



SPGoO

Scientific Programming Group of Orleans

<https://spgoo.educandgames.org/>

<https://github.com/spgoo-ecpy/>

spgoo.ecpy@gmail.com



E. Eveno (CNRS)



C. Piatecki (UO)

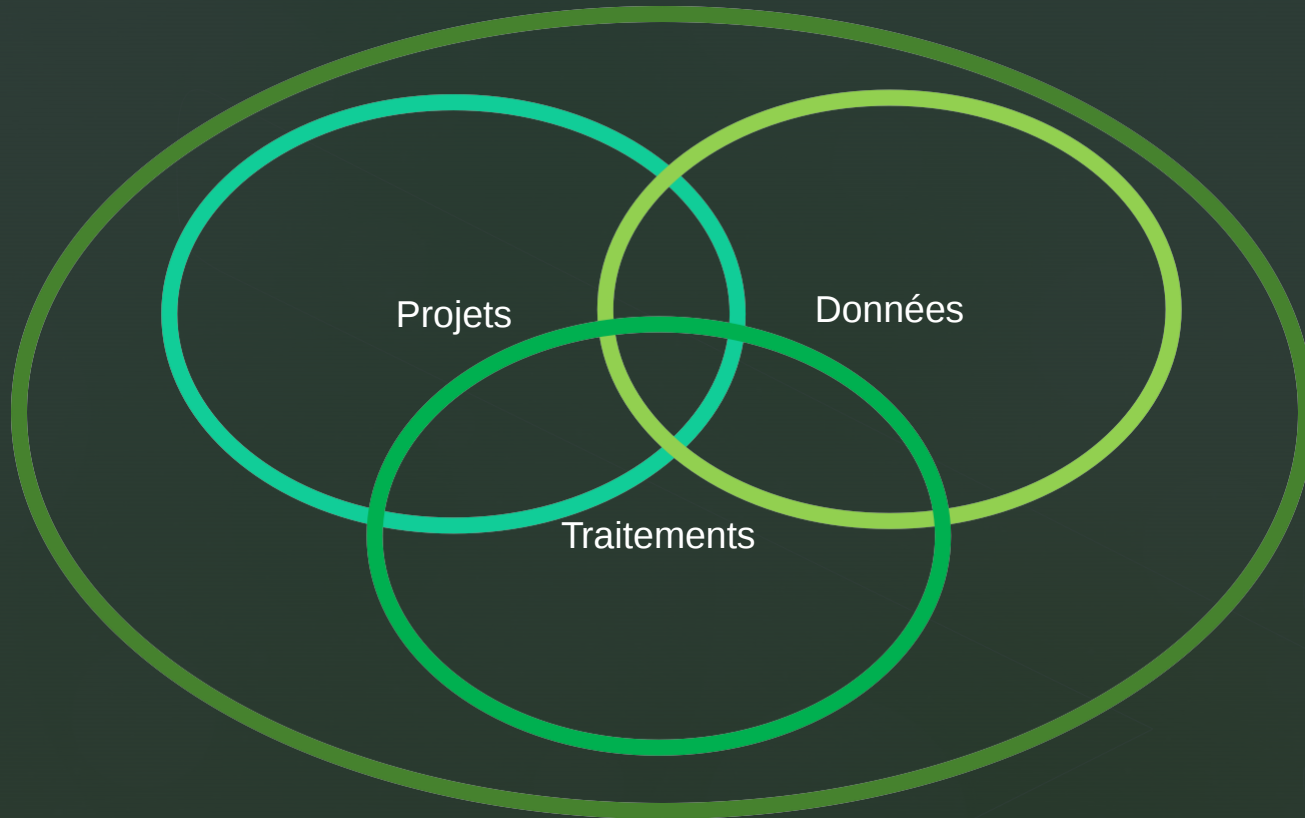


Y. Stroppa (CNRS)

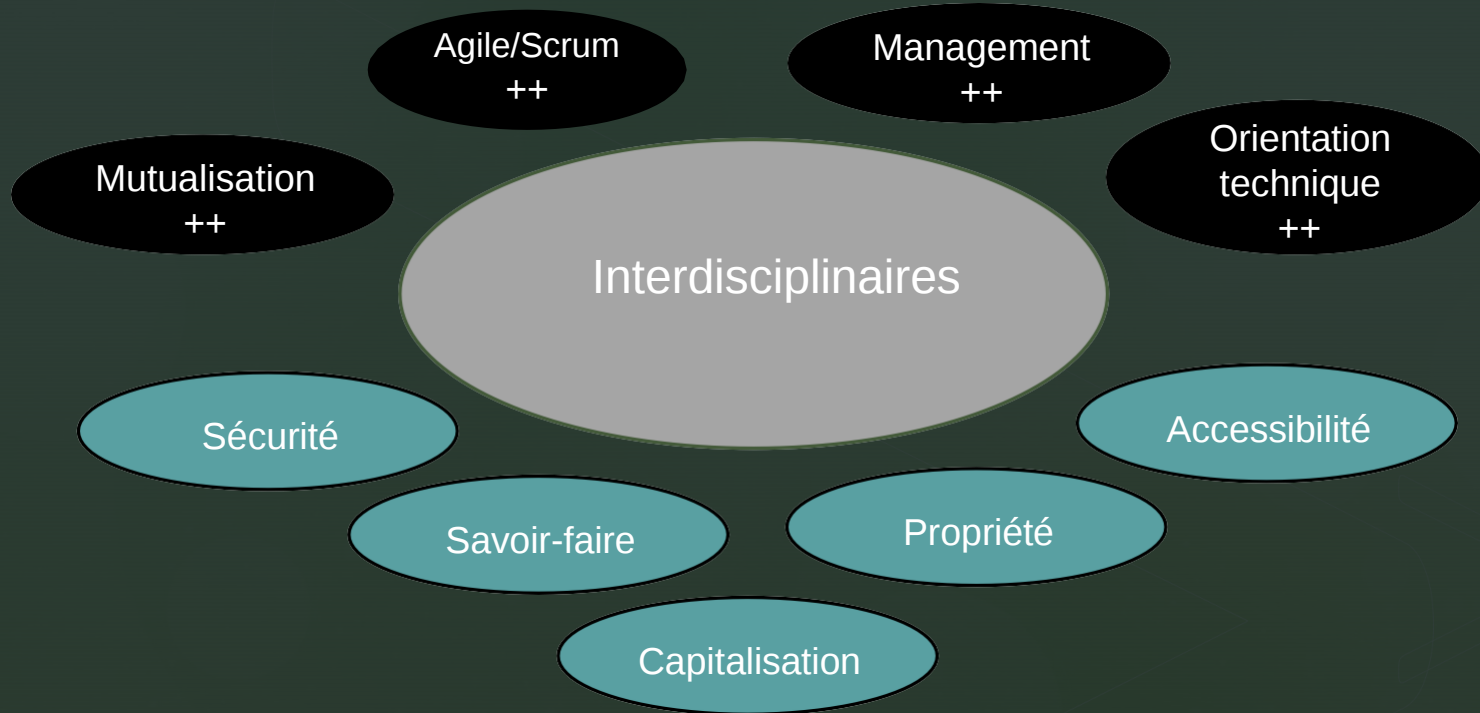


P. Krezel (CNRS)

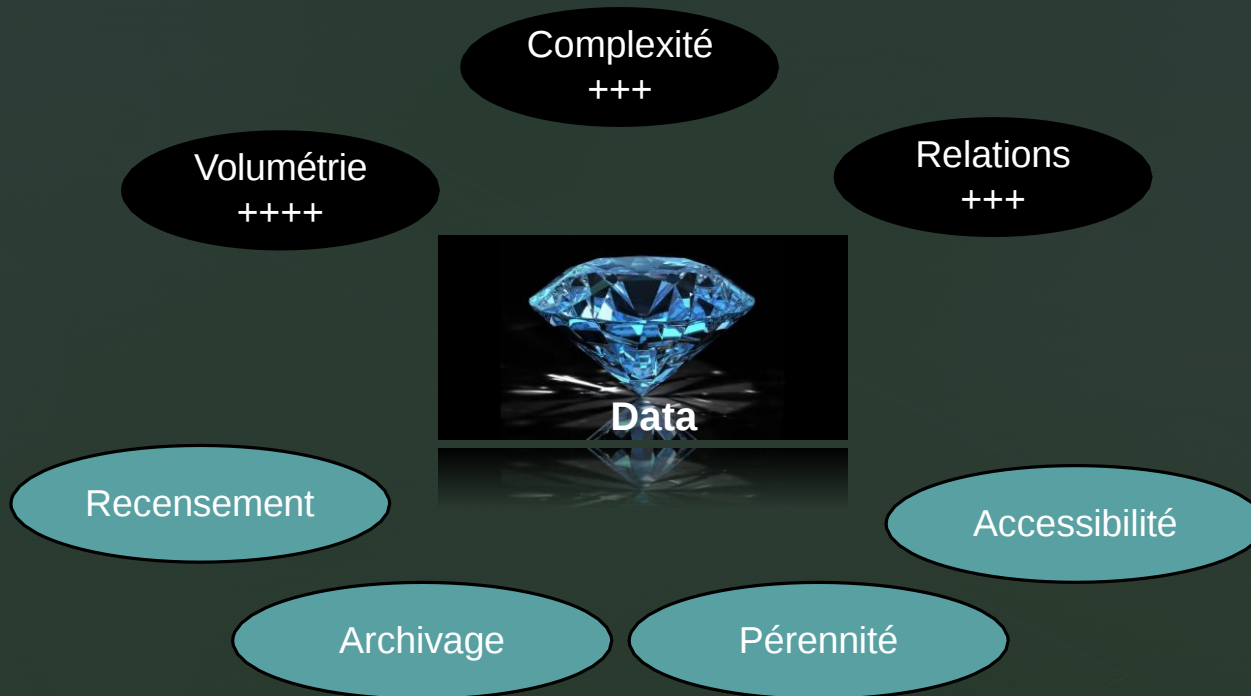
- ▀ Répondre aux besoins des chercheurs



▀ Répondre aux besoins des chercheurs : Projet/Méthodologie



▸ Répondre aux besoins des chercheurs: Données



MySQL

Postgresql

SqlServer

Mongodb

CouchBase

Neo4J

Cloud :
Openstack

Proxmox
Orchestration
K8S, Swarm

Virtualisation
Qemu,
KVM,
Hyper V
VMware

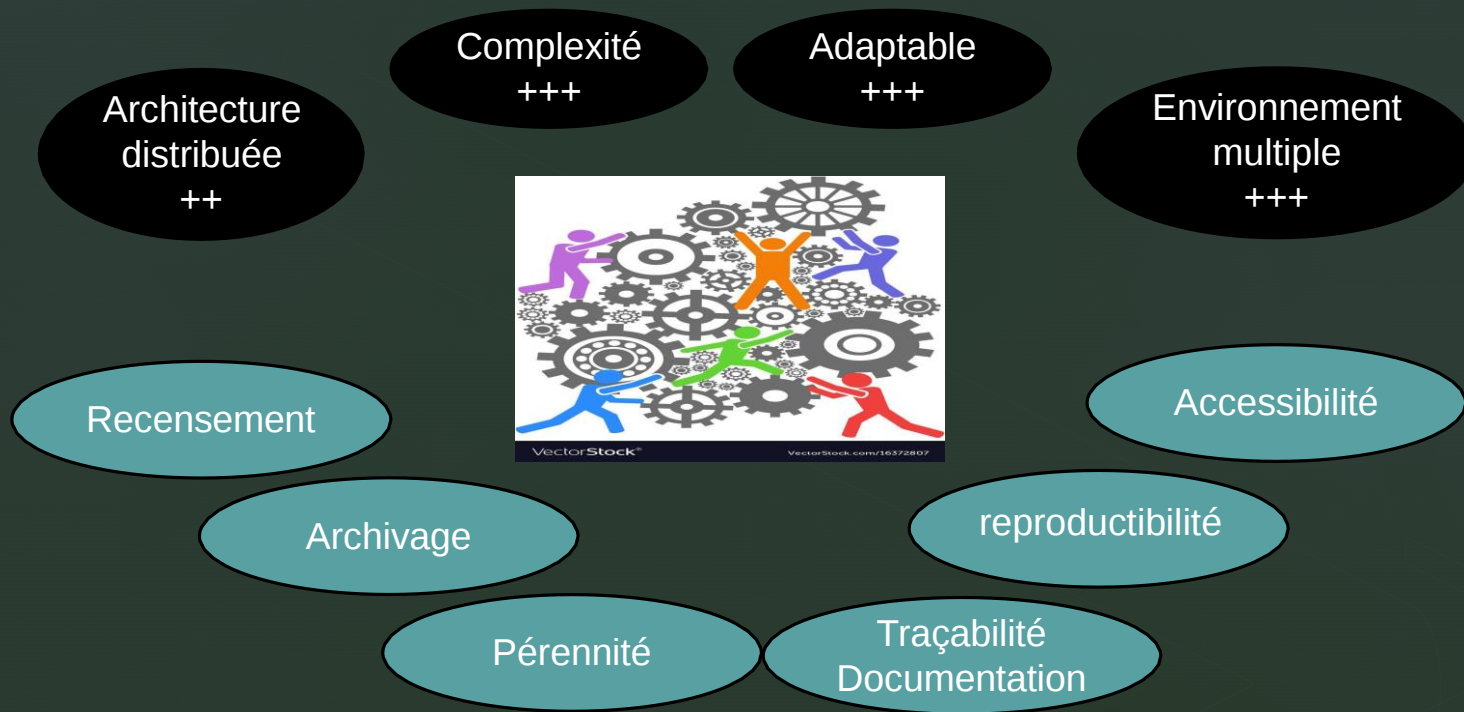
Conteneur
Docker
Podman
Cri-o

Versionning :
Git
Mercury

DevOps :
Github
Gitlab

Hébergeur :
OVH
AZURE
AWS
Google
CNRS ?

▀ Répondre aux besoins des chercheurs : Traitements



Langages :
C
C++
Python
Java
Perl
Cypher
Spring Boot
Matlab
R

Php
CSS
HTML
Typescript
Javascript
JSXGraph

Systèmes :
Ansible
Shell

Envs :
Jupyter
Eclipse
Netbeans
IntelliJ
PyCharm

Données/Traitements : Linguistique

Objets Métier

Locuteur

Interviewer

Nom
Prénom
Date_naissance
Pays_origine
Langues parlées
Langues maternelles
Sexe
.....

Nom
Prénom
Affiliation

Enregistrement

Localisation
Audio/Vidéo
Locuteur[s]
Interviewer[s]

Transcription[s]

Segments[s]

Arbres Lexicaux
Description Phonétique
Arbre[s] Syntaxiques
Description prosodique

Erreur[s]

Remédiation[s]
Correctifs

Annotation[s]

Méthode Verbo-tonale

Description

Système expert

Lexique

Entrée[s]

Graphie
Phonétique
Etymologie

Décomposition morphologique

sonore

Dictionnaire



Données/Traitements : Linguistique

- Projet FRAPéOR : corpus sonores Orléanais et élaboration d'un avatar conversationnel pour la remédiation phonétique
<https://frapeor.alimcorp.org>
 - Données : textuelles, numériques et sonores (volumétrie: 500Go)
 - Traitements : statistiques, Machine Learning/IA et plateforme Web
 - Méthodes : verbo-tonale à modéliser et à encapsuler dans l'avatar
 - Equipé projet : Internationale (France, Algérie, Argentine et Italie)
 - Accès : chercheur et public

Données/Traitements : Biologie

Objets Métier

Remodelage de l'ARN et régulation des gènes

- ❑ **Modèle: Terminaison de la transcription Rho-dépendante (mécanisme bactérien essentiel)**

- ❑ Mécanismes moléculaires
- ❑ Cibles physiologiques (gènes)
- ❑ Conservation phylogénétique
- ❑ Exploitation biotechnologique

ANR SMALLRHOSE(2013-18)

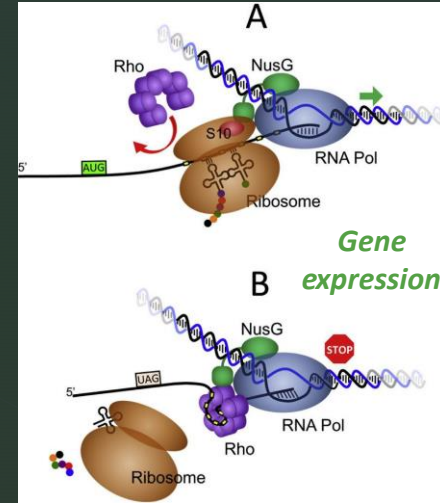
ANR INTEGHROME (2016-20)

ANR CONOCO (2019-23)

ANR HELISWITCH (2020-24)

ANR MYCHRO (2023-27)

(Approches: biochimie, biologie moléculaire & structurale, bio-informatique, microbiologie)



■ Données/Traitements : Biologie

- Projet IAML :
 - Données : textuelles, images, spectre (volumétrie: 500Go X NbrANR)
 - Traitements : statistiques, Machine learning et IA pour élaboration de modèles prédictifs ; plateforme Web pour annotations spécifiques des objets par les chercheurs et constitution d'une base de connaissance.
 - Equipé projet : Orléans
 - Accès : chercheur, Collaborateurs ANR

Données/Traitements : Chimie

Objets Métier

Organisme

Gène

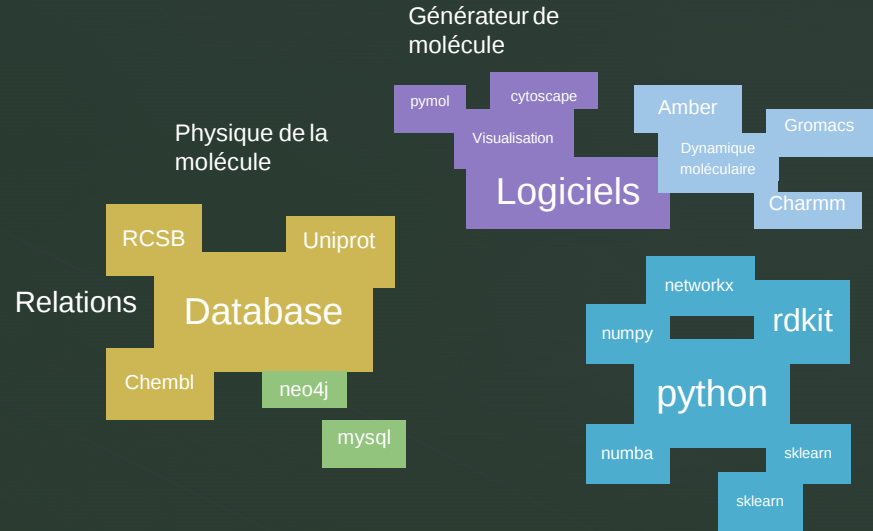
Protéine

Molécule

Composition
Propriétés
Attributs
Caractéristiques

Expérimentation

Objet de
l'expérimentation
Conditions de
l'expérimentation
Caractéristiques et
relevés



Comportement d'une molécule
dans un milieu donné

Evaluation du déplacement d'une
molécule sur une protéine

Pathologie/Maladie

Données/Traitements : Chimie

- Projet EmolGINE : plateforme de génération et d'extraction de molécules thérapeutiques .
 - Données : textuelles, numériques (volumétrie: 1To)
 - Traitements : statistiques, Machine Learning et IA, plateforme Web
 - Equipé projet : France
 - Accès : chercheur, Industrie

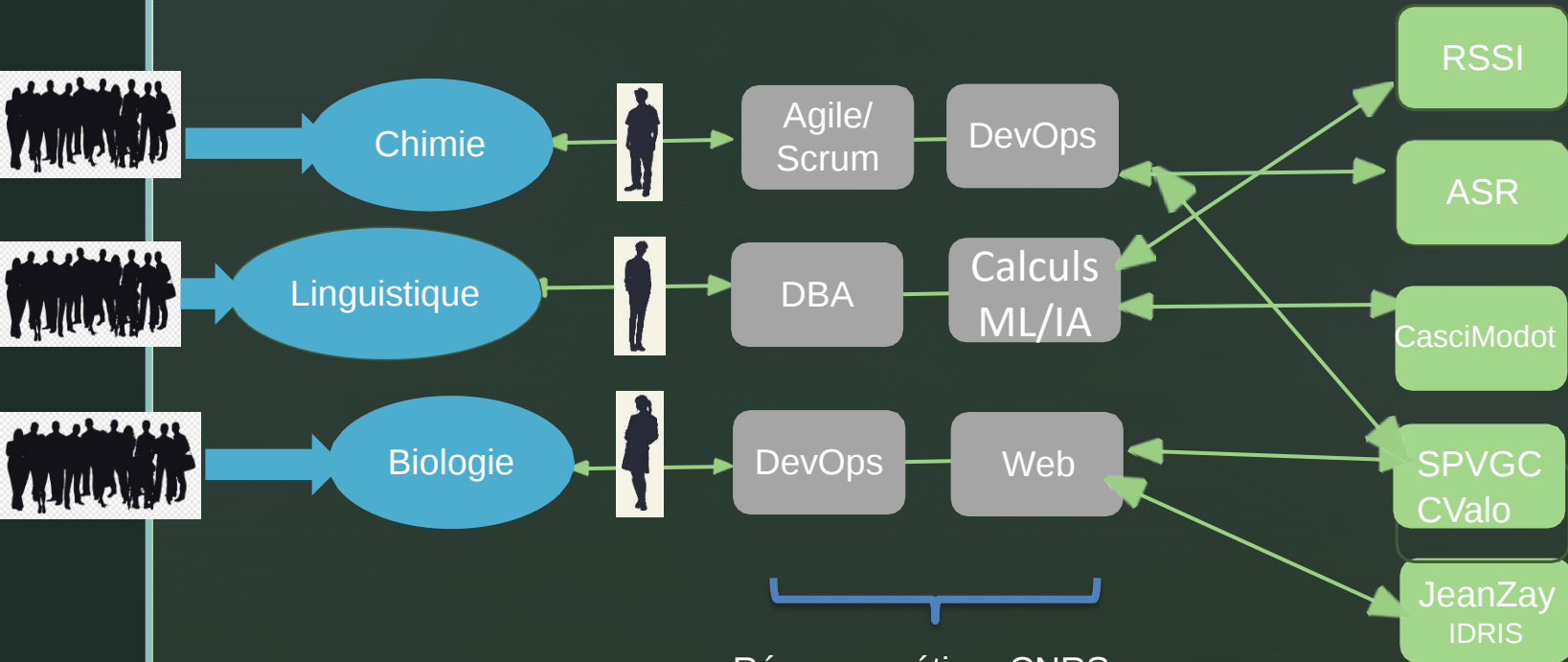
Verrous scientifiques

- Fusion de données hétérogènes (images, textes, sons).
- Intégration d'approches d'intelligence artificielle pour la modélisation.
- Gestion des différences méthodologiques entre disciplines.

Verrous organisationnels

- Les projets scientifiques sont composés d'acteurs de différentes disciplines.
- Localisés à des endroits différents (National et international) avec des tutelles différentes.

Contexte initial



Réseaux métiers CNRS
DevLogs, RBdd, Mathrice

...

▀ Problématique et justification



Pourquoi SPGoO ?



Points clés :

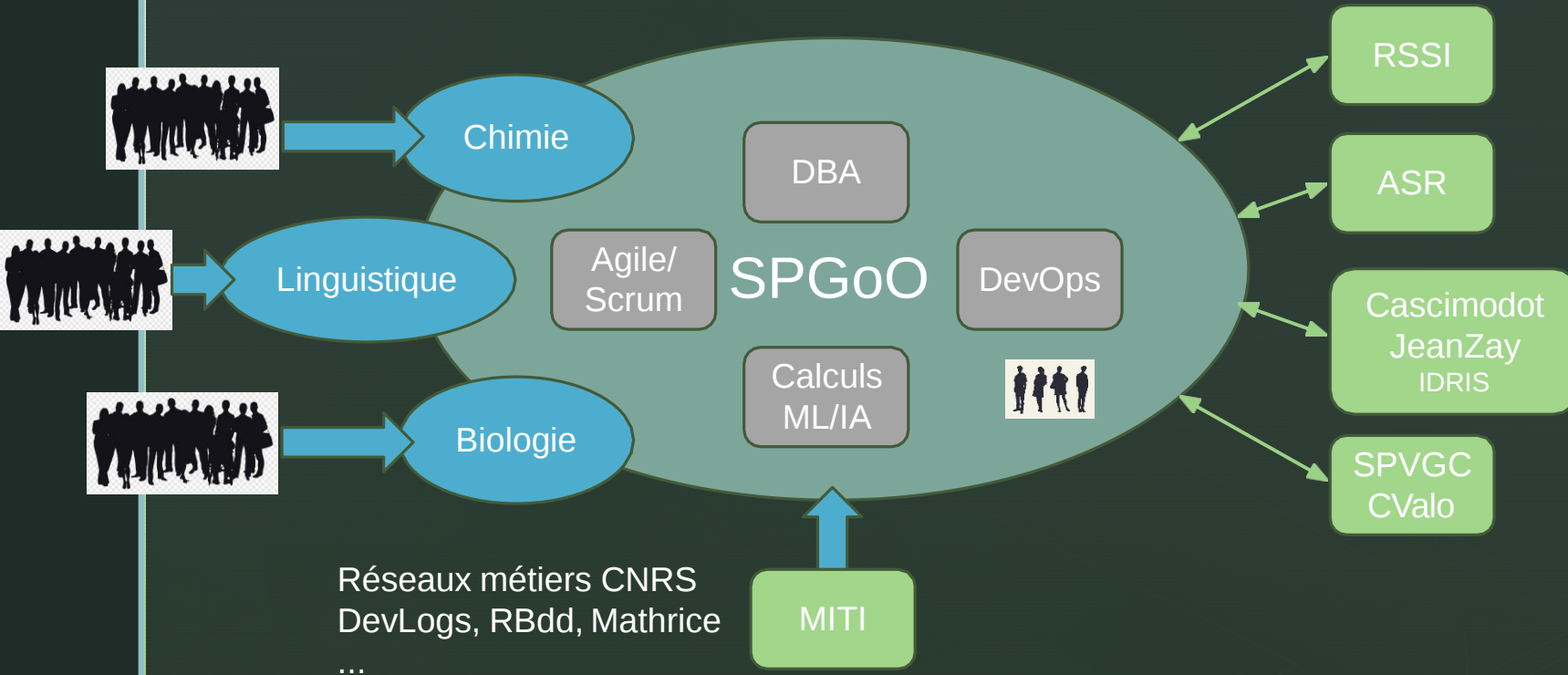
Augmentation de la complexité des données scientifiques

Ingénieur isolé au sein de leur laboratoire

Activités très diversifiées nécessitant des compétences sur des domaines connexes et variés

...

Proposition :



Création d'une cellule de mutualisation

- Centrée dans un premier temps sur les Labos : CBM, ICOA et LLL
- Avec un Comité d'orientation : DUs des structures impliquées
- Et comme objectifs :
 - Analyse et exploration de solutions innovantes dans le contexte scientifique chercheur ingénieur
 - Faciliter l'analyse, la modélisation et les interfaces utilisateurs des données issues de disciplines variées comme la biologie, la chimie, la linguistique, etc...
 - Mettre en place une plateforme centralisée et collaborative, conçue pour être accessible et conviviale, qui optimise les synergies interdisciplinaires et offre un lieu de travail collectif où les personnes peuvent interagir de manière fluide et efficace.

■ Méthodologies et Développement

- Principales Méthodes :
 - Développement d'une interface pour générer et annoter les données à partir de données brutes.
 - Intégration de pipelines de traitement des données avec Jupyter Notebooks.
 - Utilisation de Kubernetes et Docker/podman pour l'orchestration des "jobs".
- Outils d'IA :
 - Techniques de deep learning pour l'analyse prédictive et la modélisation non supervisée.
 - Support pour les frameworks de ML existants et importation de solveurs personnalisés.

▀ Valeur ajoutée pour les Laboratoires du Campus d'Orléans

- Bénéfices Concrets :
 - Soutien aux équipes dans le traitement de données complexes.
 - Mutualisation des ressources et expertise en analyse de données et IA.
- Accélération des projets grâce à un outil adapté aux exigences scientifiques modernes.
 - Alignement avec les politiques CNRS :
 - Favorise l'Open Science et le partage de données.
 - Intègre les normes de gestion éthique des données sensibles.

■ Résultats attendus

- À court terme (6 mois - 1 an) :
 - Déploiement initial de la plateforme.
 - Mise en place des fonctionnalités de base : pipelines, analyses statistiques, gestion des données.
- À moyen terme (1 – 2 ans):
 - Modèles prédictifs avancés en biologie, chimie et linguistique.
 - Aval du CDL: Pérennisation de la structure et ouvertures aux autres unités.
 - Création de publications interdisciplinaires et communication des découvertes.
- Impact à long terme : > 2 ans
 - Optimisation des collaborations interdisciplinaires.
 - Réplicabilité accrue et disponibilité des données pour d'autres chercheurs.

Plan de déploiement et ressources

- Étapes clés :
 - Phase de prototypage (6 mois) : déploiement au ? .
 - Évaluation et ajustement de l'infrastructure pour les besoins des utilisateurs.
- Défis anticipés :
 - Gestion de l'infrastructure (serveurs, clusters).
 - Maintien de la traçabilité et de la sécurité des données.
- Ressources nécessaires :
 - Mise à dispositions de serveurs, clusters, VM etc...
 - Collaboration avec les unités pour les compétences spécifiques.

■ Nos besoins

- Allocation de temps pour l'équipe mutualisé :
 - Afin de garantir l'efficacité et la qualité de notre contribution à SPGoO, nous demandons 20 % de temps de travail dédié pour chaque membre clé impliqué.
 - Justification : Cette allocation permettrait aux membres de se concentrer pleinement sur les objectifs du projet, tout en assurant une continuité dans leurs autres missions au sein des laboratoires.
- Espace de travail collaboratif : Un local dédié où nous pourrions nous réunir régulièrement pour échanger et travailler en équipe.
 - Justification : Un espace commun favorise une meilleure coordination, facilite la résolution rapide des problèmes techniques et méthodologiques, et stimule les échanges interdisciplinaires en présentiel.

■ Nos besoins

- Accès à des serveurs informatiques : Pour mener à bien les développements et les traitements de données complexes, nous avons besoin de machines dédiées accessibles (à distance).
 - Justification : Ces machines offriront un environnement sécurisé et performant pour les analyses intensives, permettant d'intégrer et de tester les développements dans un cadre adapté, sans impacter les infrastructures en production.
- Alignement stratégique : Nous pensons que ces ressources permettront à SPGoO de s'implanter efficacement au sein du campus, avec un impact positif direct sur la productivité et l'innovation dans nos unités.
- Avoir des relations privilégiées avec les acteurs extérieurs : RSSI, ASR
- Constitution d'un WIKI pour capitalisation des connaissances et diffusion de bonnes pratiques.
- Accès Gitlab/hub ou autres solutions pour garantir les opérations de DevOPS (automatisation des déploiements)



Merci pour votre
attention



Disponible pour
vos questions