Conda
 - Docker
- Utiliser planemo, suite logicielle pour faciliter la validation et le déploiement d'un outils galaxy



SIWAA, Développement

Le cas d'un tool containerisé

- Tous les Codes hébergés sur forgemia
 - Recettes docker
 - Scripts python, R, ...
 - wrapper tool Galaxy
- CI/CD de gitlab
 - tester les tools galaxy
 - 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)



Outil redelacXlsx2Csv

Extrait d'un plan de rotations atomique excel des fichiers csv

- Pas de solution trouvé à ce jour pour faire du scripting en direct sur des fichiers excels.
- Impossible de recalculer les formules.
- Travailler avec des fichiers excels permet de visualiser dans Galaxy/SIWAA le contenu.
- Extraction uniquement des données nécessaires
- Entrée : Feuille excel spécifiant les entrées des Stics pour un plan de rotation
- Sorties : un fichier csv par onglet pour les données utiles à l'expansion
- Script Python



Outil redelacPlanExpand

expansion d'un plan de rotations en fonction d'une liste d'UPC

- Pas de solution trouvé à ce jour pour faire du scripting en direct sur des fichiers excels.
- Duplique la liste des USMS autant de fois que nécessaire
- Remplace les valeurs dupliquées par les valeurs statiques du plan fourni en entrée
- Se substitue aux formules excel et recalcule les cellules dynamiques
- Régénère la liste des rotations à simuler
- Entrée: des fichiers csv
- Sortie: des fichiers csv
- Script Python



Outil redelacSticsInputGenerator

génère les fichiers d'entrées XML puis TXT de toutes les situations nécessaires

- Le maillon faible dans le dispositif :) plus long que la simulation.
- Utilise la nouvelle version de SticsRFile pure XSLT
- Les fichiers d'entrée CSV sont std SticsRfile
- Un workspace de base contenant des données sol en entrée
- Donnée météo en entrée aussi(pour pouvoir changer de scénario)
- Hack de SticsRfile pour spécifier un nom de fichier USM xml
- Autorise la parallélisation de la génération des répertoire de configuration USM au format txt.
- Entrée: des fichiers csv pour les paramètre Stics + archive météo zip + workspace de base au format zip
- Script R SticsRFiles



Outil redelacSticsSimulator

Simulation en rotation en parrallel

- Utilise une version légèrement modifiée de SticsOnR pour la parallélisation avec dofuture
- Autorise la parallélisation des simulations de rotations
- Entrée: Un workspace STICS contenant des USMs au format txt et un fichier csv spécifiant les rotations
- En sortie: Un workspace simulé et un répertoire de résultats contenant des fichiers de donnée csv et R.
- Script R SticsOnR



2 containers Dockers/Singularity

pour répondre aux besoins des outils

- `docker://registry.forgemia.inra.fr/redelac/redelac-toolbox`
 - une machine virtuelle boîte à outils python
 - pour servir les outils en charge de l'expansion du plan
- `docker://registry.forgemia.inra.fr/redelac/redelac-toolbox/sticsrpacks`
 - une machine virtuelle SticsRPacks
 - pour servir les outils
 - de générations des entrées du simulateur Stics
 - de simulations des rotations.
 - très pratique pour adresser des versions particulières des outils.



GitHub & forgemia

pour tracer, automatiser et garder le contrôle

- <https://forgemia.inra.fr/redelac/e3-et-e4>
 - dépôt de référence et de spécification des plans d'expérience en R
- <https://forgemia.inra.fr/redelac/redelac-toolbox>
 - Les outils et les containers déployés sur SIWAA
- <https://github.com/SticsRPacks/SticsRFiles/>
 - <https://github.com/Chabrier/SticsRFiles>
- <https://github.com/SticsRPacks/SticsOnR>
 - <https://github.com/Chabrier/SticsOnR>
- <https://forgemia.inra.fr/stics-dev/modulostics>

```
93 # necessite une variable STICS_REPO_TOKEN donnant l'accès au depot de source de stics en lecture
94 # fournis sous la forme d'un artefact stics_modulo attendu par le job build_docker_sticssimulator
95 get_stics_modulo:
96   stage: .pre
97   image: registry.gitlab.com:5000/stics-dev/modulostics
98   variables:
99     GIT_STRATEGY: none
100    BRANCH_NAME: "plecharj"
101    EXPLICIT_COMMIT: "bc2681dcab3495123156e0bb4c2e7a39de0ededb"
102    FPM_FLAGS: "-Werror -Wimplicit-interface -fmax-errors=1 -fimplicit-none"
103   script:
104     - git clone --single-branch --branch $BRANCH_NAME --depth 1 https://oauth2:$STICS_REPO_TOKEN@forgemia.inra.fr/stics-dev/modulostics.git
105     - git -C modulostics branch sticsexplicit $EXPLICIT_COMMIT
```

Development, deployment and execution of simulation workflows to study the impact of climate change on dairy farms

XIII Séminaire STICS 14 novembre 2023 / Patrick Chabrier / RECORD

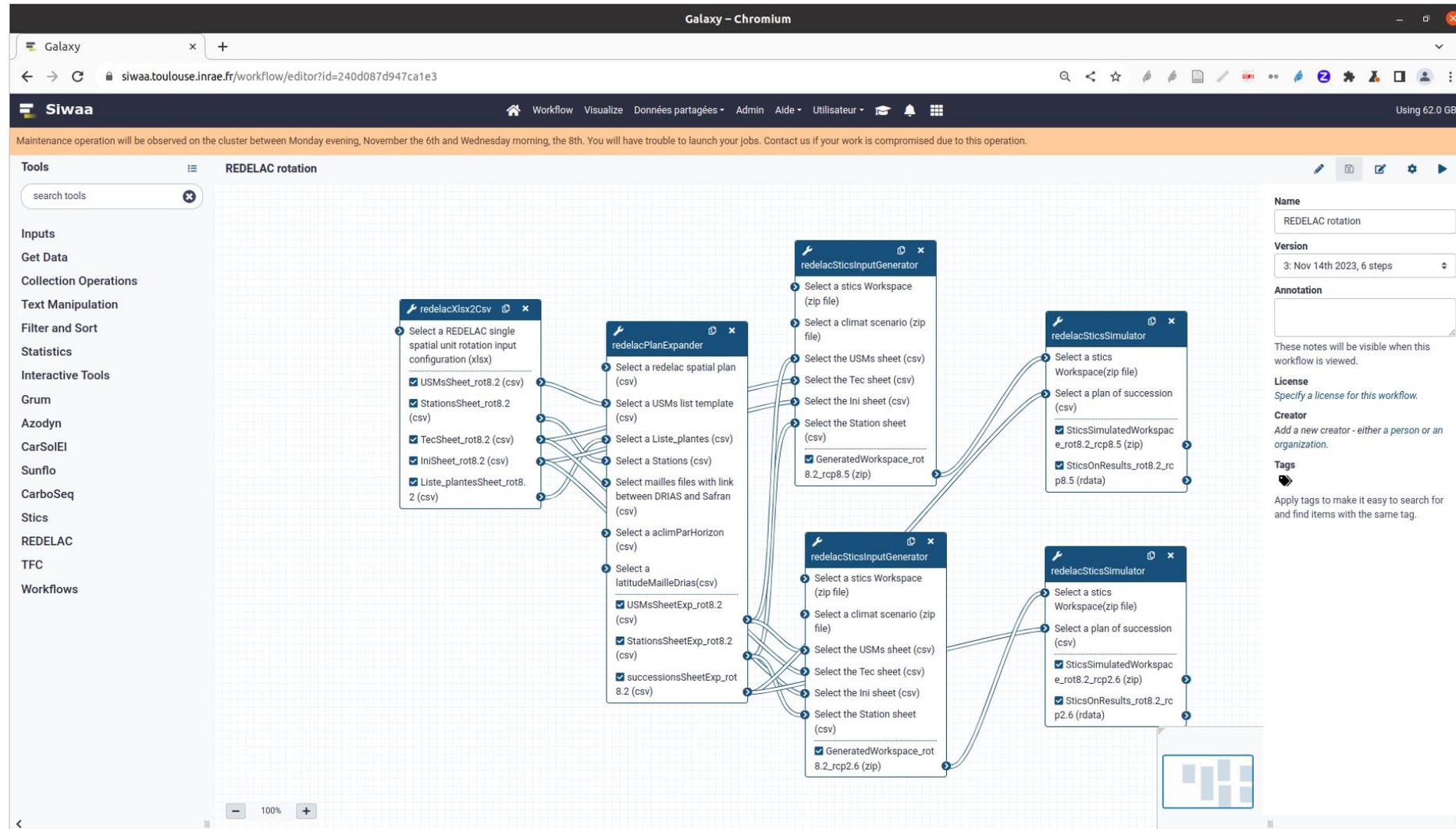
pipeline de CI/CD semi automatique

- *Mise à jour des containers.*
- *Validation*
- *Publication sur le registre*
- *Publication des outils REDELAC*
- *Installation*

The screenshot displays the GitLab CI/CD interface for a pipeline named 'Redelac / REDELAC Toolbox'. The pipeline is in a 'passed' state, triggered by Eric Casellas for commit 'affe8e8d' on the 'main' branch. The pipeline consists of several stages and jobs:

- Stage 1:** lint the inputGenerator, lint the plan expander, lint the simulator, lint the xlsx2csv, print-all-env-vars-job.
- Stage 2:** update on toolshed the inputGenerator, update on toolshed the plan expander, update on toolshed the simulator, update on toolshed the xlsx2csv.
- Stage 3:** build_docker, build_docker_sticsracks.
- Stage 4:** dockle, grype_cve, sbom, trivy.
- Stage 5:** dockle_sticsracks, grype_cve_sticsracks, sbom_sticsracks, trivy_sticsracks, test on siwaa the inputGenerator functional, test on siwaa the simulator.

Workflow par rotation



Les résultats

plan de simulation réduit 4%

- 12 plans de rotation x 4 UPC x 2 scénarios climatiques
- données et outils déployés sur SIWAA et rejouables
- exemple rot2.1a :
 - 5616 usms
 - 128 rotations simulées (30 ans)
 - avec 10 cpus en // = 16 minutes (génération des répertoires txt)



La suite et les perspectives

- Passage à l'échelle du territoire complet (x20)
- Ne pas exclure d'optimiser la génération des configurations atomiques de Stics
 - approche optimistic
- Déploiement et couplage du modèle AqualFarm
 - modèle individu centré
 - Python
- Partage des outils développés et refactoring
 - une démarche std concernant l'utilisation des containers
- Continuer à contribuer en utilisant et en proposant et en testant SticsR Packs
 - parallélisation
- Ajouter d'autres composants en amont, en aval pour capturer encore + largement le WF

