



---

## PUBLICATION DE DONNÉES DE BIODIVERSITÉ DANS LE RÉSEAU GBIF (GLOBAL BIODIVERSITY INFORMATION FACILITY)

*Documentation pour soutenir le processus d'autorisation interne afin de devenir un éditeur GBIF et GBIF France*

### Historique des versions du document

---

Version	Date	Commentaire
1.0	09/06/2020	Version initiale

### Rédaction

---

GBIF France

### Contexte

---

Document réalisé dans le cadre du projet "OpenPSD: Promoting publication and use of private-sector data on biodiversity" financé par le GBIF et co-financé par les points nodaux GBIF Spain, GBIF Colombia, GBIF France, GBIF Norway et GBIF Portugal.

### Diffusion

---

Editeurs de données français du secteur privé



## Introduction

Les étapes pour qu'une entreprise devienne éditrice de données via le GBIF sont relativement simples. Cependant, le futur éditeur doit être conscient des responsabilités des éditeurs et utilisateurs de données GBIF, ainsi que des différents aspects liés à la préparation, la publication et au suivi des données publiées.

En fonction de la taille de l'entreprise, il se peut que le premier contact avec le GBIF soit établi par le personnel des services Environnement ou Développement Durable. Cependant, la décision de devenir éditeur sera généralement prise par un administrateur ou un directeur, qui devra être informé de ce que signifie devenir éditeur GBIF et des avantages pour l'entreprise.

Le GBIF, en collaboration avec l'IAIA (International Association for Impact Assessment), a publié un guide pour les organisations du secteur privé qui réalisent des évaluations d'impact environnemental (EIE) qui couvre la plupart des avantages et des aspects techniques de la publication des données via GBIF. Ce document est en cours de mise à jour car il lui manquait certains aspects importants tels que les coûts, l'adoption de licences et les relations avec les différentes parties impliquées dans la production de données au sein d'une organisation, qui peuvent inclure des sous-traitants.

L'objectif de ce document est de servir de cadre de référence pour établir un processus de documentation interne à présenter aux organes de décision de l'entreprise pour informer et soutenir leur décision de devenir éditeur du GBIF. À cet égard, plusieurs aspects sont abordés, tels que:

- Qu'est-ce que le GBIF ?
- Quels sont les avantages de publier des données via GBIF?
- Qui publie et quelles données sont publiées via le GBIF ?
- Quelles étapes devez-vous suivre pour devenir éditeur de données ?
- Comment impliquer toutes les parties associées aux jeux de données dans votre publication ?
- Quels sont les coûts associés à la publication ?

Ce cadre de référence propose qu'une preuve de concept soit élaborée en interne. Celle-ci identifiera les étapes nécessaires à la réalisation d'une publication de données, générant une appropriation interne du modèle de publication de données, qui va du traitement des données sur la biodiversité de l'entreprise, en passant par l'évaluation des ressources internes et externes qui doivent être assignées.

## 1. Présentation

Il est proposé dans ce document d'impliquer <entreprise> en tant qu'éditeur de données de biodiversité sur la plateforme intergouvernementale GBIF (Global Biodiversity Information Facility).

Dans le contexte du déclin mondial de la biodiversité, les données d'occurrence et d'abondance d'espèces sont des outils essentiels pour planifier, mettre en place et gérer des stratégies de conservation et de développement durable. Cette importance est reconnue à l'échelle internationale par les Objectifs de Développement Durable des Nations Unies<sup>1</sup> SDG14 et SDG15, dont le GBIF permet de suivre la progression vers les objectifs fixés.

De même, l'objectif 19 d'Aichi, défini par la Convention sur la Diversité Biologique (CDB), utilise la publication de données sur le GBIF comme l'un des indicateurs pour valider l'atteinte de l'objectif d'ici 2020. Aux côtés des universités, instituts de recherche et agences gouvernementales, le secteur privé peut y jouer un rôle clé, grâce aux milliers d'études d'impacts environnementaux effectuées au niveau local dans le monde entier, qui produisent chaque année des millions de données. Cependant, la participation à cet effort est encore très faible, et les données ne sont jamais, ou très rarement, mises à disposition du plus grand nombre.

**Dans ce domaine, <entreprise> pourrait prendre une position de tête au niveau <national/international>, en partageant ses données dans le GBIF de manière organisée et systématique.**

Ainsi, nonobstant la prudence nécessaire aux aspects liés à la propriété intellectuelle et à la confidentialité, **il est proposé de réaliser une preuve de concept, afin de développer et d'évaluer un modèle de publication des données de biodiversité par <entreprise>, qui constituera une étude de cas pour le GBIF et <entreprise>. Cet exercice devrait avoir plusieurs avantages pour <entreprise> : participer à consolider sa réputation <nationale/internationale> en tant qu'entreprise pionnière dans le domaine du développement durable, et renforcer la disponibilité des données, leur intégration et les processus d'organisation qui pourraient être réutilisés dans d'autres contextes.**

## 2. Qu'est-ce que le GBIF ?

Le Système Mondial d'Information sur la Biodiversité est une organisation intergouvernementale créée en 2001 afin de faciliter le partage et l'accès libre et ouvert aux données de biodiversité. A l'heure actuelle, 59 pays ont signé le Memorandum of Understanding du GBIF. Le GBIF fournit un point d'accès unique ([www.gbif.org](http://www.gbif.org)) à plus d'un milliard de données de biodiversité à l'échelle de la planète, et constitue le plus grand réseau sur la biodiversité accessible via Internet.

Les données accessibles via le GBIF concernent des enregistrements de plus d'1,6 million d'espèces, collectées sur plus de trois siècles d'exploration de l'Histoire Naturelle, et incluent des observations récentes des programmes de science participative, de la Recherche et des programmes de suivi automatisé.

Les données téléchargées depuis le GBIF ont été utilisées en 2018 et 2019 dans plus de 1300 articles scientifiques publiés dans des revues internationales. A l'échelle mondiale, le GBIF a des accords et fournit directement des services aux initiatives décisionnaires internationales pour le suivi et la conservation de la biodiversité et de l'environnement, telles que la Convention sur la Diversité Biologique (CDB), la Plateforme intergouvernementale scientifique et politique sur la biodiversité et les services écosystémiques (IPBES) et GEO BON (Group on Earth Observation - Biodiversity Observation Network).

En France, la loi du 08 août 2016 sur la reconquête de la biodiversité a instauré l'obligation pour les entreprises du secteur public ou privé de contribuer à l'inventaire du patrimoine naturel en partageant les données primaires de biodiversité générées dans le cadre d'études d'impacts environnementaux.

Cette obligation légale s'applique depuis le 1er juin 2018. Elle concerne à la fois les études préliminaires et les études de suivi des impacts effectuées dans le contexte de :

- Projets d'aménagement du territoire soumis à l'approbation des autorités administratives, incluant un diagnostic et la collecte de données
- L'élaboration de documents d'aménagement.

Elle s'applique à tous les documents liés à un projet d'aménagement au sens général du terme, produisant ou mobilisant des données de biodiversité.

Source :

<http://www.naturefrance.fr/la-reunion/actualite/depobio-le-depot-legal-des-donnees-brutes-de-biodiversite>

### **3. Quels sont les avantages de la publication des données dans le réseau GBIF ?**

La publication de données dans le GBIF est une initiative qui peut avoir beaucoup de visibilité à l'échelle internationale, et aider à consolider la réputation de <entreprise> comme référence dans le domaine du développement durable.

La publication de données dans le GBIF est très valorisée au sein de la CBD et a même été choisie comme indicateur de la Décennie de la Biodiversité (2010-2020) pour l'un des objectifs d'Aichi<sup>2</sup> (Objectif 19).

Les données publiées dans le GBIF contribuent également à l'objectif d'Aichi n.9 (espèces envahissantes), 11 (aires protégées) et 12 (espèces menacées). De même, ces données sont utiles pour l'évaluation des SDG 14 (biodiversité sub-aquatique) et 15 (biodiversité terrestre).

En contribuant à ces objectifs, <entreprise> améliorera aussi ses profils environnemental et social en augmentant le retour sur investissement initialement effectué pour collecter ces données.

Le GBIF est très intéressé par l'utilisation et le partage de nouvelles études de cas, et la mobilisation des données issues du secteur privé est un des grands objectifs actuels du réseau. Ce processus pourrait favorablement contribuer à la performance du GBIF dans le Dow Jones Sustainability Index et à l'évaluation de l'entreprise dans le contexte des Equator Principles.

Un autre avantage de ce processus serait de poser les fondations pour améliorer le modèle de gestion des informations collectées dans le contexte des études d'impact environnementaux, ainsi que le suivi des impacts et les mesures de compensation. La collecte de ces données demande des études coûteuses, pour lesquelles la valeur ajoutée des informations collectées est souvent faible en raison de l'inefficacité de leur gestion ultérieure.

Pour cette raison, les informations sont souvent perdues ou difficiles d'accès et ne contribuent donc pas au processus d'apprentissage interne ni à la réutilisation d'informations collectées dans de nouveaux projets ou à la gestion environnementale durable des infrastructures en opération. La publication de données dans le réseau GBIF, avec l'appui d'un point nodal GBIF national, fournirait motivation et outils supplémentaires pour la gestion de ce type d'informations, ce qui serait potentiellement applicable pour le groupe <entreprise> aux niveaux national et international.

En résumé, d'après le rapport *Digitally Transforming Environmental Assessment*<sup>3</sup>, la mise en ligne de données de biodiversité par des entreprises privées entraîne :

1. Un moindre coût de recherche d'information
2. Un besoin réduit en études de terrain

3. Des informations plus complètes sur les impacts aux étapes préliminaires, pouvant réduire le recours à une étude d'impacts environnementaux (6 à 18 mois économisés)
4. Des données de suivi améliorées, réduisant le risque de suspension temporaire des opérations pour un manquement à la conformité
5. Une confiance accrue envers les entreprises soucieuses de l'environnement pour de futurs investissements et partenariats
6. Une probabilité accrue d'obtenir des marchés
7. L'utilisation de données et d'outils ouverts pour prendre des décisions environnementales rendra le processus plus transparent et permettra de gagner en confiance et en fiabilité.
8. La confiance accrue des investisseurs et de la communauté suscitera un intérêt grandissant pour les actions avec le secteur privé portées par les organismes publics.
9. De meilleurs outils pour traiter les données facilement accessibles aideront à évaluer l'ampleur de l'impact, ainsi que la valeur, le type et le succès de toute compensation proposée.

#### **4. Qui publie des données dans le réseau GBIF ?**

A l'heure actuelle, le GBIF comprend 47 478 jeux de données sur son portail, publiés par 1536 éditeurs de données enregistrés. Ces éditeurs sont pour la plupart des agences gouvernementales, des muséums d'histoire naturelle, des universités, des organismes de recherche et diverses organisations non-gouvernementales.

L'absence presque totale du secteur privé est visible, ce qui limite l'accès à un grand volume de données de biodiversité collectées par des entreprises du monde entier.

Au total, les enregistrements issus de entreprises privées s'élèvent à 1 069 940, ce qui constitue moins de 0,08 % du nombre total d'enregistrements publiés dans le GBIF.

Le tableau 1 liste les principaux éditeurs de données issus du secteur privé.

**Tableau 1. Entreprises du secteur privé publiant leurs données dans le réseau GBIF (à la date du 23 mars 2020)**

Company	Activity sector	Country	Number of datasets	Records Published
EDP - Energias de Portugal	Energy	Portugal	34	327,547
Isagen	Energy	Colombia	5	18,241
Asplan Viak AS	Engineering	Norway	14	3,779
Oleoducto Bicentenario	Energy	Colombia	3	2,074
Chevron Australia	Energy	Australia	1	2,048
Total	Energy	France	1	324
Multiconsult	Consulting	Norway	1	308
Arctic Research and Consulting DA	Consulting	Norway	1	8,914
Biofokus	Consulting	Norway	1	493,796
Biolog J.B. Jordal AS	Consulting	Norway	1	177,814
Ecofact	Consulting	Norway	1	11,273
Faun Naturforvaltning AS	Consulting	Norway	1	3,788
NNI	Consulting	Norway	2	3,116
Naturrestaurering AS	Consulting	Norway	2	515
Rådgivende Biologer	Consulting	Norway	5	15,214
SWECO Norge AS	Engineering	Norway	1	1,139
INERCO	Consulting	Colombia	1	1,090
Federación Nacional de Cafeteros	Agriculture	Colombia	6	26,840
Aigos SAS	Consulting	Colombia	3	2,404
Anadarko Colombia Company	Energy	Colombia	3	282
Hatovial SAS	Engineering	Colombia	1	1,898
Moam SAS	Consulting	Colombia	1	1,781
Terrasos	Consulting	Colombia	2	7,725
Celsia	Energy	Colombia	1	35
Stratos	Consulting	Colombia	1	849
LafargeHolcim	Mining	Spain	2	35

## 5. Quelles données une entreprise peut-elle publier dans le GBIF ?

Les entreprises qui effectuent des études d'impacts environnementaux, du suivi d'impact et des études de mesures compensatoires, collectant ainsi des données d'occurrence ou d'abondance d'espèces, peuvent les publier dans le GBIF.

Beaucoup de ces données sont collectées dans des régions dans lesquelles les efforts de prospection sont faibles et dont la biodiversité est moins connue, ou possèdent des organismes sous-représentés, et sont donc d'un grand intérêt pour la communauté scientifique et les organisations comme la CBD, IPBES ou GEO BON.

Même les données collectées dans des régions mieux étudiées peuvent être précieuses, car elles permettent de combler les lacunes d'information et d'améliorer les suivis temporels. Ainsi, toutes les données collectées par <entreprise> au cours de ses opérations peuvent être publiées dans le GBIF, en prenant en compte la protection de la propriété intellectuelle et de la confidentialité (permanente ou temporaire) des informations.

Si les données comportent des informations sensibles, telles que la localisation d'espèces menacées, sensibles ou ayant une forte valeur économique, il est recommandé d'appliquer les bonnes pratiques en floutant ces informations<sup>4</sup>.

De cette façon, les données collectées par des entreprises privées peuvent être publiées dans le GBIF au bout d'une courte période, si les aspects administratifs de la publication sont finalisés et si le format des données est adapté à la structure exigée par le GBIF (Darwin Core). Le point nodal national peut également fournir tout le support technique demandé au cours du processus de standardisation et de publication des données.

## 6. Comment une entreprise devient-elle éditeur de données dans le GBIF ?

La décision de devenir éditeur de données de biodiversité dans le réseau GBIF doit être prise par les branches administratives de <entreprise>. Il est ensuite nécessaire de compléter quelques étapes, communes à toutes les institutions souhaitant s'enregistrer comme éditeur de données :

1. Prendre des mesures au niveau institutionnel pour s'assurer que toutes les parties impliquées dans le processus, de l'administration jusqu'aux partenaires liés à la production des données, donnent leur accord pour la publication des données selon les termes indiqués ;
2. Prendre connaissance et accepter le Data Publisher Agreement (la version anglaise est valide dans un cadre juridique) ;
3. Prendre connaissance du Data User Agreement, que les utilisateurs du GBIF doivent accepter avant d'utiliser les données ;



4. Soumettre la candidature de l'institution pour qu'elle soit enregistrée comme éditrice de données GBIF, et demander la validation du point nodal national. Les demandes d'enregistrement et de validation se font en ligne via ce formulaire.

## **7. Implication des différentes parties dans le processus de publication**

Selon l'importance des projets dans le cadre desquels les jeux de données sont produits, il est possible que les données de biodiversité de l'entreprise aient été obtenues par le biais d'autres entreprises ou organisations ayant effectué le travail de collecte. Il s'agit de la configuration la plus commune dans le cas des études d'impacts environnementaux ou des études de suivi, pour lesquelles des sociétés sous-traitantes effectuent les campagnes d'échantillonnage.

L'implication dans le processus de publication des données de ces sous-traitants et techniciens de terrain, qui ont observé ou identifié les espèces, est souhaitable à chaque fois qu'elle est possible. Ces techniciens peuvent jouer un rôle important, notamment dans la vérification des données et métadonnées, et contribuer ainsi à une meilleure qualité du jeu de données. Il est également important qu'ils soient reconnus et cités pour leur travail, et associés aux enregistrements publiés.

Une autre façon de les associer, ainsi que leur organisation, au jeu de données est de les identifier dans les "parties associées" lors du remplissage des métadonnées. Ils doivent être inclus en tant que co-auteurs du jeu de données et dans la citation de ce dernier.

## **8. Comment peut être développée la preuve de concept ?**

La preuve de concept concernant la publication de données dans le GBIF pourrait être développée selon les étapes suivantes :

1. Développement des processus internes de l'entreprise menant à la décision de publier des données dans le GBIF de façon expérimentale.
2. Soumission par l'entreprise de la demande d'enregistrement en tant qu'éditeur de données
3. Mise en place d'une étude de cas sur l'implication de l'entreprise en tant qu'éditeur de données avec le point nodal national, les partenaires nationaux et le GBIF international, afin de donner une visibilité mondiale au processus et inciter d'autres entreprises à s'investir en tant qu'éditeurs de données.
4. Sélection d'un premier jeu de données pour publication sur le GBIF, issu d'études effectuées par l'entreprise. Ce jeu de données doit contenir une bonne représentation de la diversité taxonomique des groupes étudiés et des formats de données, afin d'anticiper les potentiels problèmes liés à la structure et à la disponibilité de l'information.
5. Définition du type d'informations pouvant être publiées et de toute restriction sur leur publication, par exemple la présence d'espèces sensibles, d'informations confidentielles, de données en attente de validation par des institutions

gouvernementales, etc. Les données d'occurrence (observation ou collecte d'une espèce donnée à un lieu et une date données) ou d'abondance peuvent être publiées.

6. Mise en place d'accords avec les producteurs de données (par exemple institutions et personnels embauchés par l'entreprise pour collecter des données dans le contexte d'études de terrain), afin de garantir le droit à la propriété intellectuelle.
7. Formatage des données afin qu'elles soient publiées dans le respect du standard Darwin Core utilisé par le GBIF pour préparer les bases de données pour la publication.
8. Sélection d'une licence Creative Commons à attribuer au jeu de données en cours de publication ; le choix doit s'effectuer parmi les licences CC0, CC-BY, ou CC-BY-NC<sup>5</sup>.
9. Publier données et métadonnées associées à chaque jeu de données enregistré sur le portail GBIF. Les options de mise en ligne de données seront étudiées, et utiliseront dans tous les cas la plateforme logicielle développée par le GBIF : l'Integrated Publishing Toolkit (IPT). Les points nodaux GBIF maintiennent un IPT, qu'ils rendent accessible pour héberger les jeux de données d'éditeurs de leur pays. Il est aussi possible pour une entreprise d'installer et maintenir son propre IPT. Dans les deux cas, l'éditeur de données est toujours l'institution (entreprise), et non le point nodal, et l'institution est responsable de la gestion des données (modifications, mises à jour) de façon autonome.
10. Suivre l'utilisation des données publiées durant une période d'un an après leur mise en ligne dans le réseau GBIF. Cela sera fait grâce aux statistiques fournies par l'éditeur de données concernant l'envoi de données. De plus, l'utilisation des données dans les publications scientifiques sera suivie grâce à l'attribution d'un Digital Object Identifier (DOI) à chaque jeu de données enregistré sur le GBIF ou téléchargé via le GBIF.

## **9. Quels sont les coûts pour <entreprise> ?**

Mis à part le temps consacré par les équipes de <entreprise> à la préparation de la preuve de concept, il n'y a pas de coûts supplémentaires pour l'entreprise.

Le travail nécessaire peut se faire avec l'appui du point nodal GBIF national, qui a les connaissances et l'infrastructure nécessaires pour faciliter la publication des données. Si le point nodal national met son installation IPT à disposition pour héberger et publier les données, il est recommandé d'encadrer ce service avec le Service Level Agreement entre le point nodal GBIF (fournisseur du service) et l'entreprise (utilisateur du service). Ce service n'a pas non plus de coûts associés.

De plus, le point nodal GBIF peut effectuer une formation sur la mise en ligne des données dans le réseau GBIF, contribuant ainsi à renforcer les capacités de <entreprise> dans les domaines de la gestion des informations sur la biodiversité et de la qualité des données.



## Références

<sup>1</sup> <https://sustainabledevelopment.un.org/sdgs>

<sup>2</sup> <http://www.bipindicators.net/numberofgbifrecordsovertime>

<sup>3</sup> Hatton T et al, 2019. Digitally Transforming Environmental Assessment. Report of the Digital Environmental Impact Assessment Working Group. The Western Australian Biodiversity Science Institute and the Western Australian Marine Science Institution, Western Australia, 32 pp.

<sup>4</sup> Chapman AD & Grafton O (2008) Guide to Best Practices for Generalising Sensitive Species-Occurrence Data, version 1.0. Copenhagen: Global Biodiversity Information Facility, 27 pp. ISBN: 87-92020-06-2. Available at: <https://www.gbif.org/document/80512>  
<sup>5</sup><http://www.gbif.org/terms/licences>