

L'Utilisation des données sur la Biodiversité

Secrétariat du GBIF

Version b34f741, 2022-03-30 08:40:29 UTC

Table des matières

Description	1
Audience	1
Prérequis	2
Objectifs d'apprentissage	2
Certification	2
Traitement des données	3
Données adaptées au but d'utilisation	3
Problèmes et drapeaux	5
Outils de Traitement des Données	6
Gérer l'incertitude taxonomique	6
Mauvaise identification de l'espèce	7
Synonymie	7
Nouveaux noms	8
Gérer la qualité des données	8
Filtres et problèmes géospatiaux	8
Les centroïdes des pays	9
Points le long de l'équateur ou du premier méridien	9
Localisation incertaine	9
Jeux de données maillées	10
Relevés d'absence	11
Establishment Means	11
Basis of record	11
Anciens enregistrements	12
Doublons	12
Documents clés	12
API	13
Cloud Computing	13
Darwin Core	13
Publication de données	13
Publication de données: IPT	14
La numérisation	14
GBIF	14
Géoréférencement	15
Les espèces envahissantes	15
Atlas du vivant	15
Divers	15
OpenRefine	15
Planification/Collaboration	16

Red List Assessments	16
Qualité	17
R	17
Espèces sensibles	17
Taxonomie	18
Glossaire	18
Remerciements	21
Conception du cours et instruction	21
Traducteurs	21
Espagnol	21
Français	21
Ressources	21
Colophon	22
Suggestion de citation	22
Contributeurs	22
Licence	22
URI permanent	22
Contrôle du document	22
Image de couverture	22

Description

Ce cours adopte une approche modulaire pour la formation à l'utilisation des données du GBIF et s'appuie sur le cours Utilisation des données pour la prise de décision développé dans le cadre du programme Information sur la biodiversité pour le développement (BID). Différents cas d'utilisation ont été développés pour vous aider à démarrer sur les nombreuses façons d'utiliser des données en GBIF.

Ce cours est composé de contenu en ligne jumelé à des quiz et vous devez d'abord les compléter. Vous pourrez ensuite compléter le cas d'utilisation pertinent qui s'appuiera sur ce que vous avez appris grâce à un ensemble d'exercices pratiques. Lorsqu'il est offert en tant qu'atelier sur place ou virtuel, le travail de groupe et l'interaction sociale sont encouragés.

Le cours comprend un cours obligatoire sur le traitement des données, puis des modules optionnels que vous pouvez suivre en fonction du parcours qui vous intéresse le plus. Les deux premiers modules optionnels que nous avons développés sont :

- Modélisation de niche écologique
- Évaluation de l'état de conservation d'une espèce

Les sujets incluent:

- Traitement des données
 - Traitement d'un jeu de données GBIF et le rendre "adapté à son but d'utilisation"
- Modélisation de niche écologique
 - Introduction au fonctionnement et à l'interprétation d'un modèle écologique de base pour déterminer la distribution d'une espèce
 - Explorer de la modélisation de niche dans différentes conditions environnementales
- Évaluation de l'état de conservation d'une espèce
 - Utilisation de données gérées par le GBIF pour la création de cartes de distribution d'espèces en utilisant les standards de cartographie de l'UICN
 - Utiliser les données gérées par le GBIF pour évaluer l'état de conservation des espèces à l'aide des catégories et critères de la Liste Rouge de l'UICN

Audience

Ce cours s'adresse aux personnes qui travaillent en tant que chercheurs ou techniciens dans des institutions de recherche en biodiversité ou de politiques. Les instructions fournies sont particulièrement utiles pour ceux qui ont le besoin ou le désir d'utiliser des données gérées par le GBIF dans leurs propres recherches ou analyses.

Prérequis

- [Introduction au cours du GBIF](#)

De plus, pour tirer le meilleur parti des activités de ce cours, les participants devraient posséder les compétences et connaissances suivantes:

- Compétences de base en informatique et dans l'utilisation d'Internet et, en particulier, dans l'utilisation de feuilles de calcul.
- La compréhension de base des outils d'analyse géographique et statistique par ordinateur tels que SIG et R, et devrait avoir déjà mené des analyses à l'aide de ces outils.
- Connaissances de base sur la géographie et l'informatique sur la biodiversité: concepts de géographie et de cartographie, taxonomie et règles de nomenclature de base.
- La volonté de diffuser les connaissances acquises dans l'atelier avec des partenaires et des collaborateurs dans votre projet en adaptant les données sur la biodiversité utilisent des matériels de formation à des contextes et des langues spécifiques tout en maintenant leur valeur pédagogique.
- Une bonne maîtrise de l'anglais. Bien que des efforts soient faits pour fournir du matériel dans d'autres langues, les instructions / vidéos seront en anglais.

Objectifs d'apprentissage

- Accéder aux données du GBIF via une gamme de points d'accès
- Comprendre les problèmes courants de qualité des données dans les téléchargements GBIF qui peuvent affecter l'utilisation des données
- Appliquer des routines de traitement des données sur les téléchargements GBIF pour créer des jeux de données spécifiques
- Apprenez la différence entre les niches fondamentales et réalisées
- Expliquer comment délimiter une région d'entraînement
- Générer un modèle de niche simple
- Expliquer les résultats du modèle de niche
- Identifier les zones d'incertitude dans la projection
- Développer une stratégie de communication et des arguments convaincants pour l'intégration de la biodiversité dans les processus décisionnels
- Appliquer le critère B des catégories de l'UICN et des critères pour l'évaluation de l'état de conservation d'une espèce à l'aide d'un jeu de données adapté aux besoins
- Appliquer les protocoles de cartographie de l'UICN pour la production de cartes de distribution des espèces de la liste rouge

Certification

Les certificats sous la forme de badges numériques seront délivrés aux participants qui ont terminé

avec succès une affectation de cas d'utilisation. Il y a plusieurs cas d'utilisation et vous devriez compléter ceux qui sont les plus importants pour votre travail ou qui vous intéressent.

Unresolved directive in index.fr.adoc - include::downloads.fr.adoc[]

Traitement des données



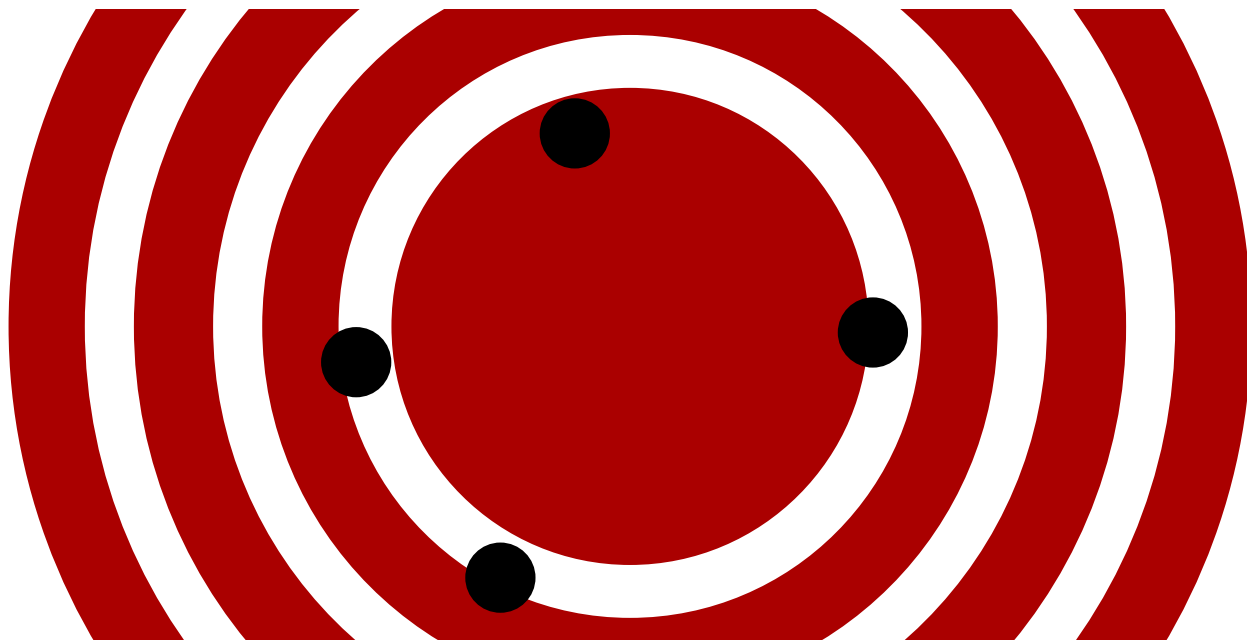
Dans le module de traitement des données, vous vous familiariserez avec le concept de jeux de données adaptés à vos besoins et avec certains filtres de données couramment utilisés que vous voudrez peut-être envisager pour créer votre propre jeu de données adapté à vos besoins.

Données adaptées au but d'utilisation

Presque toujours vous voudrez post-traiter votre téléchargement du GBIF d'une manière ou d'une autre pour correspondre à vos objectifs. Parfois, vous devrez faire des jugements difficiles pour votre cas particulier d'utilisation. Chaque fois que vous traitez avec des milliers ou des millions d'enregistrements, vous ne connaîtrez jamais la vraie qualité des données source. Il est important de garder à l'esprit que vous ne faites qu'atténuer les problèmes de qualité des données, et non les éliminer.

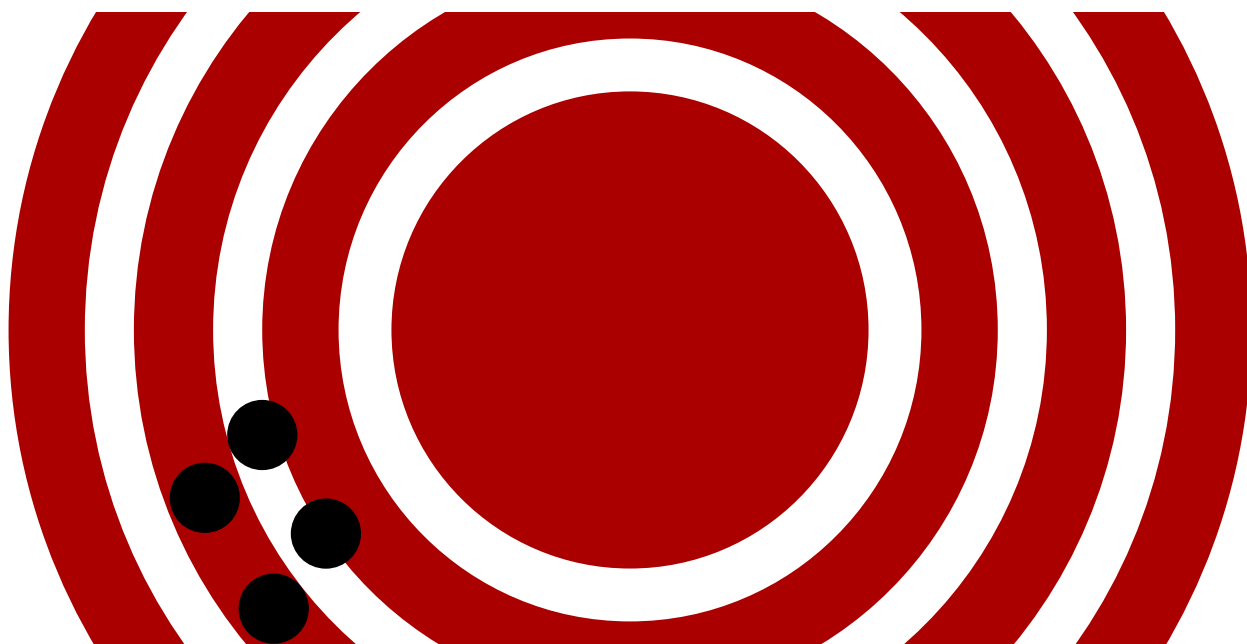
Les données que nous obtenons dans le téléchargement du GBIF, contiendront des données provenant d'une gamme de sources et les données varieront probablement en termes de justesse et de cohérence. La justesse et la cohérence sont deux façons de documenter les erreurs de données et sont des mesures de la qualité des données. Ce sont des mesures qui montrent à quel point le collecteur de données a été en mesure de saisir la véritable valeur étudiée. La nature du flux de publication des données du GBIF signifie que la justesse et la cohérence des données peuvent varier selon les éditeurs de données et la source des données. Si vous connaissez ces propriétés des données que vous avez, cela vous aidera à comprendre comment vous pouvez et ne pouvez pas nettoyer, valider et traiter les données.

- Justesse (Précision) - proximité des valeurs mesurées, des observations ou des estimations de la valeur réelle ou réelle e. . - les espèces ont été identifiées correctement ou la localité de collecte a été correctement identifiée.



Par exemple, si nous étudions la biogéographie végétale en Indonésie, et voulons faire une analyse spécifique pour une seule des îles de l'archipel, alors une question appropriée pourrait être - Les localités de l'île ont-ils été correctement géoréférencés?

- Cohérence (Précision) - niveau de résolution des données, par exemple précision des coordonnées, détermination taxonomique.

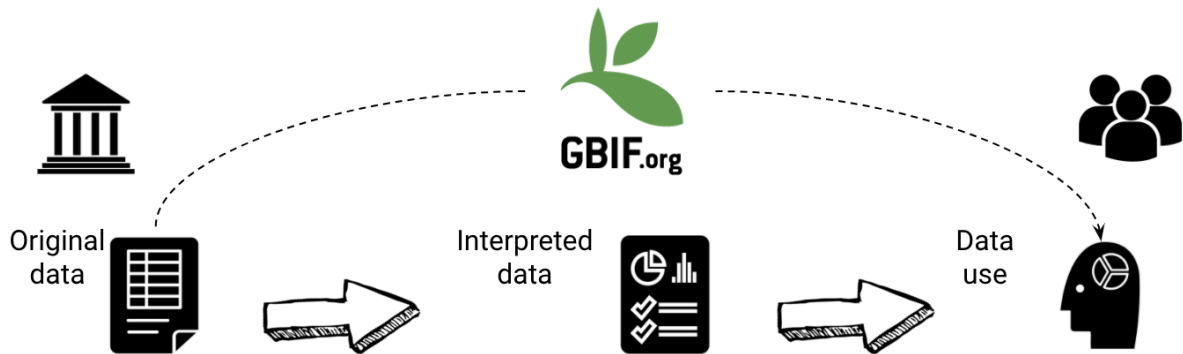


Dans l'exemple indonésien, une question appropriée pourrait être - L'incertitude dans l'estimation des coordonnées permet-elle que l'occurrence ne soit pas sur l'île ?

En règle générale, pour la plupart des analyses, vous voulez des données très correctes, mais le niveau de précision peut varier en fonction de votre analyse. Le GBIF peut vous aider à déterminer la justesse et la précision des données à travers, par exemple, des filtres et la signalisation de tickets, cependant, vous devez toujours vérifier à deux fois !

Problèmes et drapeaux

Le réseau GBIF publie des ensembles de données, les intégrant dans un système d'accès commun. Les utilisateurs peuvent y récupérer des données par le biais de services communs de recherche et de téléchargement. Au cours du processus d'indexation des données brutes, le GBIF ajoute des problèmes et des drapeaux aux enregistrements présentant des problèmes courants de qualité des données. Ceux-ci peuvent contenir des informations utiles pour vous, en tant qu'utilisateur, afin de créer des ensembles de données adaptés.



Des remarques sont indiquées sur les pages individuelles des occurrences pour expliquer le processus effectué après l'interprétation:

*Exclues signifie que les données originales n'ont pas pu être interprétées, donc elles sont exclues des champs interprétés.

*Modifiée signifie que les données originales ont été modifiées lors du processus d'interprétation afin d'être indexée sur GBIF.org.

*Inferred means the Using other record information the data indexed is inferred, if the original is empty.

L'interface de recherche ne permet pas actuellement d'exclure tous les enregistrements comportant un problème particulier. Il est possible de filtrer tous les enregistrements qui ne vous intéressent pas et qui comportent des problèmes en sélectionnant le problème en question et en cliquant sur le bouton d'inversion. Cependant, l'inversion ne vous donnera que toutes les autres occurrences marquées et non les enregistrements sans problème. Le GBIF s'efforce d'améliorer ce point. (sur la recherche d'occurrences)

The screenshot displays the GBIF Occurrences search interface. On the left, a sidebar lists various filters with their respective counts. The filter 'Basis of record invalid' is selected and highlighted with a red box, showing a count of 12,214,242. Below the list are 'CLEAR' and 'REVERSE' buttons, also highlighted with red boxes. The main area shows a table of search results with columns for Scientific name, Country or area, Coordinates, and Month & year. The search results are filtered to show only records with 'Basis of record invalid'.

Scientific name	Country or area	Coordinates	Month & year
Flabelligena hakuhoae Jimi			2020 January
Flabelligena hakuhoae Jimi			2020 January
<i>Monacha cartusiana</i> (O.F.Müller, 1774)		47.8N, 13.0E	2020 January
<i>Physcia aipolia</i> (Ehrh. ex Humb.) Fümnr.	Austria	47.0N, 12.4E	2020 January
<i>Abida secale</i> (Draparnaud, 1801)		47.8N, 13.0E	2020 January
<i>Cochlostoma septemspirale</i> (Razoumowsky...)		47.8N, 13.0E	2020 January
<i>Abies alba</i> Mill.		47.6N, 13.2E	2020 January
<i>Rubus idaeus</i> L.		47.6N, 13.2E	2020 January
<i>Polystichum braunii</i> (Spenn.) Fée		47.6N, 13.2E	2020 January

Un aperçu complet de tous les problèmes et drapeaux peut être trouvé ici: <https://data-blog.gbif.org/post/issues-and-flags/>

Outils de Traitement des Données

Bien que les filtres GBIF permettent un certain traitement des données, par exemple de sélectionner uniquement les données qui répondent à certains critères pour les télécharger, il est fortement recommandé de procéder à un traitement supplémentaire des données et une gamme d'outils est disponible. Le choix de l'outil que vous utiliserez sera basé sur l'accès personnel, la familiarité et l'utilité de ces derniers. Les outils qui peuvent être utilisés à cette fin sont les suivants:

- Logiciel d'édition de feuilles de calcul par exemple Excel, Google Sheets (petits jeux de données)
- OpenRefine
- Paquetages et scripts R, par exemple rgbif, CoordinateCleaner, scrubr et biogeo (traitement automatisé des données)
- Systèmes d'information géographique (SIG), par exemple ArcGIS, QGIS et MapInfo

Il est toujours important d'inclure une étape de visualisation des données dans votre traitement des données afin de pouvoir identifier les points de données anormaux qui ont pu être manqués pendant l'étape de traitement. Pour ceux qui traitent leurs données avec R, cette étape peut être réalisée dans R, mais avec OpenRefine ou un logiciel d'édition de feuilles de calcul, vous devrez peut-être utiliser un logiciel SIG ou alors des outils tels que Google Earth pour la visualisation des données.

Gérer l'incertitude taxonomique

L'incertitude entourant la taxonomie d'une donnée peut survenir pour plusieurs raisons:

- Mauvaise identification de l'espèce

- Synonymie
- Nouveaux noms

Mauvaise identification de l'espèce

L'identification des espèces est un processus complexe, avec des espèces typiquement décrites à partir d'un certain ensemble de caractères identifiés dans une description publiée et liée à un spécimen-type qui appartient à une collection scientifique et qui sera utilisée pour valider l'identification des espèces. Lorsque les taxons sont très similaires ou qu'un ensemble de traits complexes sont requis pour une identification correcte, une expertise taxonomique spécifique que les éditeurs de données ne possèdent pas peut être requise, menant à une identification erronée d'une espèce. En tant qu'utilisateurs, vous devez avoir une compréhension claire de la façon dont les décisions taxinomiques pour votre groupe d'intérêt sont faites:

- Quels sont les caractères utilisés pour définir les espèces ?
- Ces caractères sont-ils facilement confondus ou capturés lorsque l'espèce est observée ou collectée?
- Y a-t-il des espèces connexes qui pourraient facilement être confondues avec les espèces qui vous intéressent?

Si vous pensez qu'il y a un risque que l'espèce ne soit pas identifiée de manière correcte, vous pouvez adopter une approche prudente avec les données que vous utilisez, et utiliser uniquement les données liées à des spécimens dans des collections où la validation taxonomique serait possible et en éliminant d'autres sources de données. Une autre approche peut être d'utiliser des données associées telles que des informations sur le collecteur, les médias, les séquences d'ADN, etc. pour valider la détermination taxonomique.

Synonymie

La synonymie peut survenir lorsque la même espèce a été décrite plusieurs fois et qu'un nouveau nom est donné à l'espèce chaque fois qu'elle est décrite, ou, lorsqu'il y a un changement dans la taxonomie d'une espèce, par exemple, une espèce est déplacée d'un genre à l'autre. Un seul nom d'espèce peut être accepté, et d'autres noms sont ce que nous appelons des synonymes. Ces synonymes peuvent toujours être utilisés dans une mesure variable et vous devez être sûr que lorsque vous récoltez des données depuis GBIF d'obtenir des données pour le nom taxonomique dont vous avez besoin. La colonne vertébrale taxonomique du GBIF différencie les noms scientifiques et les synonymes acceptés, et les identifiants uniques sous la forme de clés de taxon. Les recherches sur les espèces <https://www.gbif.org/species/search> permettent de filtrer les noms acceptés, les synonymes et les clés de taxon qui peuvent être utilisés pour les recherches programmatiques sur GBIF.

Taxon Keys. Les noms scientifiques peuvent être désordonnés. Si vous accédez aux données du GBIF via un programme plutôt que par le site Web, les clés taxonomiques ou taxon keys constituent un moyen efficace de définir des recherches basées sur la taxonomie. Les clés taxonomiques sont émises au niveau de l'espèce, du genre, de la famille, de l'ordre, du phylum et du royaume. Des identifiants uniques sont attribués aux noms acceptés et les synonymes de ces noms acceptés

reçoivent le même identifiant. Il peut donc être utile de trier les espèces en fonction de leurs clés taxonomiques uniques fournies lors de l'indexation du jeu de données par le GBIF.

Nouveaux noms

Il peut arriver que le nom scientifique ne corresponde à aucun nom de l'ossature taxonomique du GBIF, peut-être parce que l'espèce est nouvellement décrite ou qu'elle ne figure pas dans une liste de contrôle utilisée par le GBIF pour construire son ossature. Ces noms sont signalés par l'indicateur TAXON_MATCH_HIGHERRANK qui indique que le nom scientifique n'a pas été reconnu mais que la donnée a été appariée à un niveau taxonomique supérieur, par exemple le genre ou la famille. Cet indicateur peut être utilisé pour identifier et filtrer ces données. Lorsque les noms ont été mal orthographiés ou mal formatés, il existe également un indicateur TAXON_MATCH_FUZZY qui peut être utilisé pour identifier et filtrer les noms qui ne peuvent correspondre à l'ossature taxonomique qu'en utilisant une correspondance floue et non exacte.

Gérer la qualité des données

Le filtrage des données vous permet, en tant qu'utilisateur, d'obtenir les données les mieux adaptées à votre objectif. Toutes les recherches disposent d'un ensemble de filtres qui peuvent être utilisés pour trouver les données dont vous avez besoin, et les recherches sur les occurrences disposent d'un ensemble de filtres de recherche "avancés" supplémentaires pour les utilisateurs qui ont besoin d'effectuer un filtrage plus poussé. Bien que les filtres puissent vous permettre d'éliminer les données qui ne sont pas pertinentes ou qui sont de qualité insuffisante pour vos besoins, un filtrage supplémentaire peut être nécessaire, soit manuellement, soit par programmation, pour traiter les problèmes de qualité des données qui surviennent dans le modèle de publication des données du GBIF. Vous trouverez ci-dessous quelques filtres de données courants que vous pouvez envisager en tant qu'utilisateur pour rendre les données plus adaptées à vos besoins.

Filtres et problèmes géospatiaux

Les données peuvent être filtrées spatialement dans une recherche d'occurrences de l'une des 3 façons suivantes:

- Pays ou zone/Continent - les données sont filtrées par pays et incluront des données dans la Zone Economique Exclusive (EEZ)
- Zone administrative - ce filtre utilise la base de données GADM <https://gadm.org/data.tml> des zones administratives de tous les pays du monde pour permettre au GBIF de supprimer les problèmes géospatiaux communs par défaut si vous choisissez d'avoir des données avec une localisation.
- Localisation - ce filtre vous permet de filtrer les données avec des coordonnées et/ou de dessiner vos propres polygones ou d'utiliser un fichier GeoJSON pour délimiter votre propre buffer. Si vous filtrez pour ces données avec des coordonnées, un certain nombre de problèmes géospatiaux associés à la publication des données seront éliminés. Parmi lesquels:
 - Zéro Coordonnées - Les coordonnées sont exactement (0,0) ou ce qui est parfois appelé "l'île nulle". La coordonnée zéro est un problème géospatiale très courant. Le GBIF supprime (0,0)

lorsque 'hasgeospatialissue' est paramétré sur FALSE.

- Inadéquation des coordonnées du pays - Les éditeurs de données fourniront souvent au GBIF un code de pays (US,TW,SE,JP...). Le GBIF utilise le système de codage ISO 3166-1 alpha-2 à deux lettres - https://fr.wikipedia.org/wiki/ISO_3166-1_alpha-2. Lorsqu'un point ne fait pas partie du polygone ou de la ZEE, mais qu'il indique se situer dans le pays, il est signalé comme ayant «country coordinate mis-match» et sera supprimé si les données sont filtrées pour leur localisation.
- Coordonnée invalide - Si le GBIF est incapable d'interpréter les coordonnées.
- Coordonnée invalide - Les coordonnées sont en dehors de l'intervalle pour les valeurs possibles de lat/lon décimales (-90,90), (-180,180).

Les centroïdes des pays

Les centroïdes de pays sont ceux pour lesquels l'observation est épinglée au centre du pays à la place du lieu où le taxon a été observé ou enregistré. Les centroïdes de pays sont généralement des enregistrements auxquels on a attribué rétrospectivement une valeur lat-long basée sur une description textuelle de l'emplacement de l'enregistrement original. Les logiciels de géocodage utilisent des répertoires géographiques, des dictionnaires ou des annuaires utilisés en conjonction avec une carte ou un atlas, pour attribuer des coordonnées aux noms de lieux. Ainsi, si la notice indique simplement "Brésil", certains éditeurs placeront la notice au centre du Brésil. De même, si la notice indique simplement "Texas" ou "Paris", elle sera placée au centre de ces régions. Il s'agit d'une caractéristique presque exclusive aux données de musée (PRESERVED_SPECIMEN), mais cela peut également se produire avec d'autres types d'enregistrements.

L'identification des données dérivées de centroïdes de pays n'est pas possible actuellement avec les filtres GBIF, cependant, le programme R CoordinateCleaner peut être utilisé pour identifier et filtrer les centroïdes de pays.

Points le long de l'équateur ou du premier méridien

Certains éditeurs considèrent que zéro et NULL sont équivalents ainsi les champs de latitude et de longitude vides pour un enregistrement reçoivent une valeur zéro. En conséquence, les records finissent par être tracés le long de l'équateur et des premières lignes méridiennes.

Localisation incertaine

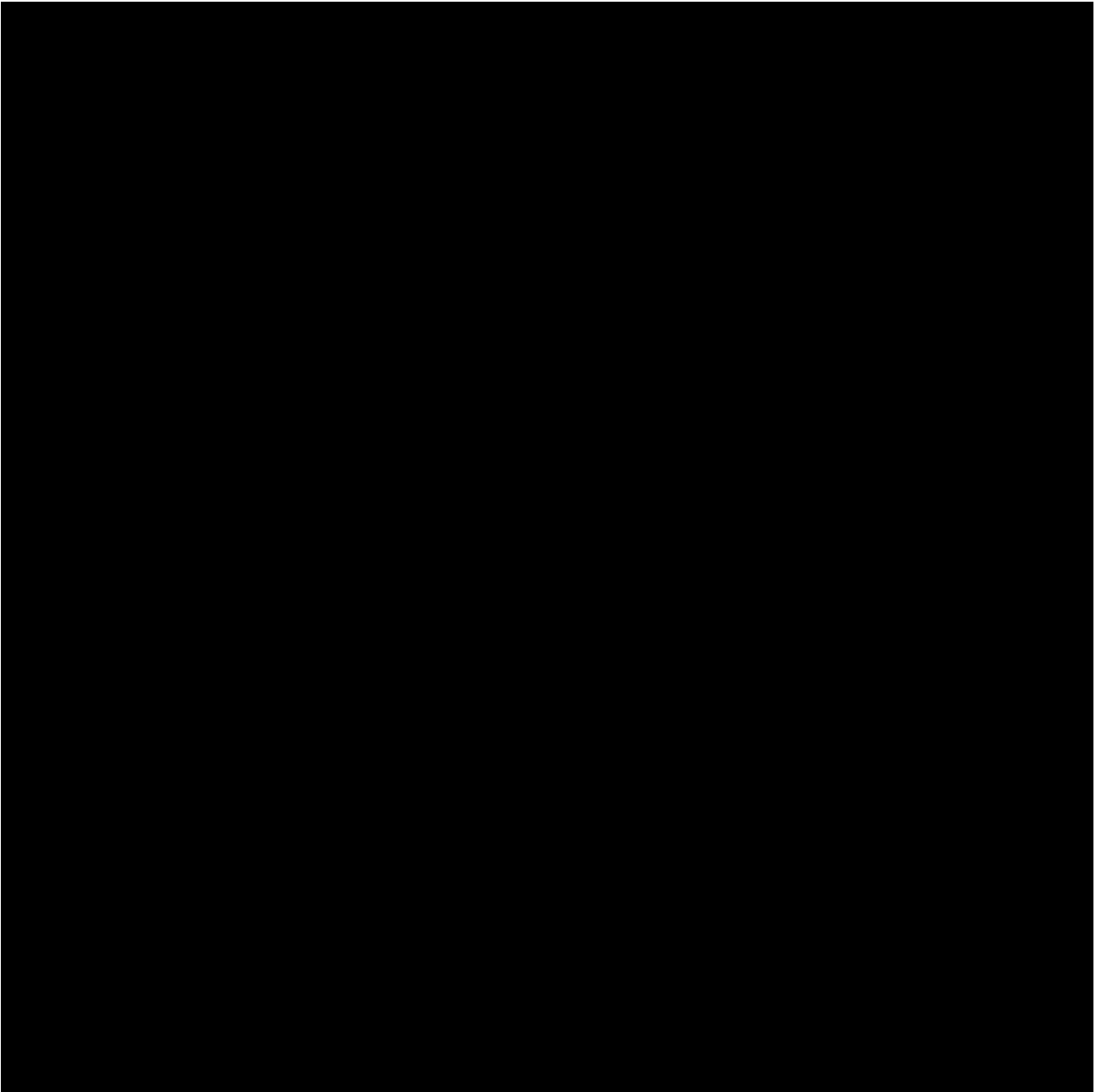
Souvent, vous voudrez être sûr que les coordonnées donnent un certain emplacement et ne sont pas réellement à des milliers de kilomètres de l'endroit où l'organisme a été observé ou collecté. Il existe deux champs - coordinate precision et coordinateUncertaintyInMeters - dans Darwin Core que vous obtenez avec un téléchargement CSV SIMPLE. Vous pouvez les utiliser pour filtrer par "incertitude". Cependant, ces champs ne sont pas très souvent utilisés par les éditeurs qui estiment que leurs enregistrements sont assez certains (à partir d'un GPS) et nous recommandons de ne pas écarter les valeurs manquantes.

Il existe également quelques valeurs "fausses" pour l'incertitude des coordonnées que vous devez connaître. Ces valeurs sont des erreurs produites par les logiciels de géocodage et ne représentent pas de véritables valeurs d'incertitude. Ces "fausses" valeurs sont 301, 3036, 999 et 9999. Dans le cas

de la valeur 301, l'incertitude est souvent beaucoup plus grande que 301 et représente en fait un centroïde de pays.

Jeux de données maillées

Les jeux de données maillées sont un problème connu sur GBIF. De nombreux jeux de données ont des points espacés de manière égale . Ces jeux de données sont généralement des enquêtes nationales systématiques ou des données obtenues à partir de certains atlas (« schémas de collecte matriciels »). Les occurrences géoréférencées sont positionnées sur un point central



La plupart des éditeurs de jeux de données maillées remplissent en fait l'une des colonnes suivantes : `coordinateuncertaintyinmeters`, `coordinateprecision`, `footprintwkt`. Ainsi, filtrer ces colonnes peut être une bonne manière de retirer les jeux de données maillées. Le package R `Coordinate cleaner` a également une fonction pour supprimer les jeux de données maillées. Le GBIF a une API expérimentale pour identifier les jeux de données qui exposent un certain niveau de "maillage". Vous pouvez en lire plus ici: <https://data-blog.gbif.org/post/finding-gridded-datasets/>

Relevés d'absence

Par défaut, les relevés de présence et d'absence sont affichés lorsque vous faites une recherche sur www.gbif.org. Les relevés d'absence confirment qu'une espèce n'a pas été trouvée dans une localité spécifique qui a été échantillonnée et que cette information peut être utile, par exemple, pour le développement de modèles de niche écologique. Cependant, vous pourriez n'être intéressé que par les enregistrements de présence et dans ce but vous pouvez filtrer pour garder seulement les enregistrements de présence en utilisant le filtre statut d'occurrence.

Establishment Means

Le terme Darwin Core establishmentMeans identifie le processus par lequel le ou les individus biologiques représentés dans les occurrences se sont établis à cet endroit. En tant que tel, il peut servir d'outil de filtrage utile pour identifier les enregistrements qui se trouvent en dehors de l'aire de répartition d'une espèce, les termes acceptés pour ce champ étant natif, natif réintroduit, introduit, introduit par colonisation assistée, vagabond et incertain. Actuellement, les enregistrements du GBIF peuvent être recherchés à l'aide des anciens termes de vocabulaire natif, introduit, naturalisé, indigène, géré et incertain - https://rs.gbif.org/vocabulary/gbif/establishment_means.xml, et ceux-ci seront mis à jour à la fin de 2022. Dans certains cas, la suppression des enregistrements "MANAGED" entraînera la suppression des enregistrements de zoo.

Utilisez ce filtre avec prudence, car la plupart des enregistrements ne contiennent pas cette information et seraient donc exclus d'une recherche avec ce filtre. Nous vous recommandons d'utiliser l'information dans Establishment Means pour le filtrage après le téléchargement.

Basis of record

Basis of record est un terme Darwin Core qui fait référence à la nature spécifique de l'enregistrement et peut se référer à l'une des 6 classes:

- Living Specimen - un spécimen vivant, par exemple, une plante vivante dans un jardin botanique ou un animal vivant dans un zoo.
- Preserved Specimen - un spécimen qui a été préservé, par exemple, une plante sur un planche d'herbarium ou une quantité cataloguée de poissons dans un pot.
- Fossil Specimen - un spécimen préservé qui est un fossile, par exemple, un fossile corporel, un coprolite, un gastrolithe, un ichnofossile ou un morceau d'arbre pétrifié.
- Material Citation - Une référence à, ou une citation d'une partie, d'un, ou de plusieurs spécimens dans des publications scientifiques, par exemple, une citation d'un spécimen physique dans une collection scientifique, un traitement taxonomique dans une publication scientifique, ou une occurrence mentionnée dans un carnet de terrain.
- Human Observation - le résultat d'un processus d'observation humaine, par exemple, la preuve d'une occurrence prise à partir de notes de terrain ou de la littérature, ou encore un enregistrement d'une occurrence sans preuve physique ni preuve capturée par une machine.
- Machine Observation - Une sortie d'un processus d'observation de machine, par exemple, une photographie, une vidéo, un enregistrement audio, une image à distance ou un enregistrement

d'occurrence basé sur la télémétrie.

Basis of record devrait permettre aux utilisateurs de filtrer ces individus dans les collections ex situ telles que les zoos, les jardins botaniques ou les fossiles, ainsi que de filtrer ces enregistrements en fonction de la base ou de l'observation, qui peut supporter la validation taxonomique. Vous devriez noter que, même si cela peut être un filtre utile, les éditeurs de données ne remplissent pas toujours correctement le champ Basis of record, ou alors, il peut y avoir des nuances dans les données qui peuvent ne pas être immédiatement évidentes pour un utilisateur. . <https://data-blog.gbif.org/post/living-specimen-to-preserved-specimen-understanding-basis-of-record/> et vous devriez toujours vérifier vos données avant de les utiliser.

Anciens enregistrements

Le GBIF possède de nombreux enregistrements de musées qui pourraient être plus anciens que ce qui est souhaité pour certaines études.

Doublons

La duplication des enregistrements peut se produire lorsque plusieurs enregistrements d'un même individu sont effectués. Cela peut se produire, par exemple, lorsqu'un chercheur dépose plusieurs spécimens d'un même arbre dans des herbiers du monde entier qui publient ensuite ces données sur le GBIF, ou lorsqu'un individu a été déposé dans une collection d'histoire naturelle et que cet individu a également été échantillonné pour son ADN. Dans ce cas, il y aura un enregistrement pour le spécimen dans les collections et un autre pour la séquence d'ADN.

Le GBIF a récemment introduit une fonction de regroupement dans sa recherche avancée qui permet aux utilisateurs d'identifier des groupes d'enregistrements, c'est-à-dire des enregistrements qui semblent provenir de la même source. Cela permet aux utilisateurs d'identifier les données potentiellement dupliquées et de les filtrer lors du téléchargement. Notez que si vous filtrez les enregistrements qui sont dans un groupe, vous perdrez tous les enregistrements trouvés dans ce groupe et vous perdrez des données potentiellement utiles. Le filtre peut être mieux utilisé pour indiquer l'étendue de la duplication dans l'ensemble de données, ou pour des téléchargements indépendants des ensembles de données groupées et non groupées pour comparaison.

Unresolved directive in index.fr.adoc - include::R_Data_Processing_Pipeline.fr.adoc[]

Unresolved directive in index.fr.adoc - include::Ecological-Niche-Modelling.fr.adoc[]

Unresolved directive in index.fr.adoc - include::Use_Case_ENM.fr.adoc[]

Unresolved directive in index.fr.adoc - include::Red-List-Assessments.fr.adoc[]

Unresolved directive in index.fr.adoc - include::Use_Case_Conservation_Status.fr.adoc[]

Documents clés



Les références suivantes fournissent plus de détails sur les sujets abordés dans ce cours. Tous les liens s'ouvrent dans une nouvelle fenêtre/onglet.

API

- [Guide débutant de l'API GBIF](#)
- [\(-Presque\) tout ce que vous voulez savoir sur l'API des Espèces du GBIF](#)

Cloud Computing

- [GBIF et Apache-Spark sur le tutoriel Microsoft Azure](#)
- [GBIF et Apache-Spark sur le tutoriel AWS](#)

Darwin Core

- [Termes Darwin Core : Un guide de référence rapide](#)
- [DarwinCore Simple](#)
- [Questions & Réponses sur le Darwin Core](#)
- [Extensions Darwin Core enregistrées sur GBIF](#)

Publication de données

- [Guide rapide pour publier des données via GBIF](#)
- [Comment publier des données sur la biodiversité via GBIF.org](#)
- <https://www.gbif.org/devenez> un éditeur de données[Devenir un éditeur de données avec GBIF^]
- [Bonnes pratiques pour publier des données sur la biodiversité à partir des études d'impact environnemental + Secrétariat du GBIF et IAIA : Association internationale pour l'évaluation des impacts \(2020\).](#)
- [Guide à destination des entreprises privées pour devenir éditeurs de données via le GBIF : modèle de document pour appuyer le processus de validation interne afin de devenir un éditeur de données GBIF + Rui Figueira, Pedro Beja, Cristina Villaverde, Miguel Vega, Katia Cezón, Tainan Messina, Anne-Sophie Archambeau, Rukaya Johaadien, Dag Endresen & Dairo Escobar \(2020\).](#)
- [Publier des données dérivées de séquences génétiques via des plateformes de données sur la biodiversité + Anders F. Andersson, Andrew Bissett, Anders G. Finstad, Frode Fossøy, Marie Grosjean, Michael Hope, Thomas S. Jeppesen, Urmas Kõljalg, Daniel Lundin, R. Henrik Nilsson, Maria Prager, Cecilie Svenningsen & Dmitry Schigel \(2020\).](#)
- [Les différents types de jeu de données supportés par GBIF](#)
- [Pré-requis GBIF sur la qualité des données pour la publication](#)

- Licences de données GBIF
- Fichiers modèles pour les listes taxonomiques
- Fichiers modèles pour les listes d'occurrences
- Fichiers modèles pour les événements d'échantillonnage
- Bonnes pratiques pour les événements d'échantillonnage
- Partager des images, sons et vidéos sur GBIF
- Data papers
- Data papers publiés

Publication de données: IPT

- The GBIF Integrated Publishing Toolkit: Facilitating the Efficient Publishing of Biodiversity Data on the Internet + Robertson et al. (2014)
- Installer l'IPT ou non
- centres d'hébergement de données IPT
- Installation de l'IPT / Configuration du webinaire
- Installation de l' IPT en vidéo
- Demonstration pratique d'utilisation de l'IPT en vidéo

La numérisation

- Ressources de numérisation iDigBio
- Flux de travail de numérisation des collections iDigBio
- Workflows et Protocoles de numérisation iDigBio
- guide de capture d'images iDigBio
- Guide en 10 étapes du Canadensys pour la gestion des images avec vos données sur la biodiversité

GBIF

- Qu'est-ce que le GBIF
- Plan stratégique
- Devenir membre
- Revue Scientifique
- Établir un point nodal effectif du GBIF : Concepts et considérations générales + Secrétariat du GBIF (2019).

Géoréférencement

- [Meilleures pratiques de Géoréférencement](#) + Arthur D. Chapman & John R. Wieczorek (2020).
- [Guide de référence rapide du Géoréférencement](#) + Paula F. Zermoglio, Arthur D. Chapman, John R. Wieczorek, Maria Celeste Luna & David A. Bloom (2020).
- [Manuel du calculateur de géoréférencement](#) + David A. Bloom, John R. Wieczorek & Paula F. Zermoglio (2020).
- [Ressources de Géoréférencement](#)

Les espèces envahissantes

- [GRISS - Registre mondial des espèces introduites et envahissantes](#)
- [TrIAS - Suivi des espèces exotiques envahissantes](#)

Atlas du vivant

- [Atlas du vivant](#)
- [Documentation technique clé ALA](#)

Divers

- [Guide VertNet pour ouvrir des fichiers texte dans Excel](#)
- [Guide de licence de données VertNet](#)

OpenRefine

- [Documentation d'OpenRefine](#)
- [Les expressions régulières d'OpenRefine](#)
- [Guide de nettoyage des données de biodiversité avec OpenRefine](#) + Paula F. Zermoglio, Camila A. Plata Corredor, John R. Wieczorek, Ricardo Ortiz Gallego & Leonardo Buitrago (2021).
- [Utiliser Google Refine et des bases de données taxonomiques \(EOL, NCBI, uBio, WORMS\) pour nettoyer les données désordonnées](#) + iPhylo blog post. Rod Page 2012.
- [Réconcilier les noms des auteurs en utilisant Open Refine et VIAF](#) + article de blog iPhylo. Rod Page 2013.
- [Valider les noms scientifiques avec l'API du portail GBIF](#) + Le message d'invité a été rédigé par Gaurav Vaidya, Victoria Tersigni et Robert Guralnick 2013.
- [Nettoyage des données iDigBio avec OpenRefine](#) + iDigBio 2013.
- [Avons-nous les noms « Corrects »?](#) + Canadensys 2014.
- [Nettoyage des données avec OpenRefine](#) + Desmet et Brosens 2016 TDWG.

- [Liste rouge EasyOpen](#) + Interrogation de la liste rouge de l’UICN, en utilisant une liste d’espèces, OpenRefine, et un code pré-écrit. Olly Griffin Juillet 2019.

Planification/Collaboration

- [Agile](#)
- [Qu’est-ce que SCRUM](#)
- [Cadre SCRUM](#)
- [méthodologie Kanban](#)
- [Guide Scrum](#)
- [GitHub](#)

Red List Assessments

- **IUCN Red List Categories and Criteria**
 - [IUCN Red List Categories and Criteria v3.1 - ENGLISH](#)
 - [Categorías y Criterios de la Lista Roja de la UICN v3.1 - ESPAÑOL](#)
 - [Catégories et Critères de la Liste Rouge de l’UICN v3.1 - FRANÇAIS](#)
- **Red List Guidelines**
 - [Guidelines for Using the Red List Categories and Criteria version 15 - ENGLISH](#)
 - [Directrices de uso de las Categorías y Criterios de la Lista Roja de la UICN Versión 14 - ESPAÑOL](#)
 - [Lignes directrices pour l’utilisation des Catégories et Critères de la Liste rouge de l’UICN Version 14 - Français](#)
- **Criteria Summary Sheet**
 - [Criteria summary sheet - ENGLISH](#)
 - [Resumen de los Criterios - ESPAÑOL](#)
 - [Résumé des Critères - FRANÇAIS](#)
- **Mapping Standards**
 - [Mapping Standards - ENGLISH](#)
 - [Standard Attributes for Spatial Data - ENGLISH](#)
 - [Attribute codes for Presence](#)
- **Regional and National Levels Guidelines**
 - [Guidelines for application of IUCN Red List Criteria at Regional and National Levels Version 4 - ENGLISH](#)
 - [Directrices para el uso de los Criterios de la Lista Roja de la UICN a nivel regional y nacional Versión 4 - ESPAÑOL](#)

◦ [Lignes directrices pour l'application des Critères de la Liste rouge de l'IUCN aux niveaux régional et national Version 4. - FRANÇAIS](#)

- [Spatial Tools and Data for Red List Assessments](#) + Includes brief overviews of some tools developed to aid mapping of spatial data and estimation of Red List metrics such as EOO and AOO, plus developer information, links to help files, support networks and associated research publications.
- [Supporting information guidelines](#) + The Documentation Standards and Consistency Checks for IUCN Red List Assessments and Species Accounts (also known as the Supporting Information Guidelines) provides guidance on the required and recommended supporting information for Red List assessments. It also provides guidance on the writing style and format that should be used for all IUCN Red List assessments.

Qualité

- [Principes de la qualité des données](#) + Arthur Chapman 2005.
- [Principes et méthodes de nettoyage des données: Espèces primaires et Données d'Occurrence des espèces](#) + Arthur Chapman 2005.
- [Soyez prudent avec les dates dans Excel](#) + DataOne 2014.
- [Encodage de caractères pour les débutants](#)
- [Guide MVZ d'enregistrement des localités dans les notes de terrain](#)

R

- [Data Carpentry-Data Analysis and Visualization in R for Ecologists](#)
- [Datacamp- Range of courses in R, Python and SQL](#)
- [Manuel d'introduction à R](#)
- [Manuel rgbif](#)
- [Manuel de CoordinateCleaner](#)
- [Télécharger les occurrences d'une longue liste d'espèces avec R et Python](#)
- [Des choses communes à surveiller lors du post-traitement des téléchargements GBIF](#)
- [Trouver des jeux de données dérivés de grilles & Mise à jour des jeux de données dérivés de grilles](#)
- [Centroides de Pays](#)
- [Utiliser des shapefiles sur les données GBIF avec R](#)
- [téléchargement sans oiseaux](#)

Espèces sensibles

- [Meilleures pratiques actuelles pour la généralisation des données d'occurrence des espèces](#)

Taxonomie

- listes d'espèces du GBIF et déficits de données
- Laboratoires GBIF - Analyseur de noms
- Laboratoires GBIF - Correspondance d'espèces
- Résolveur Global des Noms
- Stratégie de correspondance de nom d'espèces marines pour le contrôle de qualité taxonomique
- Nomenmatch

Glossaire

ALA

Atlas of Living Australia. Le nœud australien du GBIF, qui a développé un portail de données open source maintenant largement utilisé au sein de la communauté GBIF et des partenaires pour les portails nationaux sur la biodiversité.

API

Interface de programmation d'applications. Un ensemble de méthodes de communication clairement définies entre différents composants logiciels.

BID

Information sur la biodiversité pour le développement. Un projet financé par l'UE et coordonné par le GBIF, dont le but est d'accroître la capacité de mobilisation des données dans les régions d'Afrique, des Caraïbes et du Pacifique.

BIFA

Fonds pour la biodiversité en Asie.

Licences CC

Creative Commons. Il s'agit d'une série de licences mises en place par l'organisation Creative Commons qui permettent le partage et la réutilisation de la créativité et du savoir par la fourniture d'outils juridiques libres. Trois d'entre eux peuvent être assignés à des jeux de données partagés sur GBIF : CC0, CC BY et CC BY-NC.

Vocabulaire contrôlé

Il s'agit d'un ensemble restreint de termes qui sont utilisés comme valeurs possibles pour un champ donné. On peut voir cela comme une liste de recherche ou une liste déroulante pour un champ particulier. Par exemple, le champ DwC basisOfRecord ne devrait contenir qu'une de ces valeurs : "PreservedSpecimen", "FossilSpecimen", "LivingSpecimen", "HumanObservation", "MachineObservation". Nous dirons que la liste des valeurs est un vocabulaire contrôlé pour ce champ.

DwC

Darwin Core est un standard de données sur la biodiversité, maintenu par le TDWG et largement utilisé au sein de la communauté et des partenaires du GBIF. Il s'agit d'un ensemble de termes normalisés (ou noms de champs) et de leurs définitions, qui sont utilisés pour partager des informations sur la biodiversité.

DOI

Identificateur d'objet numérique. Un identifiant ou une clef persistante utilisé pour identifier de manière unique les objets. Les DOI sont principalement utilisés pour identifier les informations académiques, professionnelles et gouvernementales, telles que les articles de revues, les rapports de recherche et les ensembles de données, ainsi que les publications officielles.

DwC-A

Archive Darwin Core. Un fichier compressé (zippé) contenant toutes les informations nécessaires pour partager sur GBIF, pour une ressource particulière. Chaque zip contient trois types de fichiers:

1. les données elles-mêmes, dans un ou plusieurs fichiers texte: occurrence.txt/event.txt/measurmentoffact.txt etc
2. un fichier de mappage: rtf.xml
3. un fichier de métadonnées (EML) : eml.xml Lorsque vous publiez avec l'IPT, il crée une archive Darwin Core qui est partagée avec GBIF. En outre, lorsque vous téléchargez des données à partir du site Web du GBIF, vous pouvez également choisir un format DwC-A.

GUID

Identificateur Globalement Unique

IPT

L'Integrated Publishing Toolkit (IPT) est une application web libre et open source (logiciel) pour la publication des données sur la biodiversité. Le logiciel lui-même est hébergé sur un serveur (que ce soit dans votre institution ou ailleurs) qui doit avoir accès à Internet 24/7. Il est utilisé pour créer et gérer des fichiers d'Archives Darwin Core qui peuvent être partagés et utilisés par toute personne inclus dans le GBIF.

Prêt

Dans le contexte des collections d'histoire naturelle, il s'agit de la procédure de prêt de spécimens entre institutions.

LSID

Identificateur des sciences de la vie. Ils sont des identifiants persistants et uniques au niveau mondial pour les objets biologiques.

Publication des données

En ce qui concerne le GBIF, nous avons une définition très spécifique de la publication des données. Il s'agit de rendre les jeux de données sur la biodiversité accessibles au public et découvrables, sous une forme normalisée, via un point d'accès, généralement une adresse web (une URL).

Ressource

Une ressource est le terme utilisé pour désigner un jeu de données particulier et ses métadonnées une fois qu'il a été téléchargé dans une instance IPT.

TDWG

Groupe de travail sur les bases de données taxonomiques, maintenant renommé Standards d'information sur la biodiversité.

URN

Numéro de Ressource Uniforme

UUID

Identifiant Universel Unique

Remerciements

Conception du cours et instruction

Le succès de ce cours dépend fortement du soutien apporté aux participants par le réseau des mentors en renforcement des capacités du GBIF. Visitez la page du GBIF sur [mentorat pour le renforcement des capacités](#) pour en savoir plus sur ces personnes et leurs contributions.

Les personnes suivantes sont reconnues pour leurs contributions significatives à la conception, au matériel, et à l'enseignement du cours:

- Nadine Bowles Newark
- Andrea Baquero
- Kate Ingenloff
- Hannah Owens
- Mélianie Raymond
- Andrew Rodrigues
- Laura Anne Russell
- John Tayleur
- John Waller

Remerciements particuliers ...

Traducteurs

Espagnol

- Anabela Plos
- Paula Zermoglio

Français

- Patricia Mergen
- Anne Sophie Archambeau

Ressources

- R
- [Vignette de modélisation de la niche écologique de Wallace](#)
- [Liste rouge IUCN des espèces menacées](#)

Colophon

Suggestion de citation

GBIF Secretariat (2021) GBIF Biodiversity Data Use Course. 6th edition. GBIF Secretariat: Copenhagen. <https://doi.org/10.15468/ce-wkk4-2w26>. [Date du cours.]

Contributeurs

Le *Cours d'utilisation des données de biodiversité du GBIF* a été développé à l'origine dans le cadre d'[Informations sur la Biodiversité pour le Développement](https://europa.eu), un programme financé par l'<https://europa.eu>[Union européenne^]. Le programme initial a été créé par Andrew Rodrigues, Hannah Owens and John Tayleur, avec des contributions supplémentaires des formateurs, mentors et étudiants du GBIF.

Licence

Nom du Cours est sous licence [Creative Commons Attribution 4.0 non déclarée](https://creativecommons.org/licenses/by/4.0/).

URI permanent

<https://doi.org/10.15468/ce-wkk4-2w26>

Contrôle du document

Cinquième édition, novembre 2021

Image de couverture

Di Marco M, Ferrier S, Harwood TD, Hoskins AJ and Watson JEM (2019) Wilderness areas halve the extinction risk of terrestrial biodiversity. *Nature*. Springer Science and Business Media LLC 573(7775): 582–585. Available at: <https://doi.org/10.1038/s41586-019-1567-7>