

DATA4NATURE

PARTAGER POUR MIEUX PROTÉGER

Guide de recommandations pratiques
pour la publication des données brutes de biodiversité

UNE INITIATIVE
POUR ENCOURAGER
LE PARTAGE
DES DONNÉES
DE BIODIVERSITÉ
GÉNÉRÉES PAR
LES PROJETS DE
DÉVELOPPEMENT
VIA LE RÉSEAU
MONDIAL DU GBIF.



1 Préambule

L'AFD engage désormais les maitres d'ouvrage et les prestataires à publier les éventuelles données brutes de biodiversité générées dans le cadre des projets financés, via le réseau mondial du GBIF (Global Biodiversity Information Facility).

En complément des exigences relatives à la publication des données de biodiversité formulées dans les conventions de financement et termes de référence, ce document fournit des recommandations pratiques à l'attention des bureaux d'études ou autres acteurs en charge de la publication des données.

L'application de la clause des contrats AFD relative à la publication des données de biodiversité fait intervenir au moins trois ou quatre acteurs différents :

- L'AFD ;
- Son client, généralement un maitre d'ouvrage ;
- Un ou plusieurs bureaux d'études spécialisés dans la collecte et l'analyse de données sur la biodiversité ;
- Le GBIF.

Comme le met en évidence la figure suivante, chacun de ces acteurs joue un rôle précis dans le processus de publication des données. La mise en œuvre pratique de la publication des données est généralement déléguée au(x) bureau(x) d'études (ou autres acteurs) en charge de la collecte des données. Ces bureaux d'études doivent, ainsi, réaliser plusieurs étapes de travail successives avant de procéder à la publication des données en tant que telle.

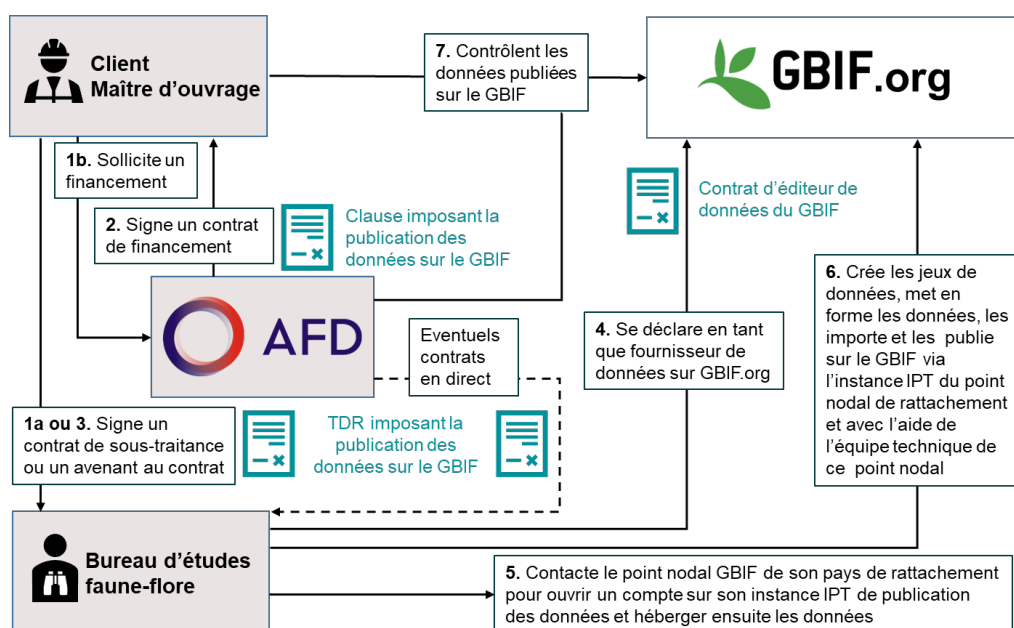


Figure 1 : Schéma du processus de publication des données brutes de biodiversité générées grâce aux financements de l'AFD

2 Comment devenir fournisseur de données du GBIF ?

2.1 Signature du contrat d'éditeur de données du GBIF

Si vous n'êtes pas encore fournisseur de données du GBIF, la première étape pour publier des données dans le réseau GBIF consiste à s'enregistrer en tant que fournisseur ou éditeur de données. Pour cela, il faut se rendre sur le portail web du GBIF <https://www.gbif.org/become-a-publisher> afin d'accepter les termes et conditions du contrat d'éditeur de données puis de remplir un court formulaire décrivant votre organisme, le type de données que vous fournirez et les modalités de mise à disposition des données.

Tableau 1 : Recommandations pour le remplissage du formulaire de demande d'enregistrement en tant que fournisseur de données du GBIF

Rubriques	Recommandations
Information sur l'organisation	Pour les coordonnées de votre organisme, fournissez celles du siège, le cas échéant. De même, pointez sur la carte la localisation du siège de votre organisme.
Nœud d'approbation / Point nodal	Conservez la valeur activée par défaut sauf si vous envisagez de fournir exclusivement des données marines.
Projets GBIF	Conservez la valeur « Non » proposée par défaut.
Contacts	Indiquez <i>a minima</i> les coordonnées de l'interlocuteur du GBIF pour votre organisme, généralement le responsable de la gestion des données de biodiversité.
Quoi et comment	Cochez <i>a minima</i> « Données d'occurrence uniquement ». Vous pouvez néanmoins sélectionner d'autres types de données si vous le souhaitez. Fournissez ensuite une courte description de tous les types de données que vous êtes susceptible de publier. Si vous ne disposez pas d'un serveur informatique à travers lequel vous donnerez directement accès à vos données au GBIF, nous vous recommandons de demander de l'aide pour publier vos données. Le personnel du point nodal GBIF dont vous dépendez vous contactera pour définir les modalités de publication de vos données.

Dès lors que vous enregistrez le formulaire, celui-ci est soumis au point nodal GBIF dont vous dépendez. Celui-ci évaluera votre demande avant de l'accepter.

Dès que votre demande d'enregistrement en tant que fournisseur de données du GBIF est acceptée, une page web dédiée vous sera automatiquement créée sur le site GBIF.org avec les informations que vous aurez fournies dans le formulaire. Depuis cette page vous aurez également accès aux jeux de données que vous aurez publiés et à des statistiques sur ces données.

Seules des organisations peuvent s'enregistrer en tant que fournisseur de données du GBIF. L'enregistrement ne peut se faire à titre individuel.

Si votre organisme est implanté dans un pays non couvert par un point nodal GBIF, l'approbation de votre demande se fait par le Comité de direction des points nodaux.

Davantage d'informations sur l'enregistrement d'un organisme en tant que fournisseur de données du GBIF sont disponibles sur la page web suivante : <https://www.gbif.org/fr/endorsement-guidelines>

2.2 Ouverture d'une instance IPT de publication des données

La publication de vos données sur le GBIF requiert :

- Un serveur informatique que vous utiliserez à la fois pour entreposer vos données brutes de biodiversité et y donner accès au GBIF ;
- Un logiciel de publication de données tel que le logiciel libre et gratuit « IPT » (Integrated Publishing Toolkit) mis à disposition par le GBIF.

Si vous n'avez pas de serveur informatique ou si vous n'avez pas la possibilité de partager vos données depuis vos propres serveurs, le point nodal GBIF dont vous dépendez ou à défaut les équipes du Secrétariat international du GBIF peuvent éventuellement héberger vos données et ouvrir une instance IPT de publication de vos données sur leurs propres serveurs. Dans ce

*Vous pouvez retrouver les coordonnées du point nodal GBIF dont vous dépendez à l'adresse suivante <https://www.gbif.org/the-gbif-network>
Si le pays d'implantation de votre organisme n'a pas de point nodal GBIF, rapprochez-vous du secrétariat du GBIF à Copenhague pour connaître le point nodal dont vous dépendez.*

cas, votre point nodal GBIF vous communiquera l'URL de l'instance IPT à utiliser ainsi que vos identifiants de connexion.

Dans tous les cas, nous vous recommandons de vous rapprocher du point nodal GBIF dont vous dépendez pour mettre en place et configurer vos outils de publication de données.

3 Recommandations pour la collecte et la mise en forme des données

3.1 Collecte des données

La publication des données brutes de biodiversité sur le GBIF ne nécessite pas la collecte d'informations supplémentaires sur le terrain. Elle implique uniquement la mise en forme des données selon un cadre défini.

Il est, par conséquent, recommandé d'anticiper cette mise en forme dès la collecte des informations sur le terrain afin de vous faciliter par la suite la publication des données. Cette anticipation de la publication des données peut se traduire par un certain nombre de bonnes pratiques qui concourront également à améliorer la qualité de vos données :

- Utilisation de fiches de saisie des observations sur le terrain permettant *a minima* de préciser : le type d'observation, le nom scientifique du taxon, sa localisation et la date d'observation (Un exemple de fiche de terrain est fourni en annexe) ;
- Localisation des relevés et des observations les plus intéressantes à l'aide d'un GPS ;
- Centralisation et numérisation des observations dans un tableau dont la structure est la plus proche possible du standard défini pour la publication des occurrences d'espèces sur le GBIF (Voir le chapitre 3.2 suivant dans le présent document).

La localisation des observations à l'aide d'un GPS n'est pas indispensable. Chaque observation doit néanmoins pouvoir être associée à une position géographique et être publiée avec la même précision de localisation que celle collectée sur le terrain, hormis pour les espèces considérées comme sensibles (Voir le chapitre 3.3 à la page 6 du présent document).

Pour les observations non précisément géolocalisées, la position géographique peut correspondre aux coordonnées d'un point situé au centre de l'aire d'étude. Dans ce cas, le rayon du plus petit cercle englobant l'aire d'étude à partir de ce point central sera évalué et intégré dans les données (champ *coordinateUncertainty* à compléter, voir le Tableau 2 à la page 5).

3.2 Mise en forme des données selon les standards du GBIF

Les termes de référence de diligences en phase d'instruction et conventions de financement de projet/programme établis par l'AFD impliquent que les données brutes de biodiversité récoltées soient publiées *a minima* sous la forme d'**occurrences d'espèces** afin de fournir suffisamment d'informations sur la localisation des observations dans le temps et l'espace.

Ces occurrences d'espèces devront être regroupées et publiées par jeux de données pour lesquels seront fournies des métadonnées, c'est-à-dire des informations complémentaires détaillant les conditions d'acquisition et d'utilisation des données. Le chapitre 4.4 à la page 8 détaille les métadonnées à renseigner.

En l'absence d'exigences spécifiques complémentaires de l'AFD pour la publication des données, **chaque projet financé donnera lieu, au minimum, à la publication d'un jeu de données regroupant l'ensemble des observations réalisées sur le terrain.**

Afin de préparer la publication de votre jeu de données, il est nécessaire de lister vos observations dans un tableur renseigné avec les informations détaillées dans le tableau ci-après. Dans ce cadre, vous pouvez exploiter le modèle de fichier Excel mis à disposition par le GBIF ici : <https://github.com/gbif/ipt/wiki/occurrenceData#templates>. Il se peut que vous ayez à ajouter manuellement certains champs dans le modèle de fichier communiqué. Dans ce cas, il est important d'utiliser exactement les noms de champ listés dans le tableau qui suit.

Le renseignement de la localisation des observations requiert des connaissances de base en cartographie. Il est, par ailleurs, vivement recommandé d'utiliser un système d'information géographique et/ou un logiciel de gestion des positions GPS pour renseigner les coordonnées géographiques des observations.

Si vous le souhaitez, vous êtes libre de publier vos données dans plusieurs jeux de données différents afin de fournir davantage de précision sur les protocoles d'inventaire utilisés pour chaque jeu de données.

Tableau 2 : Description des champs du standard "occurrence d'espèce" utilisés par le GBIF et attendus dans le cadre de la publication de données de biodiversité acquises via des financements AFD

	Nom du champ	Description
Champs requis systématiquement	occurrenceID	Identifiant unique de l'occurrence de taxon. Vous êtes libres de définir vous-même cet identifiant unique. <i>A minima</i> , il doit être unique au sein du jeu de données. Il peut s'agir d'un identifiant numérique ou alphanumérique (exemple : 2354 ou Obs0015). La bonne pratique est cependant d'utiliser un identifiant unique et permanent de type UUID (Universally unique identifier) ou éventuellement URI (Uniform Resource Identifier) généré par ordinateur.
	basisOfRecord	Nature de l'observation à l'origine de l'occurrence, il peut s'agir par exemple d'une observation humaine ou d'un enregistrement réalisé par une machine (exemple : enregistreur à ultrasons, piège photographique, balise Argos, radar...).
	scientificName	Nom scientifique complet au rang taxonomique le plus bas possible incluant l'auteur de la description du taxon et la date de publication du nom (Exemple : <i>Canis lupus Linnaeus, 1758</i>). Le nom scientifique des taxons peut être recherché sur le site du GBIF via la page suivante https://www.gbif.org/species/search ou alors grâce à l'outil en ligne Species matching tool fourni par le GBIF (voir la partie 3.4 à la page 6 du présent document pour plus d'informations sur cet outil).
	eventDate	Date ou période au cours de laquelle le taxon a été observé. La date doit être formatée selon le standard ISO 8601 (Exemple : pour le 18 novembre 2020, on écrira 2020-11-18)
	decimalLatitude	Coordonnées géographiques de l'occurrence en degrés décimaux. Ex : decimalLatitude = 48,861037 et decimalLongitude = 2,335860.
	decimalLongitude	
	geodeticDatum	Système de coordonnées géographiques utilisé pour indiquer la latitude et la longitude des occurrences. A titre indicatif, le système de coordonnées géographiques de référence utilisé par le GBIF est le WGS84 (World Geodetic System 1984).
Champs requis sous condition	coordinateUncertainty	Précision de la localisation de l'occurrence de taxon en mètres. Autrement dit, il s'agit du rayon du plus petit cercle contenant l'observation. Condition : les coordonnées fournies correspondent au centre d'une zone au sein de laquelle a été observée l'occurrence.
	dataGeneralizations	Description d'une éventuelle dégradation volontaire des données (par exemple un floutage de la position géographique). Permet à l'utilisateur de savoir qu'il existe une version plus précise des données publiées. Condition : observation d'une espèce considérée comme sensible
Champs recommandés	individualCount	Le champ <i>individualCount</i> permet de préciser le nombre d'individus observés pour le taxon en question. Il est également possible de quantifier l'occurrence de taxon avec d'autres systèmes de mesures par exemple des coefficients d'abondance dominance, une densité, un coefficient d'abondance-dominance, etc. Dans ce cas les champs, <i>organismQuantity</i> et <i>organismQuantityType</i> sont à utiliser pour indiquer la quantité et l'unité de mesure utilisée pour quantifier l'occurrence de taxon.
	organismQuantity	
	organismQuantityType	
	countryCode	Code à deux lettres indiquant le pays dans lequel le taxon a été observé. Ce code doit suivre le standard ISO 3166-1-alpha-2
	taxonRank	Rang taxonomique du nom scientifique fourni pour l'occurrence de taxon en latin ou en anglais. (Exemples : kingdom, phylum, class, order, family, genus, species)
	kingdom	Nom scientifique complet, en latin, du règne auquel appartient le taxon observé. Les valeurs possibles pour ce champ sont donc : Animalia, Archaea, Bacteria, Chromista, Fungi, Plantae, Protozoa ou Viruses.
	informationWithheld	Dans le cas d'une dégradation volontaire des données, il est utile de compléter le champ <i>informationWithheld</i> en complément du champ <i>dataGeneralizations</i> afin d'indiquer que des informations plus précises existent mais qu'elles n'ont pas été communiquées volontairement.

Davantage d'informations sur les champs associés à la description des occurrences d'espèces sont accessibles sur le portail web du GBIF : <https://www.gbif.org/fr/data-quality-requirements-occurrences#emlBasis>

Les données de chaque jeu de données peuvent être saisies dans la langue de votre choix. La langue utilisée pour décrire les données et métadonnées doit néanmoins être précisée dans les métadonnées de chaque jeu de données. Cette information est déclarée à l'étape 4 du processus de publication des données (voir la partie 4.4 à la page 8 de ce document).

Le GBIF vous permet également de regrouper vos occurrences d'espèces dans des événements d'échantillonnage. Dans ce cadre, vos listes d'observations peuvent être associées à différents relevés pour lesquels vous pouvez publier des informations complémentaires spécifiques. Bien

que ces regroupements ne soient pas exigés par l'AFD, ils augmentent encore la valeur et l'exploitabilité de vos jeux de données. Davantage d'informations sur les événements d'échantillonnage sont accessibles sur le site du GBIF ici : <https://www.gbif.org/fr/sampling-event-data>.

3.3 Dégradation de la localisation géographique des données sensibles

Afin de ne pas compromettre la survie des populations locales d'espèces menacées par le prélèvement ou la destruction intentionnelle d'individus (braconnage par exemple), vous dégraderiez volontairement la précision de la localisation des observations d'espèces dites sensibles.

La méthode la plus simple pour dégrader la localisation d'observations consiste à arrondir les coordonnées géographiques enregistrées. Cet arrondi peut être adapté en fonction du niveau de sensibilité de l'espèce. Ainsi, pour les espèces les plus sensibles vous pourriez par exemple arrondir les coordonnées géographiques aux nombres entiers les plus proches. Pour les espèces un peu moins sensibles vous pouvez éventuellement arrondir les coordonnées à un ou deux chiffres après la virgule.

Tableau 3 Correspondance entre le nombre de décimales utilisées pour les coordonnées géographiques et le niveau de dégradation de la localisation des données à l'équateur

Nombre de chiffres après la virgule pour l'arrondi des coordonnées géographiques	Valeur en degré décimaux	Equivalence en mètres sur le terrain / marge d'erreur dans la localisation des observations
0	1,0	111,32 km
1	0,1	11,13 km
2	0,01	1,11 km

Dès lors que vous dégradez volontairement la précision de localisation d'une observation, il est nécessaire de décrire la nature de la dégradation dans le champ `dataGeneralizations`. On pourra par exemple écrire dans ce champ : « Arrondi des coordonnées géographiques aux nombres entiers les plus proches ». En complément du champ `dataGeneralizations`, il est également utile de remplir le champ `informationWithheld` afin d'indiquer que des informations plus précises sur la localisation existe mais qu'elles n'ont pas été communiquées intentionnellement.

Au-delà des cercles polaires, le floutage de la localisation des espèces sensibles par dégradation des coordonnées géographiques devient moins pertinent du fait du rapprochement des méridiens.

Pour plus d'informations sur la dégradation de la localisation des espèces sensibles, vous pouvez vous référer au guide de bonnes pratiques mis à disposition par le GBIF (CHAPMAN, 2020) et téléchargeable ici : <https://www.gbif.org/document/80512/guide-to-best-practices-for-generalising-sensitive-species-occurrence-data>

3.4 Vérification de la validité des jeux de données

Pour que la publication de vos jeux de données se déroule correctement, il faut que les fichiers de listes d'observations que vous importerez dans le logiciel IPT de publication des données soient conformes aux standards utilisés par le GBIF.

Afin de vérifier que vos fichiers Excel sont bien formatés, nous vous recommandons d'utiliser successivement les deux outils suivants mis à disposition par le GBIF. Ces outils vous permettront de mettre en évidence les éventuelles erreurs que vous avez pu faire dans vos fichiers afin que vous les corrigiez plus facilement avant de les publier.

Tableau 4 : Principaux outils permettant de vérifier la validité de vos jeux de données

Etape	Nom de l'outil et lien vers l'outil	Fonction	Remarque
1	GBIF Species matching tool	Vérification de la validité des noms scientifiques des taxons et recherche des noms de taxon acceptés par le GBIF	Votre fichier doit être converti préalablement au format CSV
2	GBIF Data validator	Vérification globale de la compatibilité de vos jeux de données avec les standards de données utilisés par le GBIF	Pré-version de l'application

4 Publication des données sur le GBIF

La publication de jeux de données sur le portail web du GBIF se fait via l'application web IPT (Integrated Publishing Toolkit). Dans la mesure où vous utilisez une instance IPT mise à disposition par le point nodal GBIF dont vous dépendez, vous accédez à l'application grâce à l'URL, au compte utilisateur et au mot de passe communiqués par votre point nodal GBIF.

Afin d'utiliser facilement et correctement l'application IPT, nous vous recommandons de consulter la page <https://www.gbif.org/fr/ipt> et de visionner la démonstration vidéo mise à disposition par le GBIF. Vous pouvez également solliciter une présentation de l'application ou une formation auprès de votre point nodal GBIF.

Lorsque vous utilisez l'application IPT, à tout moment vous avez accès à un manuel utilisateur en cliquant sur le lien dédié en bas de chaque écran.

4.1 Etape 1 : Création d'un jeu de données

Dès lors que vous avez vérifié et confirmé la conformité de vos données avec les standards du GBIF vous êtes prêts à les publier depuis l'application IPT. Il vous suffit alors de créer une nouvelle ressource de type « Occurrence » ou éventuellement de type « Sampling event » depuis l'onglet « Manage Resources » de l'IPT.

The screenshot shows the GBIF Integrated Publishing Toolkit (IPT) interface. At the top, there's a header with the GBIF logo, the text 'GBIF INTEGRATED PUBLISHING TOOLKIT', and a login status 'Logged in as celleboode@biotope.fr'. Below the header is a navigation bar with 'Home', 'Manage Resources', and 'About' buttons. The main content area is titled 'Overview: test_occurrences' and contains several sections: 'Source Data' with a 'Parcourir...' button and a 'Connect to database' button; 'Darwin Core Mappings' with a dropdown menu; 'Metadata' with an 'Edit' button; 'Published Versions' with a 'Publish' button and a table showing version details; 'Visibility' with a 'Public' button; and 'Resource Managers' with an 'Add' button. Five blue callout boxes with arrows point to specific features: 'Etape 2 : Import des fichiers de données' points to the 'Source Data' section; 'Etape 3 : Mise en correspondance avec le Darwin Core' points to the 'Darwin Core Mappings' section; 'Etape 4 : Edition des métadonnées' points to the 'Metadata' section; 'Etape 5 : Publication des données' points to the 'Published Versions' section; and another 'Etape 5 : Publication des données' points to the 'Visibility' section.

Figure 2 Interface graphique de la page de gestion d'un jeu de données dans l'application IPT

4.2 Etape 2 : Import des données dans le jeu de données

Lorsque vous avez créé un jeu de données sur l'IPT, vous pouvez importer le tableau Excel contenant vos données depuis la rubrique « Source Data ». Lorsque vous importez vos données, une fenêtre intermédiaire s'ouvre afin que vous puissiez éditer si besoin le format de vos données sources. A cette étape, nous vous recommandons de prévisualiser les données que vous avez importées afin de vérifier que les formats utilisés sont correctement intégrés.

Pour prévisualiser les données importées, cliquez sur l'œil à droite du bouton « Analyze » dans la fenêtre d'édition de vos données.

4.3 Etape 3 : Mise en correspondance des données importées avec le standard Darwin Core

Le Darwin Core est le format du standard de fichier de données utilisé par le GBIF. Ainsi, cette étape vise à s'assurer que le ou les fichiers de données importées sont bien compatibles avec le Darwin Core. Si certains champs de données importés sont manquants ou ne sont pas reconnus automatiquement, il est possible d'ajouter des champs avec des valeurs prédéfinies pour le jeu de données ou alors de mettre en correspondance manuellement certains champs avec le standard Darwin Core.

Afin de simplifier cette étape de travail, nous vous recommandons d'exploiter au maximum le modèle de fichier Excel mis à disposition par le GBIF pour formater la liste de vos occurrences d'espèces et de respecter le format des champs de données listés dans la partie 3.2 à la page 4 du présent document.

Pour en savoir plus sur le standard de données Darwin Core utilisé par le GBIF vous pouvez consulter le site web suivant : <https://dwc.tdwg.org/>

4.4 Etape 4 : Edition des métadonnées

Les métadonnées vous permettent de décrire en particulier les conditions de production et d'utilisation du jeu de données. L'importance de ces informations ne doit pas être négligée car l'exploitabilité du jeu de données en dépend beaucoup. Le renseignement de certaines métadonnées telles que le nom du jeu de données, sa description ou encore les coordonnées du producteur des données est imposé par le GBIF et il s'agit d'une condition préalable à la publication des données.

D'autre part, les conventions de financement et termes de référence de l'AFD imposent le renseignement de certaines métadonnées selon des règles prédéfinies.

C'est notamment le cas des conditions d'utilisation des données, ainsi dans la rubrique « Basic Metadata » vous opterez obligatoirement pour l'un des deux niveaux de droits les moins restrictifs pour le champ « Data Licence », à savoir : la licence « Public Domain (CC0) » ou la licence « Creative Commons Attribution (CC-BY) ».

C'est également le cas des informations sur le Projet dans le cadre duquel vous avez collecté les données. Dans la rubrique « Project Data », il est indispensable de fournir le nom du projet mais aussi une courte description de celui-ci, suivi du nom du maître d'ouvrage. Dans le champ « Funding » vous préciserez obligatoirement le nom des financeurs du projet dont l'Agence Française de Développement.

4.5 Etape 5 : Publication des données

Par défaut, les données que vous publiez sur le GBIF sont privées et accessibles uniquement aux personnes à qui vous avez attribué le rôle de « Manager » de vos jeux de données. Par conséquent, avant de publier vos jeux de données, nous vous recommandons de les rendre accessibles au public en cliquant sur le bouton « Public » dans la rubrique « Visibility » de l'onglet « Manage Resources » de l'IPT.

Dès que votre jeu de données a le statut « Public », vous passez à la dernière étape du processus de publication en cliquant sur « Publish » dans la rubrique « Published version ».

L'application IPT vous permet de modifier ou compléter si besoin les jeux de données que vous avez déjà publiés. Dans ce cas, il est utile de préciser les modifications opérées au moment de la publication d'une nouvelle version d'un jeu de données.

5 Pour aller plus loin

Le GBIF peut vous accompagner tout au long de la démarche décrite dans le présent guide de recommandations. En complément des nombreuses informations fournies sur le portail web du GBIF (<https://www.gbif.org/>), vous avez la possibilité d'adresser vos questions directement à private-sector@gbif.org.

Le GBIF organise régulièrement des formations à l'attention des fournisseurs de données et met à votre disposition une plateforme d'e-learning et des supports de formation en libre téléchargement. <https://www.gbif.org/article/2IE7tH4dlcik1BnmnilPac/t raining-and-e-learning>

*Guide de recommandations
pratiques pour la publication des
données brutes de biodiversité*



CONTACTS



Secrétariat du GBIF à Copenhague (Danemark)

E-mail : helpdesk@gbif.org

Téléphone : +45 35 32 14 70



Agence Française de Développement

E-mail : data4nature@afd.fr