



# Pacific Invasive Species

Série « Invasive Species Battler » du Pacifique

## ÉCHANGE DE DONNÉES SUR LES ESPÈCES ENVAHISSANTES DU PACIFIQUE DANS LE CADRE DU SYSTÈME MONDIAL D'INFORMATION SUR LA BIODIVERSITÉ <GBIF>





### **Catalogage des données de publication pour la bibliothèque du PROE**

Échange de données sur les espèces envahissantes du Pacifique dans le cadre du Système mondial d'information sur la biodiversité (GBIF). Apia, Samoa : PROE, 2018.

28 p. 29 cm.

ISBN: 978-982-04-0785-5 (impression)  
978-982-04-0786-2 (version électronique)

1. Invasive species – Databases – Oceania.
2. Non-indigenous pests – Control – Data processing – Oceania.
3. Introduced organisms – Control – Data processing – Pacific Ocean.
4. Biological invasions – Information management – Oceania.
- I. Programme régional océanien de l'environnement (PROE)
- II. Global Biodiversity Information Facility (GBIF).
- III. Biodiversity Information for Development (BID).
- IV. European Union (EU).
- V. Titre.
- VI. Séries.

632.90995



As part of SPREP's commitment to the environment, this item is printed on recycled paper.

Secrétariat du Programme régional océanien de l'environnement (PROE)

PO Box 240  
Apia, Samoa  
sprep@sprep.org  
www.sprep.org

*Notre vision : « L'environnement océanien, garant de nos modes de subsistance et de notre patrimoine naturel, en harmonie avec nos cultures. »*

Copyright © Secrétariat du Programme régional océanien de l'environnement (PROE), 2018 La reproduction de cette publication à des fins éducatives et non commerciales est autorisée exclusivement avec l'autorisation préalable et écrite du détenteur des droits d'auteurs, dans la mesure où la source est dûment citée. La reproduction de cette publication destinée à la revente, ou à toute autre fin commerciale, est interdite sans l'autorisation préalable et écrite du détenteur des droits d'auteurs.

Photo de couverture : Stuart Chape

# TABLE DES MATIÈRES

Cher Invasive Species Battler	2
À propos de ce guide	3
Pourquoi partager les données sur les espèces ?	3
Que représente le GBIF ou, Système mondial d'information sur la biodiversité ?	4
Comment faire pour trouver des données dans le GBIF ?	5
À qui appartiennent les données ?	8
Que peuvent faire les utilisateurs agréés avec les données que je partage ?	10
Comment publier des données dans le GBIF ?	11
Voies d'accès à la publication de données	12
Comment s'inscrire en tant qu'éditeur de données dans le GBIF ?	13
Notions de base sur le géoréférencement	14
Comment préparer les données à la publication ?	15
Préparation des données à l'aide de la Boîte à outils d'édition intégrée	16
Mappage avec les termes Darwin Core	17
Définition d'un identificateur unique	19
Nettoyage des données	19
Re-publication de données	20
Comment puis-je organiser la publication des données ?	22
Glossaires de définitions	23
Pour de plus amples informations	24
Support continu	24
Sites Internet	24

## Cher Invasive Species Battler,

Nous formons un groupe diversifié dans la région du Pacifique, qui s'étend sur environ un tiers de la surface de la Terre et englobe près de la moitié des mers de la planète. Nous parlons quelque 2 000 langues différentes et notre région compte environ 30 000 îles. Le Pacifique est si varié que ses écosystèmes constituent l'un des points névralgiques de la biodiversité mondiale, avec un grand nombre d'espèces rencontrées uniquement dans le Pacifique et nulle part ailleurs. En fait, 2 189 espèces endémiques à un unique pays ont été recensées à ce jour. De ces espèces, 5,8 % sont déjà éteintes ou n'existent qu'en captivité. En outre, 45 % sont menacées d'extinction. Nous sommes confrontés à l'un des taux d'extinction les plus élevés au monde.

La principale cause d'extinction des espèces endémiques à un unique pays dans le Pacifique tient à la propagation d'espèces envahissantes. Les espèces envahissantes ont également de graves répercussions sur nos économies, nos échanges commerciaux, notre développement durable, notre santé, les services écosystémiques et la capacité de nos écosystèmes à réagir aux catastrophes naturelles. Par chance, nous pouvons y remédier.

Même dans notre région diversifiée, nous partageons de nombreux points communs. Nous sommes des insulaires, nous sommes autonomes et nous dépendons largement de notre environnement pour subvenir à nos besoins. Nous nous heurtons également à de nombreux problèmes communs liés aux espèces envahissantes, car nous sommes en fin de compte tous liés les uns aux autres. Partager nos enseignements au niveau régional nous permet, à nous et à nos familles, d'en tirer des avantages économiques, culturels mais également d'en profiter dans notre vie quotidienne.

La série « Invasive Species Battler » a été conçue pour échanger nos connaissances sur les problèmes communs liés aux espèces envahissantes dans la région. L'objectif n'est pas de traiter chaque question en profondeur, mais de fournir des informations et des études de cas qui vous aideront à prendre une décision concernant d'éventuelles mesures ou à trouver des informations plus précises.

Le Programme sur les espèces envahissantes du PROE vise à fournir un soutien technique, institutionnel et financier aux programmes régionaux sur les espèces envahissantes, en coordination avec d'autres organisations régionales. Nous coordonnons le Réseau pédagogique sur les espèces envahissantes (PILN), un réseau pour les praticiens luttant contre les espèces envahissantes dans les états et territoires du Pacifique, et le Partenariat océanien sur les espèces envahissantes (PIP), l'organisme régional de coordination des organisations travaillant sur ces espèces dans plus d'un pays du Pacifique.

Pour de plus amples renseignements sur le PROE, le Programme sur les espèces envahissantes, le PILN et le PIP, ainsi que pour avoir accès aux ressources et aux outils de sensibilisation, veuillez consulter le site Internet du PROE : [www.sprep.org](http://www.sprep.org)

Merci pour vos efforts,

L'Équipe espèces envahissantes du PROE



### À propos de ce guide

Les données sur les espèces insulaires tropicales peuvent être difficiles à obtenir et à conserver en toute sécurité, mais elles sont précieuses pour les communautés du Pacifique, les gestionnaires environnementaux et les chercheurs du monde entier. Une solution du Pacifique pour gérer et partager efficacement les données sur les espèces envahissantes peut engendrer des résultats notables en matière de biodiversité. Le but de ce guide est d'aider le praticien à publier des données sur les espèces envahissantes et à former d'autres personnes à la préparation et à la publication de rapports pour le Système mondial d'information sur la biodiversité (GBIF) en utilisant une terminologie cohérente permettant d'utiliser les données de manière répétée. Ce guide a été produit dans le cadre du projet « Mobilisation des données et de l'information sur les espèces exotiques et envahissantes nationales et régionales et renforcement des capacités dans le Pacifique », financé par l'Union européenne dans le cadre du programme Information sur la biodiversité pour le développement (DIB) du GBIF et géré par le PROE au nom des états et territoires des îles du Pacifique (ÉTIO).

## Pourquoi partager les données sur les espèces ?

Les données n'ont de la valeur que lorsqu'elles sont utilisées. La valeur des données augmente plus la quantité de données et les liens avec d'autres ensembles de données augmentent. L'échange d'information permet d'accroître la valeur de l'information.

La gestion de l'information sur la biosécurité requise par le personnel des services de quarantaine et de biosécurité peut représenter un défi, particulièrement lorsque la surveillance et le contrôle des espèces nuisibles sont effectués par différents groupes avec des systèmes d'information différents. Les agents de biosécurité bénéficient souvent des données provenant d'autres pays pour une détection et une intervention rapides au sein des frontières nationales. L'établissement de routines pour le partage des données pourrait également renforcer la fiabilité de la collecte et de l'analyse des données et contribuer à une plus grande confiance entre les secteurs et ministères concernés.

Une prise de décision éclairée dépend de la qualité de l'information. Les routines de partage des données, les systèmes et les relations de confiance favorisent la disponibilité de données de haute qualité et en facilitent l'utilisation. Des données de qualité avec un usage simplifié facilitent et accélèrent le processus d'analyse et d'interprétation pour les agents, permettant ainsi de transformer lesdites données en connaissances utiles à la réalisation des priorités nationales.

Nombreuses sont les raisons de partager les données :

- Nous partageons des données pour appuyer la recherche scientifique, les programmes d'éducation et la prise de décisions dans le monde entier.
- Le partage des données nous permet de nous faire connaître auprès de la communauté de chercheurs et d'enseignants consacrés à la biodiversité. Partager, c'est affirmer que nous sommes un membre actif de la communauté, que nous voulons aider et que nous voulons nous engager dans le domaine de la biodiversité. En partageant, vous obtenez une reconnaissance pour votre collecte de données et vos efforts de partage par le biais de références à vos données dans la recherche, ce qui démontre la valeur ajoutée que vous apportez aux connaissances mondiales sur notre biodiversité.
- Le stockage des données dans un référentiel partagé les protège contre les pannes de serveur, les pertes de matériels informatiques et la corruption de fichiers. Les métadonnées et l'information explicative requises pour le partage des données contribuent également à maintenir l'utilité des données, en cas de renouvellement du personnel.
- Le partage des données rend nos informations aussi utiles que possible - pour un plus grand impact ! Les ensembles de données que nous publions et partageons peuvent entre autres être utilisés à des fins de recherche, de formation, pour des programmes d'enseignement, la protection de la biodiversité, de même qu'ils invitent le monde à se rassembler en un seul endroit pour collaborer et communiquer grâce à un intérêt commun pour la collecte de diverses données en matière de biodiversité.
- Le partage des données invite les personnes intéressées à se réunir et à collaborer avec d'autres scientifiques, chercheurs, étudiants, etc. Ceci envoie un message fort qui traduit notre volonté d'être présents, de faire reconnaître notre travail, d'offrir notre aide et d'être ouverts aux compétences disponibles chez les autres fournisseurs de données.
- Nous pouvons encore améliorer les données que nous publions en ouvrant la communication à un réseau mondial de personnes intéressées qui ont un intérêt commun pour nos ensembles de données. L'aide d'autres intervenants nous permettrait de contourner toute limite inhérente à notre ensemble de données publiées, ce qui améliorerait la qualité de l'information et/ou viendrait alors la compléter.
- Pour un petit État insulaire, la publication de nos collectes de données sur notre environnement met en valeur le développement dans les domaines des sciences et de la recherche sur la scène mondiale.

- Dans certains cas, les donateurs ayant financé notre organisation pour faciliter la collecte de données sur les espèces envahissantes peuvent exiger que ces données soient rendues publiques. L'un des moyens de rendre public les données est le GBIF (Système mondial d'information sur la biodiversité).



## Données et ressources régionales sur les espèces envahissantes

Différentes méthodes permettent d'obtenir et de partager des données et informations pertinentes pour la gestion des espèces envahissantes du Pacifique et la prise de décisions. Outre la collecte de données GBIF, des données et des informations sur le Pacifique sont disponibles dans la [base de ressources Battler](#) et sur le [Portail des îles du Pacifique pour les aires protégées](#).

## Que représente le GBIF ou Système mondial d'information sur la biodiversité ?

Le GBIF — Système mondial d'information sur la biodiversité — est un outil de recherche ouvert financé par divers gouvernements du monde entier, visant à fournir, en accès libre, des données sur la biodiversité.

Le site Internet du GBIF à l'adresse [www.gbif.org](http://www.gbif.org) contient plus d'un milliard de documents sur la biodiversité librement et ouvertement accessibles aux utilisateurs. Le site donne accès à ces données par divers outils de recherche.

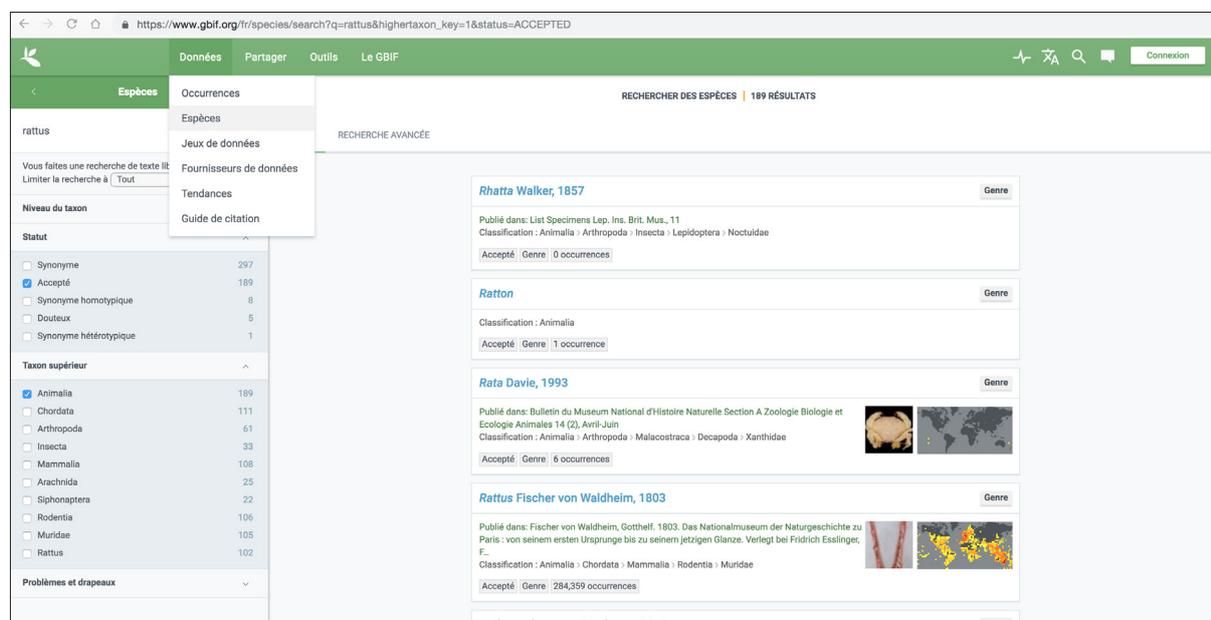
Le système de partage des données du GBIF permet de garantir une certaine uniformité entre les pays grâce à une terminologie fixe : la [norme Darwin Core](#) pour la biodiversité. Ces termes normalisés font partie intégrante d'un processus visant à assurer la cohérence et la qualité des données. La norme Darwin Core est gérée par [Biodiversity Information Standards \(TDWG\)](#) (Normes d'information sur la biodiversité), également connu sous le nom de Taxonomic Databases Working Group (Groupe de travail sur les bases de données taxonomiques), une association scientifique et éducative à but non lucratif axée sur le développement de normes pour l'échange de données biologiques/biodiversité.

Les interactions avec l'infrastructure du GBIF sont facilitées par des Pôles de participants. Dans le Pacifique, le PROE fournit un Pôle régional de participants, l'équipe de coordination désignée pour établir et renforcer les activités liées au GBIF de ses pays membres et organisations partenaires. Les Pôles de participants sont également considérés comme des centres de connaissances pour les données sur la biodiversité, guidant les parties prenantes vers les sources pertinentes d'information et de données sur la biodiversité, et offrant leur propre expertise sur la biodiversité et la gestion des données.

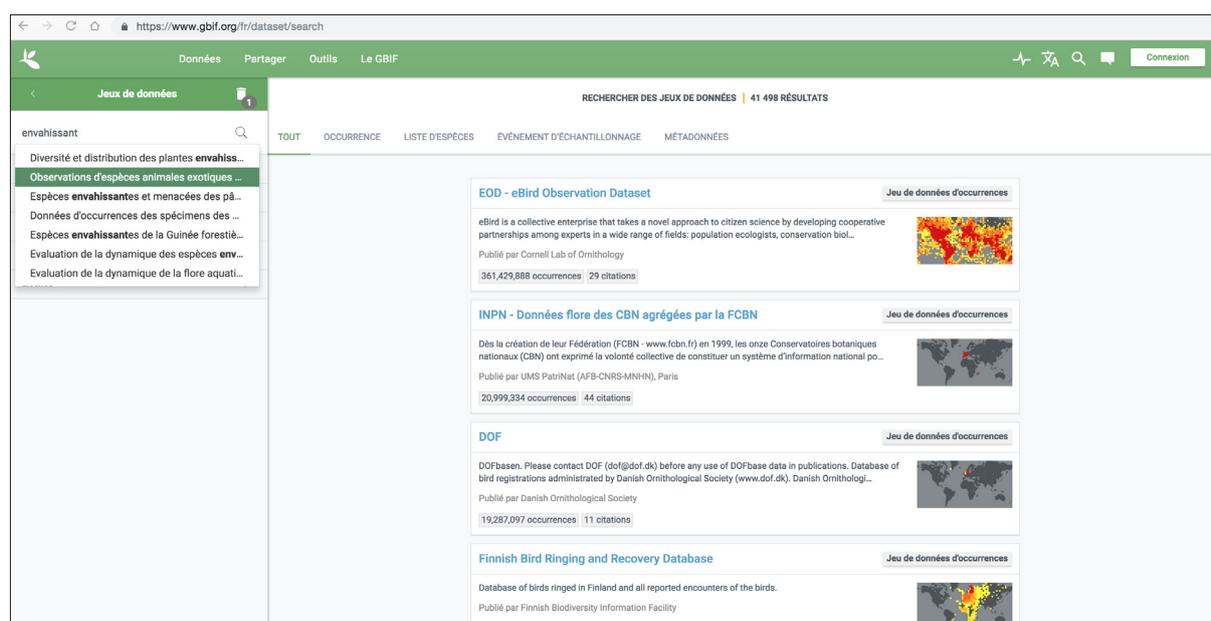
- En tant que Pôle participant, le PROE peut reconnaître les éditeurs de données sur la biodiversité. Les ministères, les organisations, les instituts ou les agences nationales désireuses de publier leurs données peuvent demander à être enregistrés auprès du GBIF. Le responsable du Pôle PROE recevra un courriel d'alerte l'invitant à amorcer le processus de vérification dès lors qu'un pays, une organisation ou un institut partenaire compétent en fait la demande. Après validation, ils peuvent publier ou actualiser des ensembles de données.
- L'objectif du PROE en tant que Pôle participant consiste à aider les organismes, les instituts et les organisations à publier des données et à les accompagner dans cette démarche.
- Remarque : pour le GBIF, l'un des critères clés veut qu'un « éditeur de données » soit une organisation plutôt qu'un individu. Les personnes souhaitant publier des ensembles de données doivent faire reconnaître leurs organisations affiliées en tant qu'éditeurs de données. Voir <https://www.gbif.org/become-a-publisher>

# Comment faire pour trouver des données dans le GBIF ?

Le site [www.gbif.org](http://www.gbif.org) contient plus d'un milliard de documents sur les espèces, les spécimens, des observations ainsi que des échantillons. Chaque document contiendra les instructions de téléchargement et de référence des données. Partez à la découverte du site ! Commencez par effectuer une recherche sur la page d'accueil ou cliquez sur « Données » dans le panneau supérieur. Vous pouvez rechercher des données d'occurrence, effectuer une recherche par espèce, rechercher des ensembles de données entiers, et plus encore.



Vous pouvez rechercher des ensembles de données entiers à l'aide de mots-clés. Dans cet exemple, nous recherchons des ensembles de données avec des titres contenant le terme « envahissant ». (Notez que vous ne trouverez peut-être pas toutes les données disponibles sur les espèces envahissantes de cette façon. Vous ne trouverez que les ensembles de données marqués avec ce mot exact.)



Vous pouvez rechercher des ensembles de données par région, appelés « Pays ou région qui publie » dans le panneau de gauche. Vous pouvez rechercher plusieurs pays à la fois et les associer à d'autres filtres : chaque case cochée sera incluse dans votre recherche.

The screenshot shows the GBIF search interface. On the left, the 'Jeux de données' panel is active, with filters for 'Fournisseur de données', 'Hôte', and 'Pays ou région qui publie'. Under 'Pays ou région qui publie', 'Tonga' (4) and 'Vanuatu' (8) are selected. Below this, a list of countries is shown with checkboxes and counts: France (1,127), Mexique (959), Royaume-Uni (825), Colombie (743), États-Unis d'Amérique (567), Japon (384), Australie (378), Brésil (339), and Bulgarie (300). The main search results area shows 12 results. The first result is 'List of the specimens of the Vanuatu National Herbarium (PVNH)' with 17,920 occurrences. The second is 'Checklist of the Vanuatu National Herbarium (PVNH)' with 1,894 enregistrements. The third is 'Priority Invasive Species in Tafea Province, Vanuatu' with 120 occurrences. The fourth is 'Reef fish species from Vava'u archipelago, Tonga, 2014' with 71 occurrences.

Pour ceux d'entre nous qui vivent dans le Pacifique, une astuce pour trouver des ensembles de données connexes de notre région consiste à consulter ceux hébergés par le Pôle du PROE. Ici, nous avons effectué une recherche en utilisant comme hôte le « Secretariat of the Pacific Regional Environment Program ». La recherche a donné 52 résultats à l'époque, et ce nombre augmente à mesure que vous et vos collègues publiez plus de données !

The screenshot shows the GBIF search interface with the 'Pays ou région qui publie' filter set to 'Secretariat of the Pacific Regional Environment Program'. The search results area shows 69 results. The first result is 'SPC NECTALIS Zooplankton/Micronekton specimens, New Caledonia 2014' with 12,454 occurrences. The second is 'SPC CoFish Invertebrates Observations Palau 2007' with 7,646 occurrences. The third is 'SPC PROCFish/C Fish Observations Fiji 2002-2003' with 7,612 occurrences. The fourth is 'Samoa Biodiversity Database' with 7,539 occurrences.

Cliquez sur le nom d'un ensemble de données pour en savoir plus. Vous pouvez également télécharger l'ensemble des données complet si vous le souhaitez, dans différents formats :

JEU DE DONNÉES D'OCCURRENCES | DATE D'ENREGISTREMENT 18 AVRIL 2018

## Invasive Species in Vanuatu

Publié par [Vanuatu Department of Environmental Protection and Conservation](#)  
 © Mimosa Bethel • Silverio Bule

ENSEMBLE DE DONNÉES | PROJET | STATISTIQUES | ACTIVITÉ | **TÉLÉCHARGER** | 12 OCCURRENCES

Archive annotée par GBIF  
 Archive d'origine Archive Darwin Core  
 Métadonnées annotées par GBIF EML

Invasive Species in Vanuatu collected during Country workshops funded by BID in 16 Feb, 2018 in Vanuatu – Thank you to GBIF and the BID programme for their support in mobilizing this dataset was funded by the European Union

DEPT OF ENVIRONMENTAL PROTECTION & CONSERVATION  
 Dernière modification des métadonnées: 15 mai 2018  
 Dernière modification de données: 15 mai 2018  
 Hébergé par: Secretariat of the Pacific Regional Environment Programme  
 Licence: CC0 1.0  
 Comment citer DOI: 10.15468/rux1ow

12 Occurrences | 100% Avec correspondance de taxon | 0% Avec coordonnées | 100% Avec l'année

Description  
 Couverture temporelle  
 Couverture géographique  
 Contacts  
 Description des données

Description  
 Invasive Species in Vanuatu collected during Country workshops funded by BID in 16 Feb, 2018 in Vanuatu – Thank you to GBIF and the BID programme for their support in mobilizing this dataset  
 Publication of this dataset was funded by the European Union

Il existe de nombreuses façons d'effectuer une recherche en utilisant des filtres, des mots-clés et plus encore. La méthode la plus courante consiste à chercher une seule espèce, et vous pouvez utiliser soit le nom scientifique (nom latin), soit le nom commun (vernaculaire). Cependant, il convient de noter que le nom vernaculaire peut varier d'un endroit à l'autre, ou que le même terme peut être utilisé pour différentes espèces. Par exemple, voici une recherche pour l'étoile de mer « Couronne du Christ ». Les résultats de la recherche recensent plus de 1 713 occurrences de l'étoile de mer. Pour obtenir les meilleurs résultats, vous pouvez effectuer plusieurs recherches en utilisant des termes différents.

https://www.gbif.org/fr/species/search?q=Couronne%20du%20Christ

Données | Partager | Outils | Le GBIF

RECHERCHER DES ESPÈCES | 1 RÉSULTAT

**Espèces**

Couronne du Christ

Vous faites une recherche de texte libre pour «Couronne du Christ »  
 Limiter la recherche à [Tout]

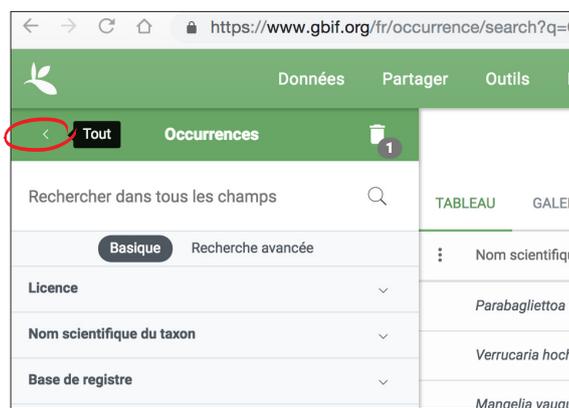
Niveau du taxon  
 Statut  
 Taxon supérieur  
 Problèmes et drapeaux

Vous recherchez des espèces dans la taxonomie GBIF.  
 RECHERCHER DANS LES LISTES D'ESPÈCES

**Lycium europaeum L.** | Espèces

Publié dans: Sp. pl. 1:192. 1753  
 Classification : Plantae > Tracheophyta > Magnoliopsida > Solanales > Solanaceae > Lycium  
 Nom vernaculaire: **couronne-du-Christ** in Français  
 Accepté | Espèces | 1,714 occurrences

Vous pouvez également rechercher dans « tout » (occurrences, espèces, ensembles de données entiers, etc.) en saisissant l'objet de votre recherche, puis en cliquant sur la flèche en haut à gauche ici :



## Mentionner le crédit dû, le cas échéant

**Toujours citer vos sources !** Chaque ensemble de données se voit attribuer un identificateur d'objet numérique (DOI) unique par GBIF (par exemple : <https://doi.org/10.15468/suj1kg>). Lorsque vous vous servez de ces données pour rédiger un rapport, indiquez également la référence donnée dans les métadonnées et/ou le DOI afin d'en préciser la source. Les résultats de votre recherche GBIF peuvent vous amener à télécharger et à utiliser des données provenant d'un grand nombre d'ensembles de données partagés par différentes organisations - par exemple, toutes les occurrences d'une espèce particulière observées ou recueillies dans un pays spécifique. Le cas échéant, le GBIF attribuera un DOI unique et spécifique au téléchargement lui-même, qui, une fois mentionné, attribuera le mérite à chacun des éditeurs de données contributeurs (y compris les liens directs vers la citation publiée à partir de la page de l'ensemble des données). Il vous suffit de suivre la citation recommandée fournie lorsque vous commandez le téléchargement.

## À qui appartiennent les données ?

Le principal moyen pour les utilisateurs d'accéder aux données du GBIF consiste à rechercher des attributs plutôt que des ensembles de données. Par exemple, vous pouvez recourir au GBIF pour connaître rapidement la distribution mondiale d'une espèce particulière. Cela signifie que les données seront souvent extraites de l'ensemble des données. Plutôt que d'extraire un ensemble de données complet, l'utilisateur trouvera des éléments de données provenant de plusieurs ensembles, et il devra ensuite déterminer l'origine de chaque élément. Pour ce faire, chaque donnée du GBIF est accompagnée de métadonnées associées.

Ces données sur les données, ou « métadonnées », doivent mentionner le propriétaire des données et la personne qui les a collectées, ainsi que la méthodologie appliquée. Veillez à rendre à César ce qui revient à César, et indiquez aux utilisateurs la personne de contact en cas de questions concernant l'utilisation de ces données.

Tous les dossiers d'événements publiés par le GBIF doivent être accompagnés d'une licence ouverte et informatisée ou d'une clause de renonciation sous les catégories Creative Commons CC0, CC-BY ou CC-BY-NC. Les conditions de licence standard permettent d'appliquer les conditions d'utilisation au téléchargement dans son ensemble.

En ce qui concerne les droits sur les données, un organisme tel que le GBIF n'a pas de droits sur les données, de même qu'il n'a pas accès à votre serveur contenant les données saisies. Les données appartiennent toujours à l'éditeur de données. L'éditeur de données octroie aux utilisateurs le droit de les exploiter, de même qu'il en détermine les conditions (selon la catégorie [Creative Commons](#)), tandis que le registre ou l'index, notamment le GBIF, fournit simplement l'infrastructure permettant de partager lesdites données. La responsabilité d'utiliser les données sous licence incombe à l'utilisateur des données. Il existe des publications expliquant aux utilisateurs de données comment être des citoyens modèles en suivant les normes communautaires d'attribution (p. ex., <http://vertnet.org/resources/norms.html>).

Lorsque vous partagez des données, celles-ci vous appartiennent toujours, mais elles seront accessibles à d'autres personnes effectuant des recherches.

Le partage des données permet à d'autres intervenants de la scène mondiale de suggérer des améliorations afin d'étayer ces données (avec votre autorisation bien évidemment), à les compléter ou à les associer différemment permettant de mieux comprendre la biodiversité mondiale.

Dans le système GBIF, les utilisateurs GBIF n'ont accès qu'aux données que l'éditeur accepte de partager en ligne. Vous pouvez publier des données ou des métadonnées uniquement. Vous pouvez également choisir de ne partager que certains éléments des données, notamment un emplacement approximatif plutôt qu'un lieu précis, ou de ne pas divulguer certains détails dans les champs contextuels pour des raisons de sensibilité culturelle ou de confidentialité.



## Le saviez-vous ?

**Les Données** sont des chiffres ou des faits bruts, sans interprétation ni analyse. En elles-mêmes, les données peuvent être difficiles à utiliser ou même à comprendre.

**L'Information** est une combinaison utile de données, couvrant contexte et interprétation. L'information est le résultat précieux que nous cherchons à obtenir sur les fondements de données pertinentes, d'analyses et de compréhensions cohérentes.



## Que sont les métadonnées ?

Les **métadonnées** sont un résumé de l'information de base d'un ensemble de données et de la façon elles ont été gérées. Elles montrent le « qui, quoi, quand et où » des données. Le GBIF impose ses propres exigences en matière de données et de métadonnées. Voir <https://www.gbif.org/data-quality-requirements>. Le PROE, en tant que Pôle participant du Pacifique, requiert les métadonnées suivantes pour chaque ensemble de données publié dans le GBIF :

### Les champs de métadonnées requis par le PROE

Intitulé de l'ensemble de données	
Description de l'ensemble de données	
Lien vers la source originale, si disponible	
Parties associées	
	Nom
	Poste
	Organisation
	Données de contact

## Que peuvent faire les utilisateurs agréés avec les données que je partage ?

Le partage contrôlé des données signifie que l'accès aux données est contrôlé, par le biais de mécanismes comme l'accès payant ou l'accès protégé par mot de passe, ou que l'utilisation des données est contrôlée, par des mécanismes comme les licences de données.

Le contrôle de l'accès, par des mécanismes tels que des services payants, des mots de passe ou un nombre limité de copies d'un ensemble de données sur un ou plusieurs ordinateurs uniquement, nécessite une capacité de stockage et de gestion des données sur le long terme, ainsi qu'une réponse aux demandes d'utilisations. Il est possible de ne publier que des informations sur les données (métadonnées) et d'inviter les utilisateurs à demander l'accès aux ensembles de données d'origine, bien que cette démarche nécessite à nouveau des ressources qui seraient responsables du suivi.

Le système GBIF repose sur le libre accès aux données et la licence [Creative Commons](#). Les utilisateurs ont le droit d'utiliser les données d'une manière ou dans un but précis. Ce système impose à l'utilisateur des données la responsabilité de comprendre et de respecter les règles de partage.

Tous les dossiers d'occurrences publiés par le GBIF doivent être accompagnés d'une licence ouverte et informatisée ou d'une clause de renonciation sous les catégories [Creative Commons CC0](#), [CC-BY](#) ou [CC-BY-NC](#), avec en général, des données gratuites servant des fins « équitables ». La licence CC0 (domaine public) est recommandée dans la mesure du possible. CC0 désigne un ensemble de données relevant du domaine public, que les utilisateurs sont libres de consulter de quelque manière que ce soit. Cependant, il est rare dans la communauté que les utilisateurs extraient et exploitent des données à des fins lucratives. En outre, il est peu probable que l'éditeur de données dispose de toutes les ressources nécessaires pour assurer le suivi des utilisateurs malveillants ou maladroits. Il est donc essentiel de vérifier la législation locale et de confirmer la licence des données auprès des personnes qui les ont collectées. Plus d'information sur : <http://www.environment.gov.au/marine/gbr/reef-trust>.

Pour plus d'informations sur les catégories de libre accès, les droits en matière de connaissances et la gestion des données sur l'information environnementale dans le Pacifique, nous vous encourageons à prendre contact avec le projets [Inform \(www.sprep.org/inform/home\)](http://www.sprep.org/inform/home) et [Access and Benefit Sharing \(ABS\) Project \(www.sprep.org/abs/about\)](http://www.sprep.org/abs/about), lesquels ont pour objet de faciliter la gestion des connaissances en vertu du Protocole de Nagoya, deux partenariats actifs avec les États et Territoires insulaires océaniques.



## Comment publier des données dans le GBIF ?

Vous devez d'abord enregistrer votre organisation auprès du GBIF, en indiquant que vous utiliserez les outils d'édition intégrés du PROE (voir ci-après). Après vous être inscrit en tant qu'éditeur de données, suivez cette procédure étape par étape pour publier vos données sur GBIF :

- 1 Saisissez les métadonnées de votre ensemble de données sur l'outil d'édition intégrée (IPT) du PROE.
- 2 Préparez votre jeu de données source Darwin Core (voir ci-après : « Comment préparer les données pour la publication »).
- 3 Téléchargez l'ensemble de données vers l'outil d'édition intégrée (IPT).
- 4 Associez votre ensemble de données aux termes de Darwin Core, sous forme de liste de vérification ou d'occurrences de données (selon vos données)
- 5 Publiez.
- 6 Validez l'ensemble de données au moyen du validateur de données GBIF, avant d'enregistrer l'ensemble de données (<https://www.gbif.org/tools/data-validator>).
- 7 Enregistrez l'ensemble de données auprès du GBIF (première fois seulement).
- 8 Sollicitez les commentaires du GBIF et des utilisateurs.
- 9 Actualisez vos données source.
- 10 Répétez à partir de l'étape 4. Notez que la responsabilité de la maintenance des ensembles de données incombe à l'éditeur de données.



### Que se passe-t-il si mes données ont été publiées par une autre institution ?

Les ensembles de données comportent des identifiants uniques. Si l'ensemble de données que vous souhaitez partager figure déjà dans le GBIF, ne poursuivez pas, car le GBIF identifiera l'ensemble de données comme étant redondant. Une meilleure solution, en tant qu'éditeur, serait d'établir un lien avec les données déjà téléchargées par l'autre institution.

N'oubliez pas que l'éditeur n'est pas nécessairement le seul ou le propriétaire original de l'ensemble de données, tel que défini dans les métadonnées, mais seulement la personne qui a effectué le travail de publication par l'entremise du GBIF. Si une autre organisation a fait le travail à votre place, ne vous privez pas et voyez-y un atout !

## Voies d'accès à la publication de données

Comment décider du contenu à publier ? Tout d'abord, tenez compte du type de données sur les espèces le plus pertinent et le plus utile pour vous. Le partage de ces données peut vous aider à obtenir une version stable et enregistrée, de même que d'autres pourraient en tirer profit.

Tenez compte des objectifs de vos projets actifs. Outre les nouvelles données, en existe-t-il d'autres, plus anciennes que vous auriez besoin de comparer avec de nouveaux résultats ? Au cours de l'élaboration du projet, un utilisateur a-t-il dû déterrer d'anciens fichiers pour tenter d'identifier l'ampleur du problème ? Ces données peuvent valoir la peine d'être partagées.

Bien que les données sur les espèces endémiques ou les nouvelles aires demeurent intéressante, celles qui portent sur les espèces présentes dans les aires courantes se révèlent tout aussi utiles pour recenser et identifier les invasions potentielles.

Les données qui se trouvent déjà dans des tableaux ou des feuilles de calcul sont idéales, mais les données sous n'importe quel format, peuvent également être présentées sous forme de colonnes et de lignes. Ainsi, il se peut que vous ayez un rapport décrivant verbalement la découverte de certaines espèces dans une ou plusieurs aires spécifiques. Vous pouvez convertir ces informations en un ensemble de données comportant le nom de l'espèce correspondant à l'aire, l'année d'observation, et ainsi de suite.

Vous pouvez décider s'il s'agit de données d'occurrence, de données en liste d'identification ou de données d'échantillons. Chacune de ces catégories nécessite des volumes et des détails de données différents. Pour de plus amples informations, voir <http://rs.tdwg.org/dwc/terms/> and [www.gbif.org/data-quality-requirements](http://www.gbif.org/data-quality-requirements).



### Le saviez-vous ?

**Les données d'occurrence** sont la preuve de l'existence d'un organisme à un endroit et à un moment précis.

**Les ensembles de données en liste d'identification** sont une liste de taxons. Ces données n'incluent que l'ID taxon (taxonID, UUID) et le nom scientifique (scientificName) en termes Darwin Core.

**Les ensembles de données provenant d'un échantillon d'événements** fournissent davantage de précisions sur une espèce présente à un endroit et à une date donnés, y compris les méthodes, les événements et l'abondance relative des espèces enregistrées dans un échantillon.



### L'étude de cas : Kosrae

Nous disposons d'un ensemble de données comportant de nombreuses variables, dont le nom scientifique, le nom vernaculaire, l'écosystème, la voie d'entrée, le statut (qui comprend entre autres « introduit ») et les estimations d'abondance. Dans ce contexte, l'ensemble de données comportait plus de variables qu'il n'en fallait pour une liste d'identification ou un ensemble de données d'occurrence dans le GBIF. Si ces données devaient être publiées sous la forme d'une liste d'identification, le nettoyage des données impliquerait d'extraire uniquement les valeurs sous « nomScientifique », d'ajouter les métadonnées sur l'organisation et l'ensemble de données, et de publier. Le nettoyage de ces données, à des fins de publication en tant qu'ensemble de données sur les événements, était un peu plus compliqué. Certains termes ont dû être normalisés : par exemple, la catégorie « Écosystème » a été remplacée par le terme « habitat » selon le terme Darwin Core car les valeurs indiquaient l'emplacement d'une espèce : les données comprenaient « terrestre », « marin », « eaux stagnantes » et « zone humide / terrestre ». Le terme Statut indiquait si l'espèce était native de Kosrae ou si elle avait été introduite, de sorte qu'il a été modifié selon le terme Darwin Core (establishmentMeans). Voir les résultats ici. <https://doi.org/10.15468/ypzayn>

# Comment s'inscrire en tant qu'éditeur de données dans le GBIF ?

La première étape pour partager des données avec le GBIF consiste à vous inscrire, en identifiant votre organisation comme ayant des données de qualité à partager. Une fois inscrit, vous pouvez publier autant d'ensembles de données que vous le souhaitez.

L'exigence du GBIF pour l'enregistrement est d'avoir accès à un Pôle - pour les îles du Pacifique, le Pôle PROE et d'accéder à l'Integrated Publishing Toolkit (IPT, Outil d'édition intégrée). Vous pouvez vous inscrire comme éditeur à tout moment, vous n'avez pas besoin de données prêtes à être publiées. Un éditeur peut également être enregistré pour donner accès à des métadonnées, même s'il n'est pas prêt à publier un ensemble de données.

Un éditeur peut être un ministère, un institut, une organisation, etc. Par souci de simplicité, nous utiliserons le terme « institution » tout au long de ces consignes.

**Étape 1** accédez au site Internet GBIF sur [www.gbif.org](http://www.gbif.org). Cliquez sur l'onglet *Partager – Devenir éditeur* et cherchez votre institution afin d'en vérifier l'enregistrement. Si votre institution n'est pas inscrite, passez à l'étape 2.

**Étape 2** Il y a un contrat avec l'éditeur de données à prendre en considération. Acceptez et continuez.



**Étape 3** Précisez les informations demandées (nom de l'institution, contact principal, adresse du site Internet, description de l'organisation, emplacement sur la carte, etc).

- N'oubliez pas de sélectionner que le Pôle de soutien, ainsi que l'accès à un IPT, se fera via le Pôle PROE. Le site vous demandera si vous avez besoin d'aide pour la publication - la réponse ici est « Non ». Le responsable du Pôle PROE vous dispense déjà cette assistance.
- Données de contact : Fournissez vos coordonnées au sein de votre institution. Il peut également s'avérer utile de fournir également un courriel de contact général (p. ex., info@.....), qui perdurera même si un membre du personnel quitte l'institution. Le point de contact technique aura besoin des coordonnées du responsable du Pôle PROE. Veuillez noter que si vos coordonnées changent, vous pouvez vous adresser directement au service d'assistance du GBIF pour les actualiser.
- Donnez une brève description de votre institution.
- Sélectionnez « Non » à la question « Avez-vous l'intention d'installer et d'exécuter un logiciel (tel que l'Integrated Publishing Toolkit – IPT) pour publier vos données directement sur GBIF.org ? »

**Étape 4** Soumettre

**Étape 5** Un courriel de notification sera envoyé à la personne de contact (Éditeur de données) ainsi qu'au PROE, en tant que Pôle participant du GBIF.

**Étape 6** Une fois que le Pôle participant aura validé la demande de l'institution pour le GBIF, l'inscription sera alors finalisée et un nouvel éditeur sera ajouté sur le site du GBIF.

**Étape 7** Le GBIF enverra à la personne de contact un mot de passe pour l'institution nouvellement inscrite. Ce mot de passe sera demandé par le responsable du Pôle en vue de faciliter l'enregistrement de nouveaux ensembles de données au GBIF par le biais de l'IPT.

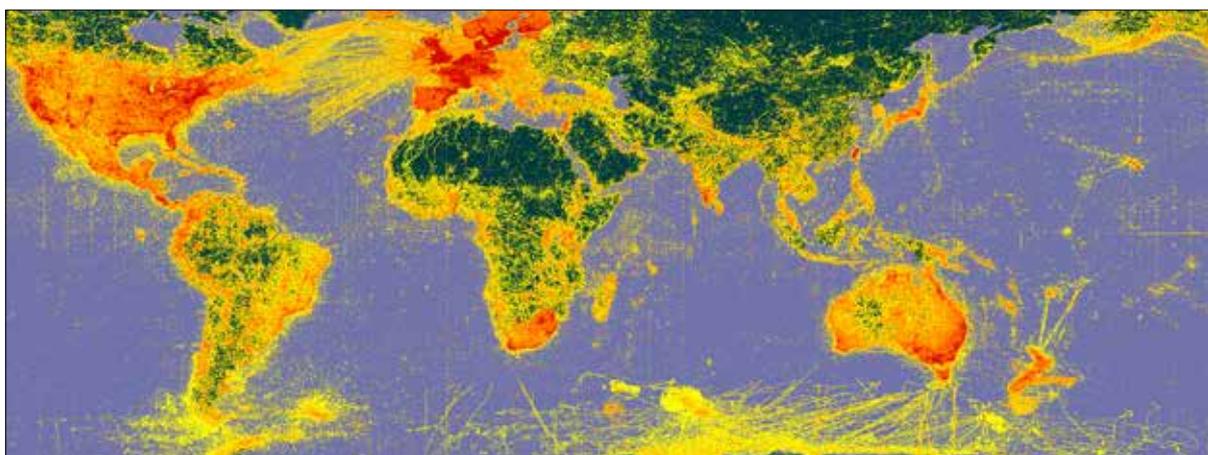


## Géoréférencement

Le géoréférencement précise le « où » pour chaque donnée de l'ensemble, ce qui permet de situer les données sur une carte et de les utiliser à des fins d'analyse spatiale. Le géoréférencement consiste à marquer un emplacement, ou une référence, pour chaque donnée. Par exemple, pour un recensement d'oiseaux, nous pourrions consigner les observations par nom d'espèce mais aussi par emplacement, en utilisant des coordonnées.

- Il est important d'ajouter des valeurs de latitude et de longitude pour donner un contexte spatial aux données.
- Les bonnes pratiques en matière de géoréférencement favorisent l'intégration d'au moins quatre champs Darwin Core primordiaux: `decimalLatitude`, `decimalLongitude`, `geodeticDatum`, et `coordinateUncertaintyInMeters`.
- Le référentiel géodésique est un modèle mathématique représentant la surface du globe et sur lequel sont basées des coordonnées. Il existe plus de 200 données géodésiques : certaines ne sont utilisées que localement, d'autres, comme le fameux WGS84 utilisé par Google Maps et d'autres systèmes, couvrent le globe.
- `Uncertainty in Meters` décrit la dimension d'un lieu décrit. La `coordinateUncertaintyInMeters` correspond à la distance maximale que vous pouvez parcourir depuis l'emplacement des coordonnées données tout en restant dans la zone qui fait l'objet de la description. Cette mesure indique aux utilisateurs des données, en une variable, à quel point l'emplacement est spécifique, et ce, d'une manière applicable à l'ensemble de la planète. L'incertitude peut être facilement déterminée à l'aide de Google Maps au moyen d'un outil de mesure (clic droit sur la carte) qui permettra de tracer une ligne sur la section la plus large du lieu. La moitié de cette distance est la `coordinateUncertaintyInMeters`, et le point médian de cette ligne correspond à l'emplacement des coordonnées.
- Le géoréférencement peut s'avérer un processus complexe. Veuillez consulter le [Guide de référence rapide sur le Géoréférencement](#) pour comprendre cette fascinante sous-discipline de l'informatique sur la biodiversité.

Il existe une variété d'outils facilitant le géoréférencement - voir [www.georeferencing.org](http://www.georeferencing.org).



## Comment préparer les données à la publication ?

La préparation des données consiste à extraire l'information de différentes sources, à créer les ensembles de données sur les occurrences et les listes d'identification, et à établir une correspondance avec les termes Darwin Core afin de faciliter la recherche et la compréhension.

Il existe 3 types élémentaires d'ensembles de données agréés par le GBIF : Occurrence, événement et Taxon ou liste d'identification. La catégorie Taxon désigne essentiellement un type d'ensemble de données contenant une liste taxonomique. Les métadonnées décriraient la portée de la liste. L'existence d'un taxon à un endroit et à un moment spécifiques - une occurrence - désigne un concept crucial, car ces données fournissent une description plus riche de l'existence des taxons dans le monde. Un ensemble de données basé sur les événements est destiné à la surveillance, où les sites sont échantillonnés ou observés de façon répétée au fil du temps, avec des informations telles que l'abondance/la densité et les mesures ou les faits concernant lesdits événements. Notez que les ensembles de données de la liste d'identification nécessitent un identificateur unique (TaxonID) et le nom de l'espèce (ScientificName), tandis que les ensembles de données sur les occurrences nécessitent l'information qui/quoi/quand/où. Vous trouverez les données requises pour les différents types d'ensembles de données sous : [www.gbif.org/data-quality-requirements](http://www.gbif.org/data-quality-requirements)

La première étape consiste à s'assurer que les données sont présentées dans un format prêt pour l'IPT, c'est-à-dire un fichier délimité tel qu'un fichier Excel ou une connexion à une base de données.



### Palaos - Étude de cas

Il est souvent possible d'extraire des données d'un texte ou de combiner des données de plusieurs rapports différents pour créer l'ensemble de données souhaité pour la publication. Ici, un article scientifique publié contient des données que nous souhaitons ajouter au GBIF. Le document comporte un identificateur unique, son DOI, nous allons donc lister cela dans la cellule associatedReferences.

**Table 4.** Detected introduced, cryptogenic and potentially introduced species during the preliminary Palau introduced species survey. Species are listed with an indication of the sampling locations/sites (wharves, moorings, vessel hulls, or "pristine" locations without commercial activity). A "•" indicates presence upon a substrate. Please note that scientific names and taxonomic authorities (for species names) were verified using the WoRMS database (<http://www.marinespecies.org/>).

Phyla	Species	Status	Wharves	Vessels	Mooring	Pristine
Porifera	<i>Haliclona caerulea?</i> (Hechtel, 1965)	Potential	•		•	•
	<i>Mycale</i> sp. (orange sponge)	Potential	•			
Hydroida	<i>Eudendrium carneum</i> Clarke, 1882	Introduced	•			
	<i>Obelia</i> sp.	Cryptogenic	•	•	•	
	<i>Thyroscyphus fruticosus</i> (Esper, 1793)	Introduced	•		•	
Polychaeta	<i>Sabellastarte</i> sp.	Potential	•			
	Serpulidiidae	Potential	•	•		
Cirripedia	<i>Amphibalanus amphitrite</i> (Darwin, 1854)	Cryptogenic (cosmopolitan)	•	•		
	<i>Chthamalus proteus</i> Dando and Southward, 1980	Introduced			•	•

Source : Campbell et al. (2016), avec l'aimable permission des auteurs.

Certaines des données que nous souhaitons intégrer se présentent déjà sous forme de tableau. Chaque espèce enregistrée que nous extrayons de ce tableau sera associée à une cellule de Références indiquant la référence de l'article en question. Dans les métadonnées, la propriété des données sera également indiquée avec les références appropriées pour chaque source de données.

### Astuce

Prenez l'habitude de joindre un identificateur (occurrenceID pour les ensembles de données d'occurrence et taxonID pour les ensembles de données taxonomiques) à chaque enregistrement de vos ensembles de données lors de la phase de préparation. Ces identificateurs seront requis pour la correspondance de vos données dans l'IPT. Vous pouvez les générer vous-même, ou utiliser un site UUID (Universal Unique Identifier) (p. ex., [www.uuidgenerator.net](http://www.uuidgenerator.net)) qui s'en chargera pour vous.

## Préparation des données à l'aide de la Boîte à outils d'édition intégrée

The IPT is a software tool installed on a server. The IPT we use in this instance is the SPREP IPT. The SPREP Node Manager is committed to assist with the initial publishing of datasets. In addition to this guide, there will be on-going technical support from the SPREP Node Manager, as well as assistance from GBIF and BID mentors as necessary.

- Une fois qu'une institution est enregistrée avec succès, le responsable du Pôle PROE l'ajoutera à la liste d'organismes habilités à publier des ensembles de données sur leur serveur, au moyen du mot de passe envoyé au contact de l'éditeur de données.
- L'éditeur convertira ses ensembles de données en une liste d'identification ou un ensemble de données d'occurrences en utilisant les termes Darwin Core. Vous trouverez un guide de référence rapide pour les termes Darwin Core ici : <http://rs.tdwg.org/dwc/terms/>

Une fois que vous aurez préparé votre ensemble de données et que vous l'aurez téléchargé dans l'IPT, vous pourrez procéder au mappage champ par champ des termes de Darwin Core et publier ces données. Notez que vos données sources sous leur forme originale seront également archivées sur l'IPT.

Le GBIF a un wiki IPT et y propose de la documentation ainsi que des tutoriels. Une démonstration vidéo de la création d'une nouvelle ressource de type Occurrence avec des extensions figure également sur la page IPT principale du GBIF : [www.gbif.org/ipt](http://www.gbif.org/ipt).

L'IPT est un logiciel installé sur un serveur. L'IPT que nous utilisons dans ce cas est celui du PROE. Le Responsable du Pôle PROE s'est engagé à contribuer à la publication initiale des ensembles de données. Outre ce guide, l'utilisateur bénéficiera d'un soutien technique continu de la part du responsable du Pôle PROE, ainsi que de l'aide des mentors du GBIF et du DIB au besoin.



### Champs Darwin Core recommandés pour un ensemble de données portant sur une simple occurrence

Termes Darwin Core et définitions :		<a href="http://rs.tdwg.org/dwc/terms/">http://rs.tdwg.org/dwc/terms/</a>	
Quoi ? Taxon	Où ? Lieu	Quand ? Événement ?	Autre
scientificName	decimalLatitude	année	occurrenceID (UUID)
	decimalLongitude	mois	basisOfRecord
	geodeticDatum	jour	organismQuantity
	coordinateUncertaintyInMeters	eventDate	organismQuantityType
	country	verbatimEventDate	establishmentMeans
	countryCode		dynamicProperties
	island		
	islandGroup		
	stateProvince		
	waterBody		

## Mappage avec les termes Darwin Core

Le mappage est l'appariement des données des champs de données sources aux champs de la norme Darwin Core. Il permet de passer du jeu de données original à un format exploitable par le GBIF.

**Les termes Darwin Core** correspondent à une liste de termes ou de domaines d'intérêt en matière de biodiversité. L'utilisation de termes normalisés permet de comparer des ensembles de données provenant de lieux, de périodes et de pays différents, souvent à l'aide de recherches automatisées. Vous trouverez un guide de référence rapide pour les termes Darwin Core ici : <http://rs.tdwg.org/dwc/terms/>.

Ce guide comprend des définitions et des exemples de chaque terme. Les termes sont organisés par catégories (en gras) dans l'index. Les catégories correspondent aux termes Darwin Core correspondant à des classes (termes assortis d'autres termes pour les décrire). Les termes qui décrivent une classe donnée (les propriétés de la catégorie) apparaissent dans la liste qui suit immédiatement le nom de la catégorie dans l'index. L'index fournit des liens vers les descriptions des termes dans le tableau en dessous de l'index.

L'IPT permet soit de mapper un champ source à un champ Darwin Core, soit de définir la valeur d'un champ sur une constante. Pour les informations qui n'ont pas de terme Darwin Core correspondant, il est toujours possible de les ajouter dans le dynamicProperties du terme Darwin Core mais elles peuvent aussi exister dans une extension Darwin Core et y être rajoutées.



### Darwin Core pour les espèces envahissantes

Certains termes Darwin Core sont particulièrement pertinents pour la description des espèces envahissantes, bien qu'ils ne couvrent pas tous les aspects d'une espèce. Par ailleurs, ils pourraient nécessiter une révision et une actualisation, de même que le vocabulaire associé.

Ce que les cadres dirigeants souhaitent savoir...	Terme Darwin Core	Suggestion de terminologie
Comment cette espèce est-elle arrivée là ?	establishmentMeans	indigène, introduit, naturalisé, envahissant, géré ou incertain
Où vit cette espèce ?	occurrenceStatus	présent, absent, commun, irrégulier, rare, douteux ou exclu (vocabulaire GBIF uniquement). (Cependant, dans la pratique, certaines personnes incluent l'état de conservation dans l'occurrenceStatus.)
S'agit-il d'une espèce indigène ?	<i>Proposition de terme :</i> Origine	<i>Proposition :</i> indigène, réintroduit, introduit (éventuellement subdivisé en avant ou pendant l'ère moderne), vagabond ou inconnu
Dans quelle mesure cette espèce est-elle bien intégrée ?	<i>Terme proposé :</i> degreeofEstablishment	<i>Proposition :</i> indigène, captif, cultivé, transporté, relâché, occasionnel, reproducteur, établi, dispersé, colonisateur, envahissant

Pour plus de détails, voir <https://github.com/tdwg/dwc-qa/wiki/Webinars#chapter4>

Voici quelques exemples de mappage des termes Darwin Core. Les définitions se trouvent dans la section [Guide de référence rapide](#) :

- Le nom d'« Espèce » devient **scientificName** (nom scientifique).
- Le « nom tongien » devient **vernacularName** (notez la perte d'information quant au terme dans sa langue initiale, ce qui devrait être spécifié dans les ensembles de données dynamicProperties pour les ensembles de données Occurrence ou dans une extension Noms vernaculaires pour un ensemble de données Taxon).
- **Propriétés dynamiques** : d'autres informations qui ne rentrent pas dans la norme Darwin Core peuvent être placées ici. Vous pouvez ainsi enregistrer des champs d'informations, des données structurées, sans perdre le contexte. Les variables des espèces envahissantes peuvent être placées dans ce champ. Des liens vers des références associées peuvent être repris ici, par exemple sous forme d'un lien permettant de télécharger l'ensemble de données sources au format PDF.
- **identificationRemarks** – toute remarque sur le processus d'association d'un Taxon à une entité observée ou collectée.
- **previousID** – tout nom scientifique précédemment appliqué à une entité, où le scientificName est le nom actuellement accepté ou valide.

« Nous élaborons un plan d'action et collaborons avec différentes organisations pour rendre ces données disponibles. Je souhaite vraiment attirer l'attention sur le problème des espèces envahissantes dans les Salomon. »  
– Josef Hurutarau, Îles Salomon



## Étude de cas : Îles Salomon

Dans ce cas, toutes les bonnes données étaient là, mais les libellés ne correspondaient pas aux termes Darwin Core. Les termes que vous utilisez dans la collecte de données, ou les libellés fournis par les chercheurs eux-mêmes, ne sont souvent pas normalisés à des fins de comparaison mondiale. Dans ce cas, le nettoyage des données était aussi simple que le changement de nom des colonnes.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Santal	scientificName	scientificName	accepted	scientificName										
2	13467	Solomon Islands	Abelmoschus manihotus	Mnäh			Plantae	terrestrial	alien	not specified					
3	53468	Solomon Islands	Acacia auriculiformis	Beotti			Plantae	terrestrial	alien	invasive					
4	53469	Solomon Islands	Acacia farnesiana	Willd. (L.)			Plantae	terrestrial	alien	invasive					
5	53470	Solomon Islands	Acacia sp.	A63			Plantae	terrestrial	alien	not specified					
6	53471	Solomon Islands	Actinidia julica	(Fernald, 1871)			Animalia	terrestrial	alien	invasive					
7	53472	Solomon Islands	Actinidia julica	(Fernald, 1871)			Animalia	terrestrial	alien	not specified					
8	53473	Solomon Islands	Actinidia julica	L.			Plantae	terrestrial	alien	not specified					
9	53474	Solomon Islands	Acacia julica	Cass. (Tw.)			Plantae	terrestrial	alien	not specified					
10	53475	Solomon Islands	Acridotheres tristis	(Linnaeus, 1766)			Animalia	terrestrial	alien	invasive					

## Définition d'un identificateur unique

Chaque enregistrement de données sur une occurrence nécessite un identificateur unique. Il est vivement recommandé que votre ensemble de données sources comporte des identificateurs globalement uniques et constants (tels que les UUID) pour les valeurs d'occurrenceID et qu'ils se conservent ensuite dans l'enregistrement. Si vous utilisez Excel, commencez par attribuer un identifiant à chacun de vos enregistrements d'occurrence (c'est-à-dire, chaque champ ou chaque ligne individuelle) ; cet identifiant se retrouvera dans l'IPT sous l'occurrenceID du terme Darwin Core. Cette étape est cruciale dans le GBIF pour les personnes cherchant à identifier un document spécifique dans votre ensemble de données une fois publié. Des identificateurs uniques globaux qui sont des UUID permettront de faire facilement référence à des documents spécifiques, et ceux-ci pointeront ultérieurement, sans la moindre ambiguïté, vers des documents donnés, et ce à tout moment.

- Le responsable du Pôle PROE peut vous aider en générant un UUID global à l'aide des ID que vous avez peut-être déjà joints à vos documents et en faisant intégrer l'UUID global dans les données sources. Ces identificateurs doivent impérativement accompagner les documents à la source et ne pas changer.
- De même, les documents d'un ensemble de données de type Liste d'identification nécessitent des identificateurs uniques pour le champ taxonID. Ceux-ci peuvent être générés à l'aide d'un site UUID, [www.uuidgenerator.net](http://www.uuidgenerator.net) et saisis dans les dossiers Taxon.
- Si vous utilisez une base de données autre qu'Excel, elle peut, peut-être, générer automatiquement un UUID. Excel peut également être programmé pour générer un UUID (voir [www.idigbio.org/wiki/images/0/03/GUIDgeneration.pdf](http://www.idigbio.org/wiki/images/0/03/GUIDgeneration.pdf)).
- Votre jeu de données dans son intégralité recevra un DOI dès son enregistrement et sa publication auprès du GBIF (par exemple, Forêt tropicale humide Toloa de Tonga = doi:10.15468/suj1kg).

## Nettoyage des données

Le nettoyage d'un ensemble de données peut aller de la simple vérification visant à s'assurer que toutes les informations sont incluses de manière standardisée jusqu'à combler les lacunes dans les données ou définir de nouveaux éléments que vous aimeriez y intégrer. Il existe de nombreuses méthodes pour nettoyer les données à l'aide de Microsoft Excel. (Vous trouverez une liste de formules utiles ici : [www.sprep.org/attachments/Publications/IOE/gbif-sprep-invasive-species-data-set-templates-resources.xlsx](http://www.sprep.org/attachments/Publications/IOE/gbif-sprep-invasive-species-data-set-templates-resources.xlsx). Open Refine est un excellent outil de nettoyage des données. Voir le *Guide d'utilisation de base Open Refine pour le nettoyage de données sur la biodiversité* sur <https://tinyurl.com/ORforBiod>.

Les étapes fondamentales sont la standardisation et la mise à jour.

- **La normalisation** consiste à prendre des champs comme les codes de pays (par exemple TO/TON pour Tonga) et à les normaliser en un seul code (dans ce cas, TO).
- On parle de **mise à jour** des données lorsque si, par exemple, vous n'avez qu'un nom scientifique pour une espèce dans un ensemble de données, et pour le rendre plus complet vous ajoutez le règne et l'embranchement

Le géoréférencement peut également être effectué pendant l'étape de nettoyage et de mise à jour – le nettoyage peut compléter tous les niveaux géographiques. Les utilisateurs recherchent toutes sortes de termes, vous pouvez donc en utiliser plusieurs pour décrire un lieu, comme le Pacifique Sud, les Tonga, les îles, etc.

Au fur et à mesure que vous nettoyez chaque partie des données, comparez-les aux termes standard utilisés ailleurs par des experts réputés.

## Re-publication de données

Une fois que l'ensemble de données sources est mappé via l'IPT, publié et enregistré auprès du GBIF, vous pouvez visualiser l'ensemble de données publié dans le GBIF. Il se peut alors que certaines des données affichées dans le GBIF n'incluent pas toutes les informations souhaitées, que vous ayez besoin d'en ajouter davantage dans le géoréférencement, que vous ayez utilisé le mauvais terme Darwin Core ou encore que les contrôles de qualité des données du GBIF aient identifié des problèmes potentiels. Vous aurez alors l'option de passer par le processus de traitement des problèmes dans les données et de les republier.

Nous présenterons un **exemple d'ensemble** de données provenant de la forêt tropicale humide de Toloa, aux Tonga. Vous pouvez télécharger le fichier Excel ici. [www.sprep.org/attachments/Publications/IOE/invasive-species-toloa-rainforest.xlsx](http://www.sprep.org/attachments/Publications/IOE/invasive-species-toloa-rainforest.xlsx)

- La forêt tropicale humide de Toloa des Tonga publiée sur le GBIF avait quelques problèmes non résolus : 3 incompatibilités potentielles avec la norme Darwin Core. Il a été décidé de republier les données après quelques améliorations.
  - Le support technique de l'éditeur de données avait supposé à tort que le champ « habitude » était le terme Darwin Core « habitat ». Cette erreur fut corrigée.
  - L'une des alertes portait sur un nom scientifique incorrect - peut-être une simple faute d'orthographe marquée en rouge par le GBIF. Pour corriger ce problème, nous pouvons utiliser le site Global names resolver : <http://resolver.globalnames.org/>). Avant toute chose, vérifiez auprès du responsable des données s'il est possible d'apporter ces améliorations.
  - La valeur « Nom tongien » devait figurer dans le champ Darwin Core vernacularName.
- **Mise à jour des données sur la forêt tropicale humide Toloa de Tonga :**
  - Connectez-vous à l'IPT.
  - Choisissez l'onglet Gérer les ressources.
  - Sélectionnez l'ensemble de données (appelé ressource dans l'IPT) à mettre à jour.
  - Dans la section Données sources, sélectionnez Fichier et téléchargez, remplaçant la version précédente.
  - L'écran de révision des Données sources vous permettra de définir les caractéristiques du fichier téléchargé et de sélectionner des feuilles de calcul spécifiques. Vous pouvez consulter les données et vérifier si elles sont interprétées comme prévu.
  - Cliquez sur Enregistrer, et les données se trouvent désormais dans l'IPT, prêtes au mappage des champs.
  - Dans les sections de Mappages Darwin Core, modifiez les mappages afin de déterminer les champs mappés et ceux qui ne l'ont pas été - par exemple, les corrections suivantes :
    - Espèces doit être répertorié sous leur scientificName.
    - Habitude doit être répertorié sous dynamicproperties (étape à réaliser avant le téléchargement).

- Nom tongien doit être répertorié sous vernacularName.
- Statut doit être répertorié sous establishmentMeans.
- Dominance 2014 doit être répertorié sous dynamicProperties (étape à réaliser avant le téléchargement).
- Invasif doit être répertorié sous dynamicproperties (étape à réaliser avant le téléchargement).
- PreviousID doit être répertorié sous previousIdentification.
- Les remarques ID doivent être répertoriées sous identificationRemarks.
- Les champs récemment ajoutés Latitude et Longitude, Date et coordinateUncertainty doivent être répertoriés sous decimalLatitude, decimalLongitude, geodeticDatum, et coordinateUncertaintyInMeters.
- Sauvegarder et revenir à la page Vue d'ensemble.

Maintenant que tout est mappé, vous pouvez publier en cliquant sur le bouton Publier dans la section Versions publiées. Nous vous invitons à joindre un résumé des démarches entreprises dans le processus de republication ; par exemple, « ajout de nouveaux champs pour les géoréférences et dynamicProperties. Deux mises à jour taxonomiques basées sur les problèmes découverts lors du processus d'intégration au GBIF. »

Vos données sont désormais republiées. Dans cet exemple de Tonga, nous avons mis à jour un ensemble de données existant.

L'outil de validation des données (<https://www.gbif.org/tools/data-validator>) permet de vérifier les problèmes de qualité avant la publication. Une fois l'ensemble de données publié, l'onglet « métriques » affiche un résumé du contenu, y compris, par exemple, la distribution taxonomique et les problèmes de qualité. Pour un exemple, voir <https://www.gbif.org/dataset/7c0cd863-8b81-4937-84f9-2f596fd3fa79/metrics>.

D'autres commentaires peuvent provenir du GBIF, soit de ses rapports sur la qualité des données, soit d'utilisateurs tentant d'utiliser vos données, y compris parfois des améliorations. Après avoir consulté ces commentaires, vous pouvez apporter les modifications nécessaires et republier au besoin.



« La tâche que nous avons accomplie ensemble représente notre héritage pour le peuple des Tonga. C'est un amour de la nature qui m'a poussé à travailler et à aider à protéger mon île natale en gérant les espèces envahissantes. »  
– Lisa Fenua, Tonga

## Comment puis-je organiser la publication des données ?

Nous vous invitons vivement à inclure la gestion des données dans le plan opérationnel de votre projet afin de maximiser la probabilité de mobiliser les données par la numérisation, la gestion et la publication en ligne. Grâce à votre plan, vous serez certain de recueillir tous les renseignements nécessaires à la publication au moment de l'échantillonnage et de tenir compte du temps à consacrer à la préparation, au nettoyage et à la publication des données. Le plan opérationnel doit comporter un échéancier pour le projet, les ressources qui y seront nécessaires ainsi que des explications sur le déroulement du projet. Demandez l'avis d'un expert lors de la rédaction de votre plan opérationnel.

Peut-être devrez-vous inclure dans les contrats de conseil et les accords d'entente (MOU) des exigences relatives à la production et à la gestion des données, si une partie de la gestion de vos espèces envahissantes se fait par l'entremise de consultants externes et/ou d'organisations partenaires.

Les modèles utilisés pour la collecte ou le traitement des données peuvent faciliter la normalisation des données en vue de leur publication ultérieure. Vous n'aurez pas à modifier tous vos systèmes de planification et de collecte de données, pour la préparation au GBIF. Cependant, si vous prévoyez de publier via le GBIF, vous devrez prendre certaines mesures de simplification, notamment l'association d'un identifiant de données d'occurrence à chaque événement lors de la collecte.

Bien que cela ne s'avère pas indispensable, il peut être utile de préparer vos ensembles de données dans un format normalisé répondant aux exigences en la matière, ou recourant à une terminologie commune, comme les termes Darwin Core, ce qui facilitera le processus de mise en correspondance par l'IPT vers le GBIF. Grâce aux termes Darwin Core, vous pouvez vous référer à des définitions cohérentes utilisées dans l'ensemble de la communauté de la biodiversité.

Un modèle de données est disponible ici : [www.sprep.org/attachments/Publications/IOE/Alien%26Invasive-Species-template.xls](http://www.sprep.org/attachments/Publications/IOE/Alien%26Invasive-Species-template.xls)

L'utilisation de termes standards revêt également une grande importance pour une gestion efficace des données. Un ensemble de définitions et de codes standard utiles pour la publication au GBIF est disponible ici : [www.sprep.org/attachments/Publications/IOE/All-lookups.xls](http://www.sprep.org/attachments/Publications/IOE/All-lookups.xls)

Vous disposez désormais des outils qui vous permettront de préparer, de publier et d'utiliser des données partagées sur la biodiversité. Nous espérons que ce guide vous encouragera à explorer les options disponibles du GBIF pour mieux défendre la diversité biologique du Pacifique.

## Glossaires de définitions

**Éditeur de données** – toute organisation / institution qui partage des données via le GBIF.

**Termes Darwin Core** – liste de termes ou de domaines d'intérêt en matière de biodiversité. Vous trouverez un guide de référence rapide pour les termes Darwin Core ici : <http://rs.tdwg.org/dwc/terms/>. Ce guide comprend des définitions et des exemples de chaque terme. Un document de 2012 décrivant la norme Darwin Core est disponible ici : <https://doi.org/10.1371/journal.pone.0029715>

**Integrated Publishing Toolkit (Boîte à outils d'édition intégrée, IPT)** – Logiciel développé par le GBIF pour la publication d'ensembles de données, hébergé sur un serveur local ou régional. Les Pôles participants ont souvent une ou plusieurs instances IPT pour les éditeurs de données. L'IPT que nous utilisons dans ce cas est celui du PROE.

**Métadonnées** – données sur les données, y compris le « qui, quoi, où et quand. » Par exemple, si la donnée que vous souhaitez publier est l'occurrence d'une espèce dans une certaine forêt en mai, vos métadonnées peuvent inclure qui vous êtes (nom de votre organisation, etc.), le but de la collecte de données, la méthodologie, la plage de dates d'échantillonnage et la portée géographique.

**Responsable de Pôle** – personne dirigeant un Pôle participant. Le responsable du Pôle est ou dirige le point de contact technique des éditeurs de données.

**OccurrenceID** – Un identificateur pour l'Occurrence (par opposition à un fichier numérique particulier de l'Occurrence). En l'absence d'identificateur unique global persistant, définissez-en un à partir d'une combinaison d'identificateurs dans le document qui rendra l'occurrenceID unique.

**Pôle participant** – une équipe de coordination (au sein d'un pays ou d'une organisation) qui travaille à l'établissement et au renforcement des activités liées au GBIF.

**TaxonID** – un identificateur pour l'ensemble des informations taxonomiques (données associées à la classe Taxon). Il peut s'agir d'un identificateur unique global ou d'un identificateur propre à l'ensemble de données.



## Pour de plus amples informations

La Battler Resource Base contient du matériel d'information et des ressources pour la lutte contre les espèces envahissantes : [www.sprep.org/piln/resource-base](http://www.sprep.org/piln/resource-base). Un modèle de base pour les ensembles de données et une liste de codes et de catégories communs à l'usage du GBIF sont disponibles dans la Battler Resource Base.

Vous pouvez contacter le Programme sur les espèces envahissantes sur le site Internet du PROE : [www.sprep.org/Invasive-Species/bem-invasive-species](http://www.sprep.org/Invasive-Species/bem-invasive-species) ou en envoyant un courriel à l'adresse [sprep@sprep.org](mailto:sprep@sprep.org).

## Support continu

**Support technique** : Responsable du Pôle Pacifique, PROE, Apia, Samoa : [gbifnodemanager@sprep.org](mailto:gbifnodemanager@sprep.org)

**Équipe du PROE pour les espèces envahissantes** : [sprep@sprep.org](mailto:sprep@sprep.org)

**Centre d'aide GBIF** : [helpdesk@gbif.org](mailto:helpdesk@gbif.org)

## Sites Internet

Page d'accueil GBIF	<a href="http://www.gbif.org">www.gbif.org</a>	Accéder, explorer et publier des données sur la biodiversité
Liste des données du projet sur les espèces envahissantes	<a href="http://www.gbif.org/dataset/search?project_id=bid-pa2016-0005-reg">www.gbif.org/dataset/search?project_id=bid-pa2016-0005-reg</a>	Section du GBIF contenant des données sur les espèces envahissantes de la région des îles du Pacifique.
Guide de référence rapide Darwin Core	<a href="http://rs.tdwg.org/dwc/terms/">http://rs.tdwg.org/dwc/terms/</a>	Page Web contenant les définitions des termes Darwin Core.
Darwin Core Hour	<a href="https://github.com/tdwg/dwc-qa/wiki/Webinars">https://github.com/tdwg/dwc-qa/wiki/Webinars</a>	Webinaires et présentations sur la norme Darwin Core
Renforcer la norme Darwin Core pour la recherche sur les espèces envahissantes	<a href="https://github.com/tdwg/dwc-qa/wiki/Webinars#chapter4">https://github.com/tdwg/dwc-qa/wiki/Webinars#chapter4</a>	Webinaire sur les projets potentiels portant sur les limites l'occurrenceStatus des termes Darwin Core et l'establishmentMeans lorsqu'il s'agit d'espèces envahissantes
Date Parser	<a href="http://data.canadensys.net/tools/dates">http://data.canadensys.net/tools/dates</a>	Outil permettant de déterminer les dates au format aaaa-mm-jjj à partir de divers autres formats
Boîte à outils d'édition intégrée (IPT)	<a href="http://www.gbif.org/ipt">www.gbif.org/ipt</a>	Description de la boîte à outils d'édition intégrée pour l'intégration des données dans le GBIF
Mode d'emploi de l'IPT	<a href="https://github.com/gbif/ipt/wiki/IPT2ManualNotes.wiki">https://github.com/gbif/ipt/wiki/IPT2ManualNotes.wiki</a>	Guide pratique sur l'utilisation de la Boîte à outils d'édition intégrée
Générateur d'identificateur unique universel (UUID)	<a href="http://www.uuidgenerator.net/">www.uuidgenerator.net/</a>	Outil de création d'identifiants uniques globaux UUID

<b>Droits d'auteur sur les données</b>	<a href="http://vertnet.org/resources/datalicensingguide.html">http://vertnet.org/resources/datalicensingguide.html</a>	Description du choix de la licence pour les données publiées
<b>Normes d'utilisation des données</b>	<a href="http://vertnet.org/resources/norms.html">http://vertnet.org/resources/norms.html</a>	Description du comportement attendu d'un bon utilisateur de données
<b>Open Refine Utilisation de base</b>	<a href="http://tinyurl.com/ORForBiod">http://tinyurl.com/ORForBiod</a>	Guide sur la façon d'effectuer le nettoyage des données communément nécessaires et relatives à la biodiversité
<b>Exporter PDF vers Excel</b>	<a href="https://acrobat.adobe.com/es/es/acrobat/how-to/pdf-to-excel-xlsx-convert.html">https://acrobat.adobe.com/es/es/acrobat/how-to/pdf-to-excel-xlsx-convert.html</a>	Si vous ne disposez pas d'une version complète d'Adobe pour ce faire, il existe un service payant en ligne pour le faire, avec un essai gratuit.
<b>Géoréférencement</b>		
<b>Bonnes pratiques en matière de géoréférencement</b>	<a href="http://www.gbif.org/document/80536/biogeomancer-guide-to-best-practices-in-georeferencing">www.gbif.org/document/80536/biogeomancer-guide-to-best-practices-in-georeferencing</a>	Description des méthodes et justification du géoréférencement
<b>Géoréférencement – Guide de Référence rapide</b>	<a href="http://manisnet.org/GeoreferencingQuickReferenceGuide.pdf">http://manisnet.org/GeoreferencingQuickReferenceGuide.pdf</a>	Description exacte de la façon de géoréférencer différents types de descriptions d'emplacement
<b>InfoXY</b>	<a href="http://splink.cria.org.br/infoxy?criaLANG=en">http://splink.cria.org.br/infoxy?criaLANG=en</a>	Outil de calcul de la géographie à partir de la decimalLatitude et de la decimalLongitude
<b>Convertisseur de coordonnées</b>	<a href="http://data.canadensys.net/tools/coordinates">http://data.canadensys.net/tools/coordinates</a>	Outil de calcul de la decimalLatitude et decimalLongitude à partir de la latitude et de la longitude dans d'autres formats
<b>Géoréférencement</b>	<a href="http://georeferencing.org/index.html">http://georeferencing.org/index.html</a>	Site contenant davantage d'informations sur le géoréférencement





# Join the Fight

Protect our islands from invasive species



Håfa Adâi

Aloha

Mogetin

Rahn Anim

Iokwe

Alii

Kaselehlie Len Wo

Ekawomir Omo

Mauri

Mālō te ma'uli

Halo

Tālofa nī

Halo

Tālofa

Halo

Tālofa

Ni sa Bula Fakaalofa lahi atu

Bonjour

Mālō e lelei

Kia Orana

Ia Orana  
Bonjour

Hello

Kia Ora

