

CDD/Post-doc de 12 mois auprès de l'infrastructure RéGEF pour mettre en place une stratégie de gestion de ses données

Candidature sur le portail emploi du CNRS (1^{ère} échéance à laquelle les candidatures seront évaluées : 21/02/2021) :

<https://emploi.cnrs.fr/Gestion/Offre/Default.aspx?Ref=UMR7358-RAPPIK0-004>

Date d'embauche prévisionnelle : 01/04/2021

Contacts et renseignements :

- Raphaël Pik, directeur de RéGEF : raphael.pik@univ-lorraine.fr | 06 03 49 22 00
- Pierre-Yves Arnould , IR bases de données OSU OTELo : pierre-yves.arnould@univ-lorraine.fr

Missions :

Le groupe de travail chargé de réfléchir à la stratégie de gestion des données de RéGEF, recrute un.e post-doctorant.e (sur un support CDD de 12 mois) pour mettre en place une stratégie de gestion de ses données. Il s'agira de coordonner les actions du groupe, assurer un relais expert vers les plateformes analytiques, et synthétiser des propositions concrètes de mise en place des plans de gestion des données types.

En suivant les nombreuses initiatives nationales et internationales autour de la « Science Ouverte », l'infrastructure RéGEF ouvre en 2021 un chantier important pour organiser la gestion des données produites par ses plateformes analytiques et préparer à leur archivage F.A.I.R. (Findable, Accessible, Interoperable, Reusable). Ce chantier sera mené par un groupe de travail constitué de spécialistes de la gestion des données, de responsables des réseaux de RéGEF, de représentant des utilisateurs (CS de l'INSU), et de représentant de l'infrastructure de recherche DataTerra. Ce groupe travaillera au cours de l'année 2021 en interaction proche avec les plateformes analytiques du réseau RéGEF, via les réseaux spécialisés d'instruments.

Le réseau RéGEF (Réseau Géochimique et Expérimental Français, www.regef.fr) est constitué de 150 plateformes analytiques dans le domaine des géosciences et sciences de l'environnement, implantées dans les UMRs de l'INSU-CNRS (www.insu.cnrs.fr). Ces plateformes analytiques produisent des données qui sont utilisées pour quantifier le fonctionnement du système Terre par une vaste communauté scientifique, de la cosmochimie, à la Terre interne, jusqu'à sa surface et ses évolutions récentes sous les forçages anthropiques et climatiques. Ces données sont très diverses (élémentaires, isotopiques, globales ou ponctuelles, sous formes d'images 2D ou 3D, etc ...) et souvent particulières à chacun des douze réseaux d'instruments spécialisés qui forment RéGEF. Ces données qui sont : (i) très nombreuses (> 10⁵/an) et très diverses, (ii) commanditées par des scientifiques de diverses spécialités, et (iii) mesurées sur des objets eux aussi très différents (roches, minéraux, météorites, sols, eaux, sédiments, bactéries, gaz, ...), sont hétéroclites et s'éloignent de la norme, ce qui les fait qualifier de « longue traîne ». A la différence des grandes séries de données produites par les activités scientifiques d'observation de la Terre et de l'Univers, leur gestion n'est pas

confrontée à des enjeux sur les capacités de stockage, mais a des enjeux sur les procédures de renseignement des métadonnées associées aux échantillons traités (souvent uniques) et aux méthodes analytiques utilisées pour les documenter (souvent nombreuses).

Activité :

Il s'agira plus précisément de :

- mettre en place une réflexion avec le groupe de travail sur la stratégie globale de RéGEF pour l'archivage des données produites par les plateformes dans l'écosystème spécifique : commanditaires-utilisateurs | plateformes | IR Data Terra | tutelles
- définir une stratégie et un cahier des charges pour mener les interactions spécifiques avec chaque réseau de RéGEF
- établir avec les différents réseaux de RéGEF le data management plan pour chacun en fonction de la spécificité de leurs données et de leurs métadonnées associées.
- établir avec les acteurs des géosciences et des sciences de l'environnement une stratégie vis-à-vis d'une nomenclature et d'un référencement unique/universel des échantillons (qui constitue dans l'exploitation ultérieure des données de « longues traine » un des points critiques à traiter de façon globale à l'échelle de la communauté).
- réfléchir à la mise en œuvre technique de cette stratégie à l'échelle de RéGEF et proposer des solutions aux tutelles.

Compétences :

Pour mener cette action de coordination auprès du groupe de travail et interagir précisément avec les plateformes et réseaux de RéGEF les candidat.e.s auront :

- un doctorat en géosciences ou sciences de l'environnement
- une forte expérience analytique acquise auprès d'une ou (mieux) plusieurs plateformes de RéGEF au cours du doctorat.
- une motivation importante pour cet aspect nouveau et important de la gestion de la science moderne et de ses données. Cette offre se positionne au niveau post-doctorat et (au-delà du cahier des charges opérationnel) sera finalisée idéalement par une publication spécialisée focalisée sur la problématique de ces données dites de « longue traine ».
- maîtrise de la langue française

Contexte de travail :

Le.a candidat.e sera basé.e au laboratoire CRPG à Nancy pour pouvoir interagir efficacement avec la cellule opérationnelle de RéGEF (Raphaël Pik - directeur | Jean-Yves Mougel - ingénieur projet | Pierre-Yves Arnould - Ingénieur de Recherche bases de données). Les missions demanderont une très forte interaction avec les différents réseaux et plateformes de RéGEF qui, en fonction de l'évolution des conditions sanitaires au cours de 2021, sera réalisée soit à distance en visio-conférence, soit pour partie en présentiel sur les différents sites d'implantation des plateformes sur le territoire.