

5 avril 2023

Formation Science Ouverte

Gestion et ouverture des données de la recherche à EcoFoG

Gaëlle Jaouen

EcoFoG = Ecologie des Forêts de Guyane



Qui sommes nous?

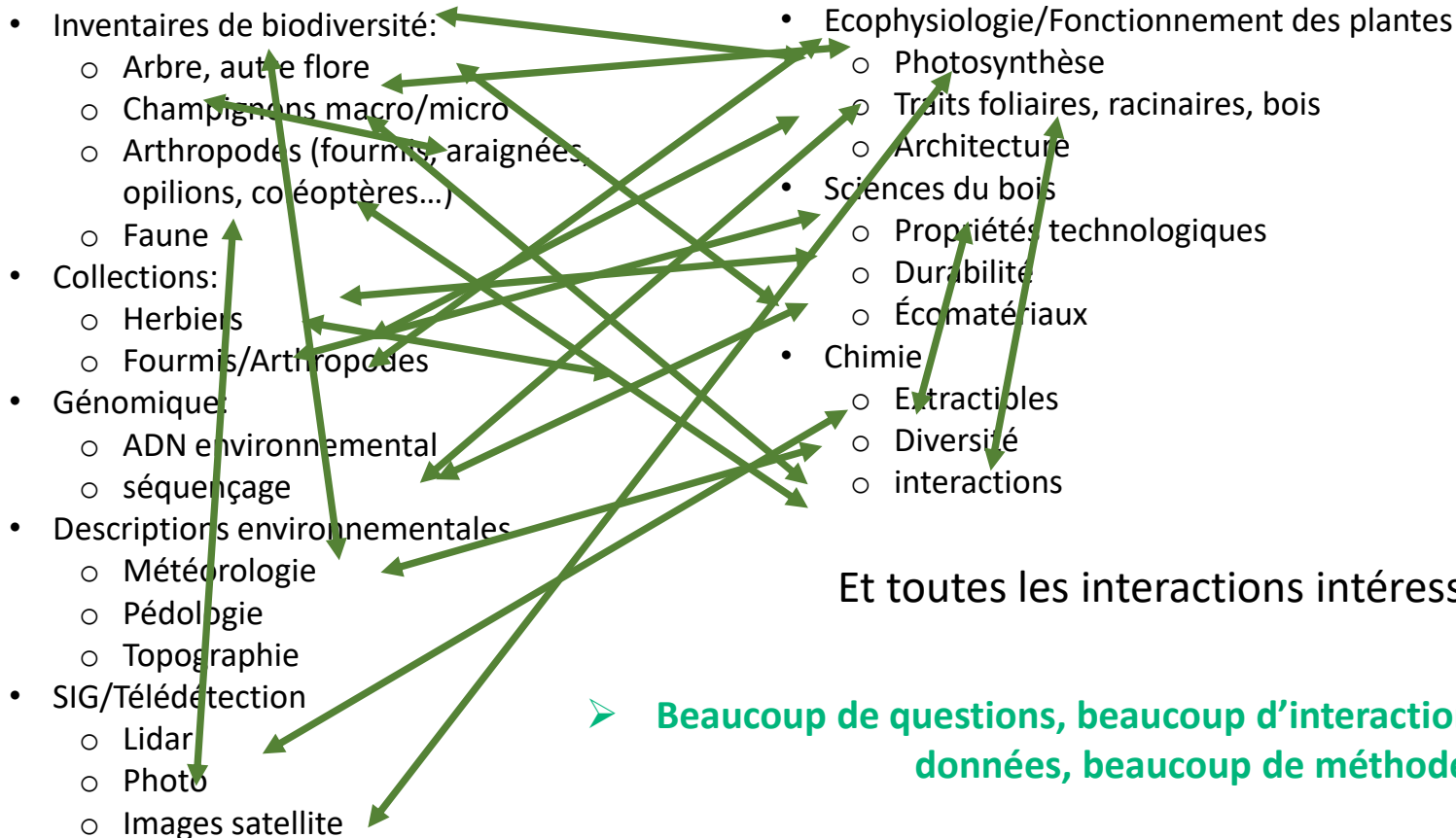
- UMR
 - 4 Organismes de Recherche (AgroParisTech, CIRAD, CNRS, INRAE)
 - 2 Universités (Université des Antilles, Université de Guyane)
- Effectifs : ~ 30 chercheurs et Enseignants-Chercheurs, ~ 30 ITA, ~ 15 doctorants (en moyenne)
- Élément majeur du LabEx CEBA (Centre d'Etude de la Biodiversité Amazonienne)

Que faisons-nous?

Ecologie et **Sciences des Matériaux** pour aborder 2 grandes questions:

- Comprendre les **relations entre biodiversité et fonctionnement des écosystèmes forestiers tropicaux**, exploités ou non, en évolution sous les pressions climatiques et anthropiques
- Susciter l'innovation dans la **valorisation des ressources forestières** sous contraintes d'utilisation liées au milieu tropical humide

Que produisons-nous et manipulons-nous comme données?



Problématique

- Grande diversité de données
 - Objets biologiques



Problématique

- Grande diversité de données
 - Objets biologiques
 - Types de données (tabulées/sorties de capteurs, images, échantillons physiques...)
 - Volumes très variables (observations<<<SIG/Lidar)
- Données anciennes, manuscrites, mal conservées : trouver comment numériser sans perdre de valeurs, d'informations
- Diversités d'acteurs, d'organismes, de besoins
- Besoin de faire interagir toutes ces données : trouver des connexions, utiliser des référentiels
- Organiser et faire fonctionner en interne + ouvrir et publier : mon rôle d'IR

Concrètement

- Faire vivre le patrimoine
 - Agrégation des sources par thématique scientifique
 - Numérisation et OCR à l'aide d'IA pour les données manuscrites
 - Documentation des données: métadonnées, traçabilité, dictionnaires de données
 - Bancarisation: conception, construction de bases de données relationnelles
 - Mise en réseau : interconnexion des BD pour interrogations croisées
- Ouvrir les accès internes
 - Interfaces clique-bouton (Access) sur mesure
 - Scripts R d'extraction/manipulation des données pour utilisateurs avancés
 - Connexions SIG (Qgis)
 - Gestion fine des droits (accès, lecture, écriture etc...) et des besoins selon les utilisateurs

Concrètement

- Ouvrir les accès externes
 - Développement de plateformes d'accès sur demande aux données en ligne (inventaires et carto)
 - DataPapers dès que possible
 - Dépôt des données dans des entrepôts thématiques (géocatalogues divers, GBIF, GenBank, AntWeb...)
 - Création d'un DataVerse d'unité avec arborescence thématique et espaces projets
- Cadrer et faire circuler les informations
 - PGD par thématique scientifique, avec les acteurs concernés
 - PGD d'unité et politique des données de l'unité
 - Charte de partage des données CEBA
 - Livret d'accueil pour les nouveaux arrivants (en projet)

Concrètement

- Sauvegarder!
 - Les BD sont sur un serveur centralisé sauvegardé chaque jour, semaine, mois, année localement
 - Chaque semaine, des sauvegardes automatiques se font automatiquement vers la métropole
- Former/Se Former
 - Accompagner les étudiants/doctorants/agents pour la gestion des données
 - Anticiper la gestion des données dans les montages de projets
 - Discuter avec le DU pour les orientations de l'unités sur les DR
 - Continuer à apprendre et découvrir les nouveaux outils utiles!

Des questions?



Contact

Gaëlle Jaouen

gaelle.jaouen@agroparistech.fr