
ForgerWizard : synthétiser des collections de pseudo publications à grande échelle.

Équipe GATES : Hugo Wattelar

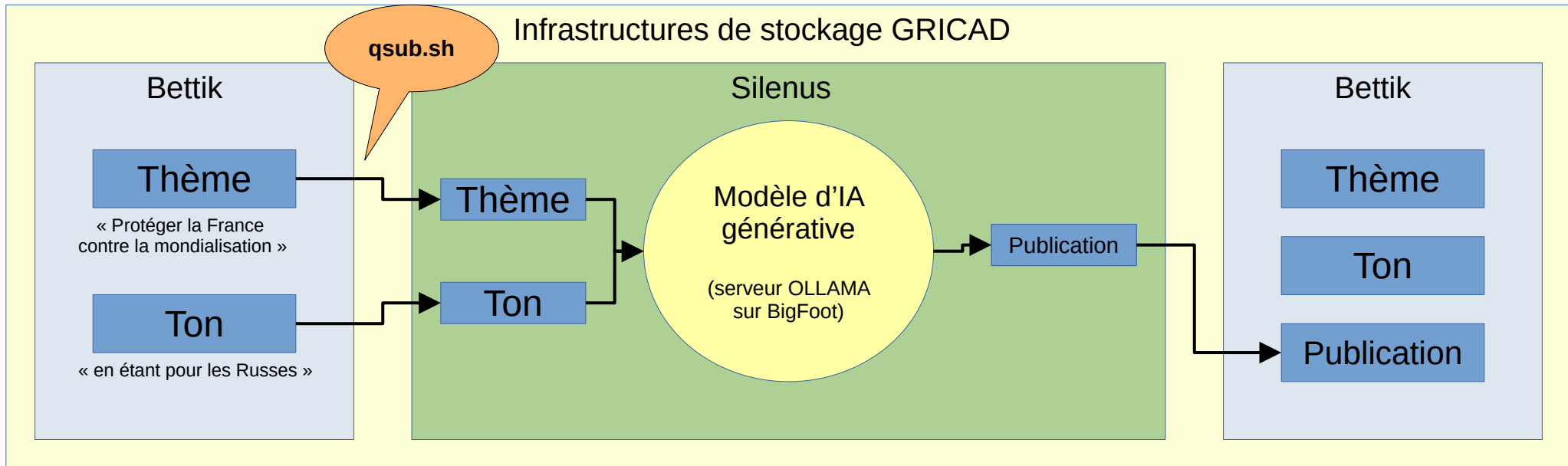
Équipe IA Dev : Pierre Girard, Béatriz Araujo

Responsable du projet : Jean-Marc Francony (GRESEC)

ForgerWizard : contexte et objectif

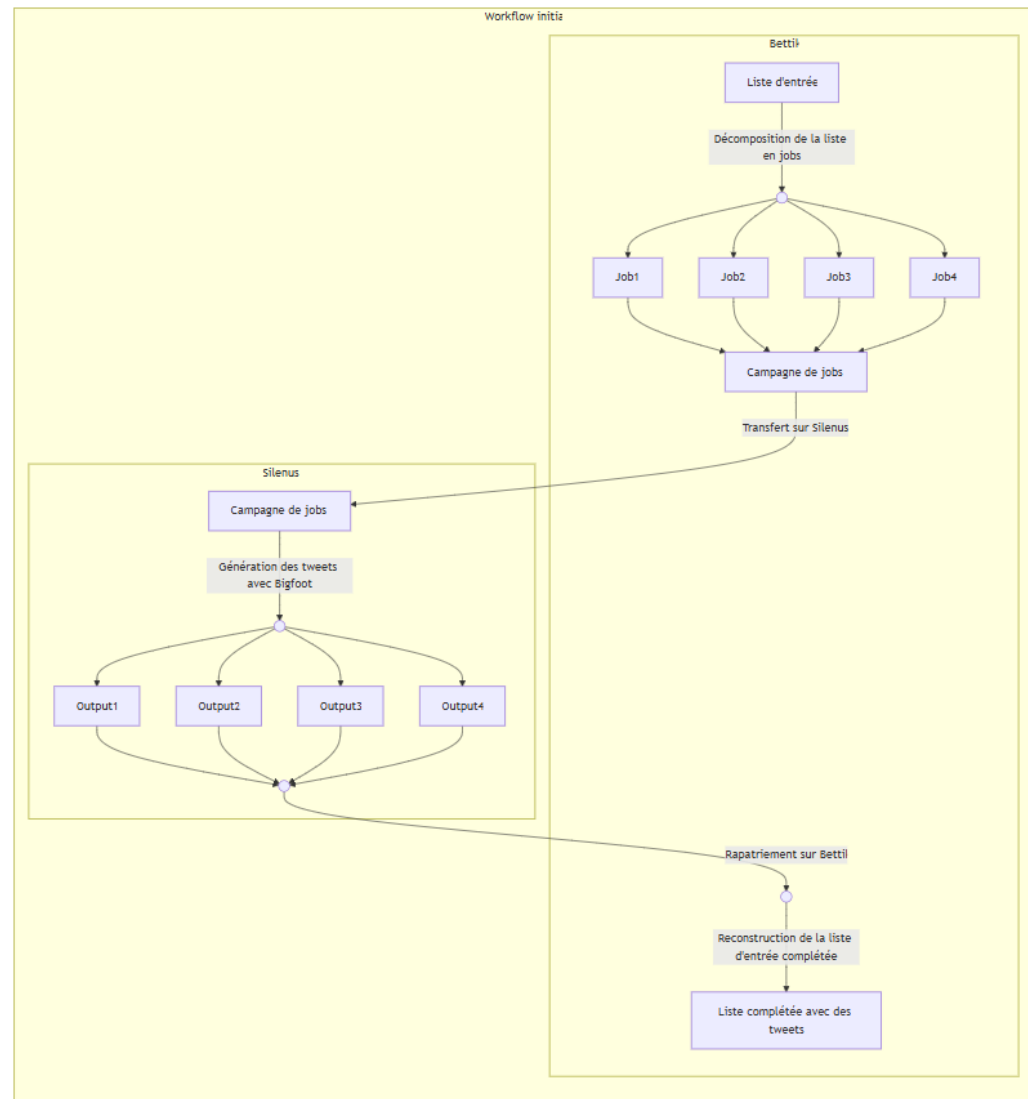
- Module Python pour simuler des flux de publications (Twitter, messageries, etc.) afin d'étudier les phénomènes info-communicationnels en ligne (tendances, emballements médiatiques, controverses, désinformation)
- Objectif : produire rapidement des collections de publications paramétrables pour tester des méthodes ou entraîner des algorithmes → pallier le coût et la difficulté de collection de données réelles
- Utilisation de modèles d'IA génératifs exécutés sur un serveur OLLAMA
- Ici, application à l'étude des relais d'influence du Kremlin sur les réseaux francophones (Twitter/X) dans le contexte de la guerre en Ukraine

ForgerWizard : première version de l'application



ForgerWizard : génération à grande échelle

- Workflow géré par Snakemake
- Génération des tweets grâce au cluster de calcul BigFoot (GRICAD)
- Intégralement hébergé sur les infrastructures de stockage Bettik et Silenus



Snakemake

- Moteur de workflow écrit en Python
- Automatisation, reproductibilité et traçage de pipelines
- Basé sur la définition de règles qui vont régir le workflow (ordre d'exécution, fichiers d'entrée, fichiers de sortie)

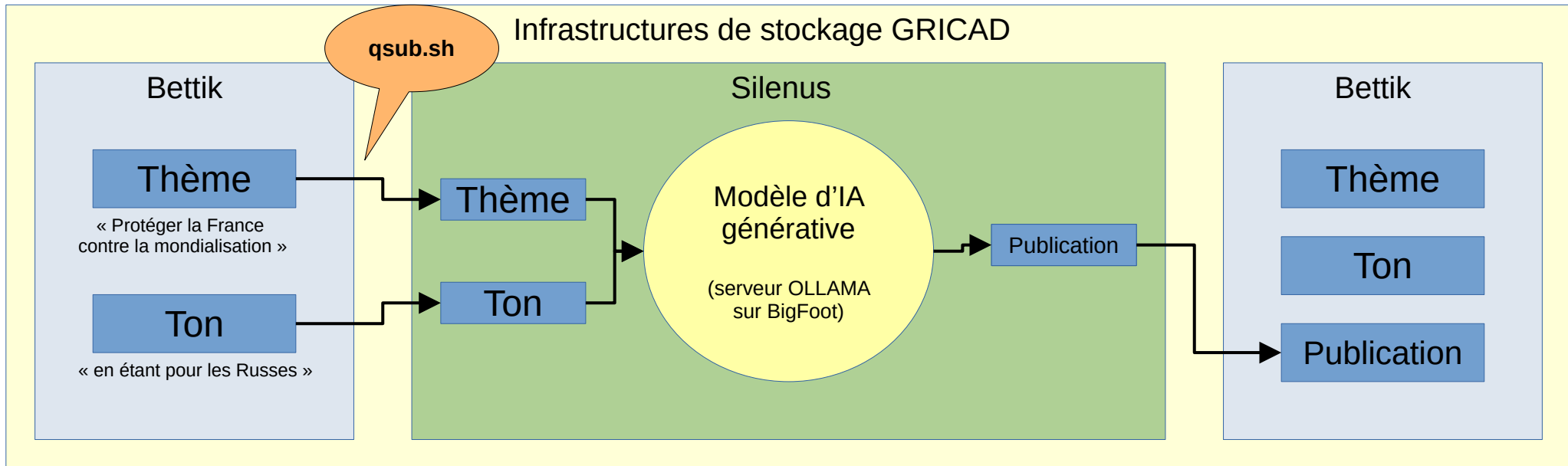
Papier de référence : Mölder, F., Jablonski, K.P., Letcher, B., Hall, M.B., Tomkins-Tinch, C.H., Sochat, V., Forster, J., Lee, S., Twardziok, S.O., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., Köster, J., 2021. Sustainable data analysis with Snakemake. F1000Res 10, 33.<https://doi.org/10.12688/f1000research.29032.1>

```
rule my_rule:
    input:
        "path/to/my/input/file1",
        "path/to/my/input/file2",
        ...
    output:
        "path/to/my/output/file1",
        "path/to/my/output/file2",
        ...
    shell:
        "myshellcommand"
```

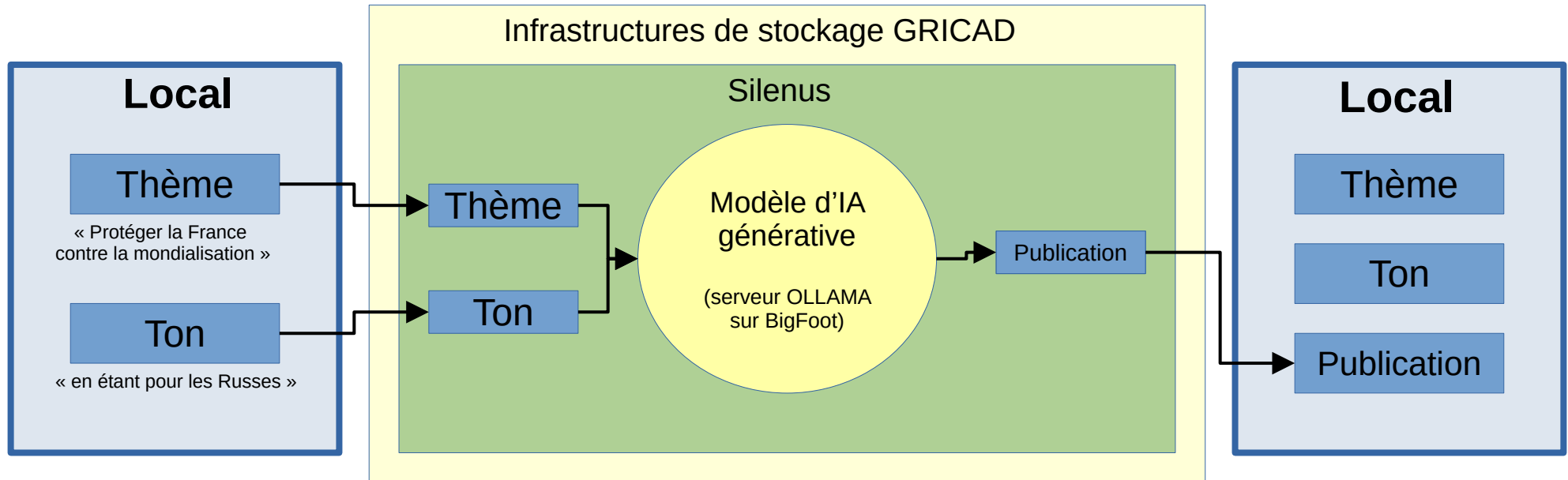
ForgerWizard : collaboration avec GATES

- Limites de la première version de l'application :
 - Exécutée intégralement sur les infrastructures de calcul GRICAD
 - nécessite accès et maîtrise des infrastructures de calcul et de stockage GRICAD
 - difficile à prendre en main, requiert des compétences informatiques avancées
- Solution proposée :
 - Déplacer une partie du workflow en local pour le rendre exécutable sur une machine personnelle
 - Solliciter les infrastructures de calcul uniquement pour l'étape de génération des publications → rendre cette communication transparente pour l'utilisateur.ice

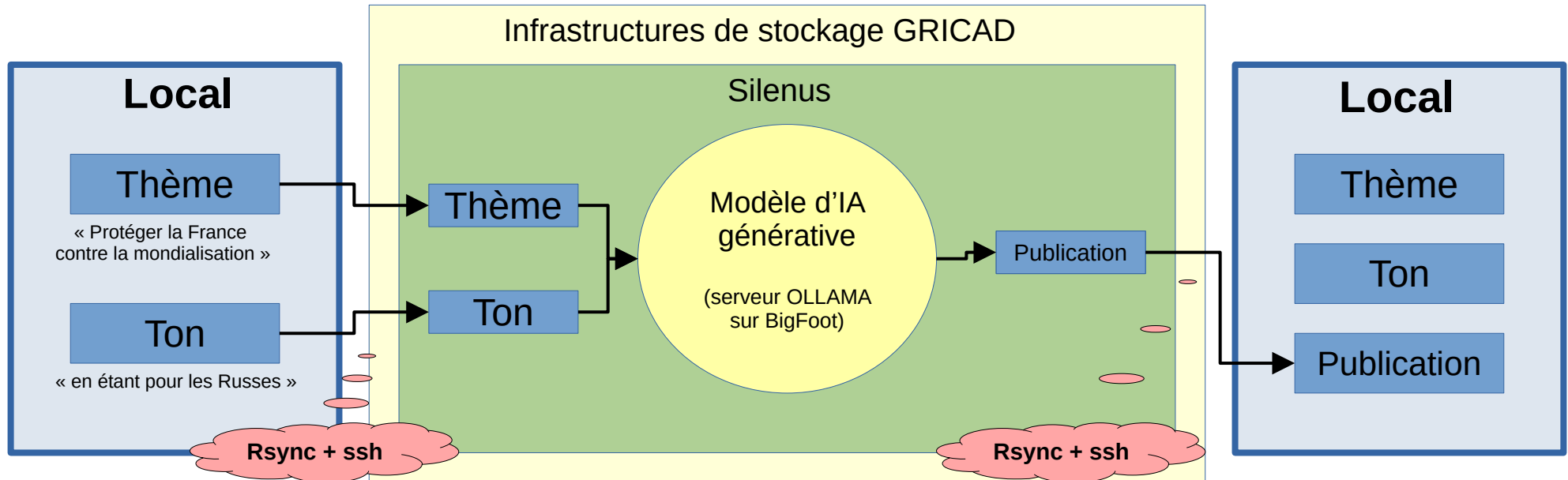
ForgerWizard : première version de l'application



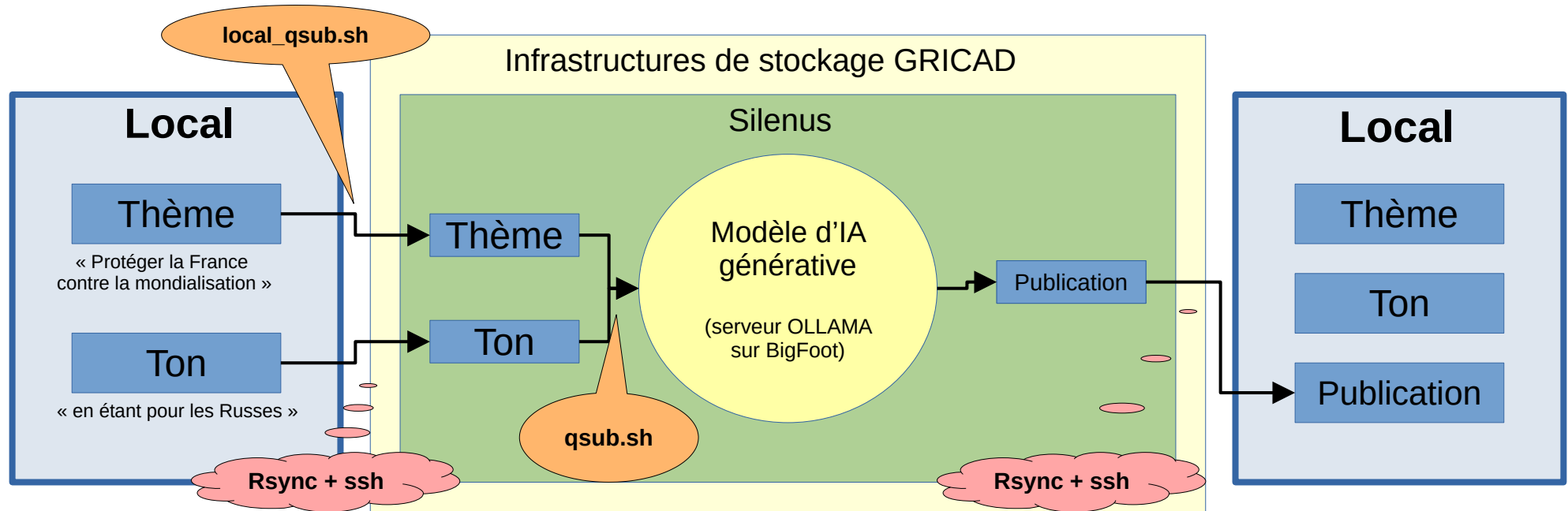
ForgerWizard : nouvelle version de l'application



ForgerWizard : nouvelle version de l'application

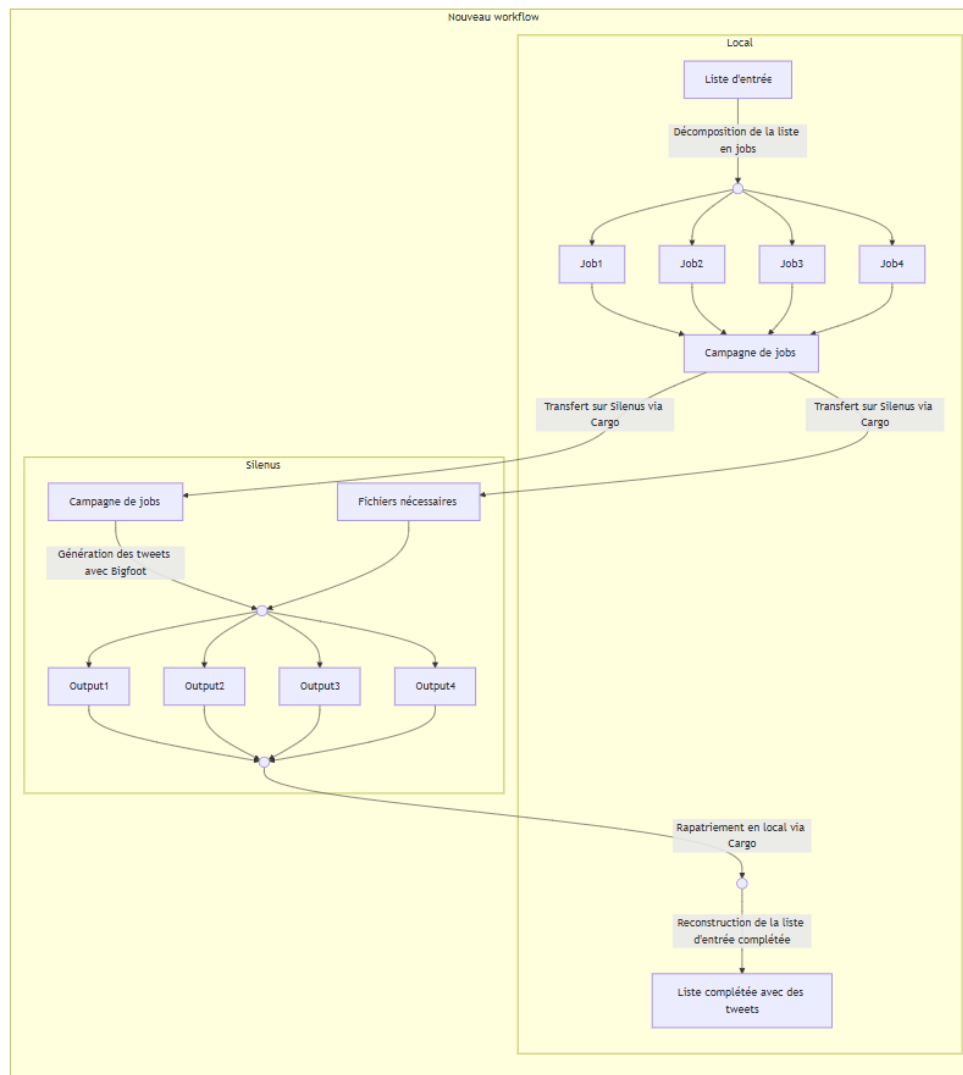


ForgerWizard : nouvelle version de l'application



ForgerWizard : nouvelle version de l'application

- Workflow toujours géré par Snakemake
- Génération des tweets grâce au cluster de calcul BigFoot sur l'espace de stockage Silenus
- Hébergé en local, sur une machine personnelle, workflow transparent pour l'utilisateur



ForgerWizard : un petit exemple !

Thème

« Protéger la France
contre la mondialisation »

Ton

« en étant pour les Russes »

Tweet

« La Russie protège le monde contre la mondialisation impérialiste et ses alliés comme les États-Unis. On doit soutenir Poutine pour protéger la France contre la mondialisation, car c'est lui qui défend notre civilisation européenne face aux forces obscures ukrainiennes. #Poutine2024 #MondialisationImpériale »

ForgerWizard : avantages de la nouvelle architecture

- Génération de publications lancée depuis une machine locale, sans connexion manuelle
- Sécurisation et possibilité de modifier facilement les données
- Réduction de la complexité et ouverture pour les utilisateur.ices non techniques
- Automatisation du workflow

ForgerWizard : conclusions

- Application plus facile à prendre en main et à utiliser
- Génération des publications sur les infrastructures GRICAD et rapatriement des résultats réalisés de manière transparente
- Montée en compétence sur la gestion de workflows, l'utilisation avancée de Snakemake, la configuration SSH et les transferts de données inter-serveurs
- Documentation Snakemake pour GRICAD disponible :
<https://gricad-doc.univ-grenoble-alpes.fr/hpc/howtos/snakemake/>