

## Formation sur la qualité, la publication et l'utilisation de données sur la biodiversité

### Description de l'événement :

Cette formation vise à donner aux participants un aperçu des différentes façons de préparer, de publier et d'accéder aux données sur la biodiversité, via le réseau, les standards et les outils GBIF.

Lancé par l'OCDE en 2001, le système mondial d'informations sur la biodiversité (Global Biodiversity Information Facility, ou GBIF) est l'infrastructure la plus importante au monde proposant un accès libre et gratuit aux informations sur la biodiversité, afin de soutenir la science, les politiques de conservation et le développement durable.

Ces données (plus de 620 millions en septembre 2016) décrivent où et quand des centaines de millions d'organismes ont été observés ou récoltés, proposant ainsi une cyber-infrastructure pour affiner la connaissance de la répartition des espèces.

Elles sont accessibles librement, gratuitement et universellement, tout en restant attachées de façon permanente à leur source d'origine, ce qui assure visibilité et contrôle à leur fournisseur.

### Nombre de participants et public cible:

Un maximum de 20 participants. Nous visons la participation de scientifiques, de gestionnaires de bases de données naturalistes ou de techniciens ayant une expérience dans les sciences biologiques et la biodiversité, et dans l'utilisation et la gestion des données numériques sur la biodiversité.

**Lieu :** Montpellier (salle encore indéterminée)

**Frais :** Formation gratuite (financée par le LabEx CeMEB) ; les repas et l'hébergement des stagiaires sont à la charge de leur unité.

### Les participants travailleront sur un ou plusieurs jeux de données qu'ils auront apportés :

Prérequis pour les jeux de données : pour pouvoir publier des données dans le réseau GBIF, il faut au moins cinq champs obligatoires (identifiant unique pour chaque occurrence, nom scientifique, date de collecte/observation, pays, type de donnée (observation, spécimen, fossile...)) ainsi qu'une courte description du jeu de données et au moins deux personnes contacts (administratif et technique) qui pourront répondre à d'éventuelles questions sur les données une fois les jeux de données publiés. En théorie, il n'y a pas de taille minimale ni maximale pour les jeux de données, mais pour faciliter la manipulation des données durant la formation, il faudrait des jeux de données contenant au plus quelques milliers d'enregistrements (des datasets de quelques dizaines ou centaines de ligne sont bien sûr tout à fait publiables aussi).

Il sera également peut-être nécessaire de déclarer de nouveaux fournisseurs de données (laboratoire, UMR, département de l'Université...) préalablement à la formation afin de pouvoir publier les données dans le GBIF ; pour ce faire, les participants qui le souhaitent peuvent le faire en utilisant le [formulaire en ligne](#) (en anglais) du site [www.gbif.org](http://www.gbif.org) ou en remplissant le document ci-joint (en français). Attention à ne pas faire de déclaration en double si des participants viennent de la même structure.

**Equipement** : Chaque stagiaire devra venir avec son ordinateur et vérifier que Java est bien à jour, et si possible installer le logiciel de nettoyage des données OpenRefine avant la formation (une aide pourra être donnée pour l'installation pendant la formation mais cela nous fera gagner du temps s'il est déjà sur la plupart des ordinateurs) ; la version d'OpenRefine à télécharger [sur cette page](#) est la 2.6-beta.1.

**Inscription** : Demande d'inscription à faire auprès d'Emmanuelle Artige, coordinatrice du Groupe « Collections » du LabEx CeMEB par mail : [emmanuelle.artige@inra.fr](mailto:emmanuelle.artige@inra.fr)

Le nombre de place étant limité, cette formation est prioritairement ouverte aux agents des unités appartenant au LabEx CeMEB et une place est réservée pour une personne de l'OSU OREME.

## PROGRAMME DE FORMATION

**Jeudi 23 février 2017 : Introduction de la formation - Qualité des données (principes généraux et exercices).**

09 :00 - Introduction et aspects pratiques par [l'équipe GBIF France](#)

Informations sur le contexte et les raisons de faire cette formation. Tour de table : présentation des participants, examen du programme et attentes des participants. Informations pratiques concernant les pauses-café et déjeuners, le calendrier, l'accès internet, les ressources, etc.

09:15- Présentation du GBIF France par [l'équipe GBIF France](#)

09:30 - Une introduction à la qualité et l'adaptation des données à leur utilisation par [l'équipe GBIF France](#)

Les concepts de base liés à la qualité des données et à leur aptitude à être utilisées seront présentés aux participants. Comprend 15 min. de discussion.

10:30 - Pause café/thé

11:00 - Exercice pratique de mise en forme des données : manipulation des tableurs GBIF par [l'équipe GBIF France](#)

12:00 - Pause déjeuner

13:30 - Exercice pratique de mise en forme des données : manipulation des tableurs GBIF par [l'équipe GBIF France](#)

15:00 - Pause café/thé

15:30 - Méthodes et outils pour améliorer la qualité des données - Démonstration de OpenRefine par [l'équipe GBIF France](#)

16:00 - Exercice pratique sur OpenRefine par [l'équipe GBIF France](#)

17:00 - Fin de la 1ère journée

**Vendredi 24 février 2017: Publication des données (standards et outils) – Data Papers et exploration du site [www.gbif.org](http://www.gbif.org).**

09:00 - Comment publier des données d'occurrence et enregistrer des jeux de données au GBIF par [l'équipe GBIF France](#)

Les données d'occurrences circulent dans le réseau GBIF depuis le début, mais de nouvelles méthodes ont été mises en place pour rendre ce processus plus simple et accessible pour tous types d'éditeurs. Les nouveaux modèles de tableurs et outils web ainsi que les fonctionnalités spécifiques

de l'IPT v.2.3.2 seront examinés en détail, ainsi que les différentes marches à suivre possibles pour connecter des données au nouveau portail GBIF.

09:30 - Publication des données - Démonstration de l'IPT par [l'équipe GBIF France](#)

10:00 - Exercice pratique sur l'IPT : création d'une ressource et chargement du fichier source par [l'équipe GBIF France](#)

10:30 - Pause café/thé

11:00 - Exercice pratique sur l'IPT : remplissage des métadonnées par [l'équipe GBIF France](#)

11:30 - Exercice pratique sur l'IPT : mapping par [l'équipe GBIF France](#)

12:00 Pause déjeuner

13:30 - Exercices pratique sur l'IPT : publication par [l'équipe GBIF France](#)

14:30 - Présentation des Data papers par [l'équipe GBIF France](#)

14:45 - Pause café/thé

15:15 - Présentation du site GBIF.org par [l'équipe GBIF France](#)

Récapitulation des principales fonctionnalités du nouveau portail GBIF, en soulignant les requêtes et les fonctionnalités de téléchargement.

15:45 - Exercices pratiques sur le GBIF.org (création de compte, recherche et téléchargement de données) par [l'équipe GBIF France](#)

17:00 - Fin de la 2ème journée et clôture de la formation