

Cours de mobilisation des données sur la biodiversité

Secrétariat du GBIF

Version 12, May 2021



Table des matières

Description du cours	1
Public	1
Prérequis	1
Objectifs d'apprentissage	1
Certification	2
Fichiers à télécharger	2
Vidéos	2
Données des exercices	3
Feuilles d'exercice	3
Fondations	3
Terminologie	3
Définitions	3
Logiciel	3
Structures	4
Qualité des données	4
Documentation	8
Workflows de numérisation	9
Outils logiciels	10
Installer OpenRefine	11
Conditions requises pour l'installation	11
Installation sur MS Windows	11
Installation sur Mac	13
Installation sur Linux	19
Compte-rendu des fondations	21
Cas Pratique I - Spécimens d'herbier	25
Scénario	25
Description	26
Collecte de données	27
Description du jeu de données	27
Exercices	27
Planification	27
Capture de données	28
Gestion de données	28
Publication de données	28
Feuille d'exercice	28
Planification	28
Ressources	28
Organiser	28
Exercice 1a-c	28
Exercice 1a	30
Exercice 1b	30
Exercice 1c	31
Compte-rendu	31

Capture de données	32
Standards et Darwin Core	32
Origines et types de données	32
Saisie, traitement et qualité des données	33
Exercice 2	33
Compte-rendu	33
Gestion de données	39
Principes de gestion des données	39
Outils de gestion des données	39
Exercice 3a-c	39
Exercice 3a	39
Exercice 3b	40
Exercice 3c	40
Conseils pour l'exercice	41
Contrôles de validation	41
Outils utiles	42
Compte-rendu	42
Publication de données	43
Concepts de publication de données	43
Présentation générale de l'IPT	43
Installations IPT pour la formation	43
Démonstration de l'IPT	44
Exercice 4	44
Compte-rendu	44
Evaluation et certification	45
Rubrique de planification	45
Rubrique de saisie de données	47
Rubrique de gestion des données	49
Rubrique de publication des données	52
Cas Pratique II - Espèces envahissantes	54
Scénario	55
Description	55
Collecte de données	56
Description des données numériques	58
Feuille d'exercice des espèces envahissantes	58
Exercice 1	58
Exercice 1a	58
Exercice 1b	59
Exercice 2	59
Exercice 3	59
Exercice 4	60
Cas Pratique II - Observations de Lépidoptères	60
Scénario	60
Description	61
Collecte de données	62

Exemple de capture de données analogiques	62
Description des données numériques	63
Feuille d'exercice Lepidoptera	64
Exercice 1	64
Exercice 1a	65
Exercice 1b	66
Exercice 2	66
Exercice 3	66
Exercice 4	67
Cas pratique III - Oiseaux dans la littérature	67
Scénario	67
Description	68
Collection de données originale	69
Description des données analogiques	69
Description des données analysées et traduites	70
Description des données numériques	71
Feuille d'exercices sur les oiseaux de la littérature	71
Exercice 1	71
Exercice 2	72
Exercice 3	73
Exercice 4	73
Tâches finales	74
CAS PRATIQUE II	74
CAS PRATIQUE III	74
Soumission du travail	74
Évaluation du cours	74
Documents clés	74
Darwin Core	75
Publication de données	75
Publication de données : IPT	75
La numérisation	76
GBIF	76
Géoréférencement	76
Espèces invasives	76
Living Atlases	76
Divers	76
OpenRefine	77
Planification/Collaboration	77
Qualité	77
Espèces sensibles	78
Taxonomie	78
Glossaire	78
Annexe : Articles sur les données (data papers)	81
Annexe : Solutions	82
Solutions de révision des fondations	82

Solutions de la révision sur la planification	85
Solutions de vérification de la saisie des données	85
Solutions des révisions sur la gestion des données	87
Solutions des révisions sur la publication des données	88
Solution suggérée pour le Cas pratique I	90
Remerciements	90
Conception du cours et instruction	90
Traducteurs	90
Français	90
Portugais	91
Espagnol	91
Ressources	91
Support des ressources	91
Colophon	91
Suggestion de citation	91
Contributeurs	92
Licence	92
URI permanent	92
Contrôle de document	92

Description du cours

Ce cours permet aux participants de planifier et de mettre en œuvre efficacement les efforts de mobilisation des données sur la biodiversité en utilisant les standards communautaires acceptés. Son objectif est d'augmenter le volume, la richesse et la qualité des données publiées via le réseau GBIF. Ce cours a été développé pour la première fois dans le cadre du programme **Biodiversity Information for Development (BID)** financé par l' **Union Européenne**.

Sujets inclus :

- Gestion de projet
- Capture de données
- Gestion de données
- Publication de données

Ce cours comprend des instructions sous forme de vidéos associées à des quiz et des exercices pratiques. Lorsqu'ils sont proposés sous forme d'atelier sur place ou virtuel, le travail de groupe et l'interaction sociale sont encouragés.

Public

Ce cours s'adresse aux personnes qui travaillent en tant que chercheurs ou techniciens dans des institutions de recherche ou politiques dans le domaine de la biodiversité. Les instructions fournies sont particulièrement utiles pour ceux qui doivent ou souhaitent planifier des projets de mobilisation et/ou mobiliser des informations sur la biodiversité pour leurs institutions respectives.

Prérequis

1. **Introduction au cours du GBIF**
2. 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.
 - 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.
 - Volonté de diffuser les connaissances acquises lors de l'atelier avec les partenaires et collaborateurs de votre projet en adaptant les supports de formation sur la mobilisation des données sur la biodiversité à des contextes et des langues spécifiques tout en conservant 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

- Apprenez les concepts clés (fondements) de l'informatique de la biodiversité, en particulier la gestion des données numériques sur la biodiversité.

- Recevez une introduction au Standard Darwin Core et à ses composantes.
- Apprenez à comprendre les différentes étapes de la planification d'un projet de mobilisation et comment adapter ces étapes à un projet spécifique.
- Évaluer une stratégie de mobilisation des données pour identifier les lacunes, inefficacités et pièges potentiels.
- Développez une stratégie de mobilisation de données adaptée à un cadre institutionnel donné.
- Apprenez à identifier les types de données et comment saisir au mieux les informations pertinentes en utilisant les meilleures pratiques, les logiciels, outils et techniques existants.
- Utilisez des outils logiciels conçus pour faciliter la saisie des données sur la biodiversité afin de produire des données numériques sur la biodiversité à partir de sources analogiques.
- Apprenez les concepts de qualité des données et recevez une introduction aux outils utilisés pour standardiser, valider et nettoyer les données.
- Utilisez des outils logiciels pour évaluer l'adéquation d'un ensemble de données sur la biodiversité à son utilisation prévue.
- Utilisez des logiciels conçus pour le nettoyage des données (sur la biodiversité).
- Apprenez le processus permettant de mettre à disposition des données sur la biodiversité en format ouvert en ligne, également connu sous le terme de publication de données, en utilisant le kit de publication intégré (IPT) du GBIF.
- Définissez les types et sous-types (le cas échéant) de données publiables pour un ensemble de données sur la biodiversité.
- Utilisez le GBIF IPT pour publier des ensembles de données sur la biodiversité en utilisant les extensions appropriées.
- Renforcez les capacités des autres dans la planification, la saisie, la gestion et la publication de données sur la biodiversité.

Certification

Une fois le cours terminé avec succès et une évaluation réussie des devoirs (par des formateurs et des mentors), les participants ont la possibilité de recevoir un certificat officiel sous la forme d'un badge [Open Badge](#). Voir [Evaluation et Certification](#) pour de plus amples détails.

Fichiers à télécharger

Tous les fichiers du cours peuvent être téléchargés à partir de cette page. Si vous préférez, tous les fichiers sont également liés individuellement dans le cours au fil de leur apparition dans le programme. De plus, les fichiers vidéo sont intégrés dans le cours et peuvent être lus à partir de YouTube. Les sous-titres sont disponibles lors de la lecture depuis YouTube pour la plupart des vidéos. Si vous avez des difficultés à accéder aux vidéos intégrées, veuillez télécharger les fichiers mp4 pour les lire localement sur votre ordinateur.

Vidéos

Les vidéos sont en anglais. Les sous-titres ne sont pas disponibles pour les vidéos téléchargées.

[Foundations1.zip](#) (73,7 Mo)

[Foundations2.zip](#) (90,2 Mo)

Planning.zip (51,3 Mo)

Capture.zip (63,1 Mo)

Management.zip (30.2 MB)

Publishing.zip (77.9 MB)

Appendix-Data-Papers.zip (97,5 Mo)

Données des exercices

Ce **fichier compressé** (ZIP 37,7 Mo) contient les données des exercices pour les cas pratiques.

Feuilles d'exercice

Ce **fichier compressé** (ZIP 1,4 Mo) contient les fiches d'exercices pour tous les cas pratiques. Les fiches d'exercices sont rédigées en anglais et doivent être complétées en anglais.

Fondations



Ce module comprend des instructions pour s'assurer que tous les participants sont au même niveau avant de se plonger dans les sujets de mobilisation des données. Vous recevrez une introduction au langage, à la terminologie et aux définitions de certains des concepts, fonctions et processus de base que vous allez mettre en oeuvre pendant le reste du cours. Vous recevrez également une introduction à la qualité des données et découvrirez l'importance de la documentation. Enfin, il vous sera demandé d'installer le logiciel OpenRefine dans le cadre de ce module.

Terminologie

Définitions



Dans cette vidéo (12:02), vous passerez en revue la terminologie utilisée dans ce cours. Si vous ne parvenez pas à regarder la vidéo intégrée, vous pouvez la **télécharger** localement. (MP4 - 38,5 Mo)

► <https://www.youtube.com/watch?v=FZAF5Sy8Nsc> (YouTube video)

Logiciel



Dans cette vidéo (05:58), vous passerez en revue des exemples des différents types d'applications et de logiciels disponibles dans le domaine de l'informatique sur la mobilisation de la biodiversité. Si vous ne parvenez pas à regarder la vidéo intégrée, vous pouvez la **télécharger** localement. (MP4 - 18,9 Mo)

► <https://www.youtube.com/watch?v=vYfDIgBBKXY> (YouTube video)

Structures



Dans cette vidéo (13:10), vous passerez en revue les types de champs et de données qui contiennent des données, les structures qui aident à organiser et à protéger ces données et ce que cela signifie pour l'intégrité et la sécurité de vos données. Si vous ne parvenez pas à regarder la vidéo intégrée, vous pouvez la [télécharger](#) localement. (MP4 - 38,8 Mo)

► <https://www.youtube.com/watch?v=msnVbZvly2E> (YouTube video)

Qualité des données



Dans cette vidéo (12:26), vous passerez en revue la terminologie utilisée dans ce cours. Si vous ne parvenez pas à regarder la vidéo intégrée, vous pouvez la [télécharger](#) localement. (MP4 - 44,5 Mo)

► <https://www.youtube.com/watch?v=5o7TcS2K7Cw> (YouTube video)



Vous trouverez ci-dessous une sélection de recommandation de lecture du guide d'Arthur Chapman "Principles of data quality". [Le document complet](#), des références et des traductions peuvent être trouvées sur GBIF.org.

Avant qu'une discussion détaillée sur la qualité des données et son application aux données d'occurrence des espèces puisse avoir lieu, il y a un certain nombre de concepts qui doivent être définis et décrits. Ceux-ci incluent le terme de qualité des données lui-même, les termes exactitude et précision qui sont souvent mal appliqués, et ce que nous entendons par données primaires sur les espèces et données d'occurrence des espèces.

Données d'occurrence d'espèces

Les données d'occurrence d'espèces sont utilisées ici pour inclure des données d'étiquettes de spécimens attachées à des spécimens ou des lots hébergés dans des musées et des herbiers, des données d'observation et des données d'enquêtes environnementales. En général, les données sont ce que nous appelons «basées sur des points», bien que des lignes (données de transect provenant d'enquêtes environnementales, des collectes le long d'une rivière), des polygones (observations à l'intérieur d'une zone définie comme un parc national) et des données de grille (observations ou les enregistrements d'enquête d'une grille régulière) sont également inclus. En général, nous parlons de données géoréférencées - c'est-à-dire d'enregistrements avec des références géographiques qui les lient à un endroit particulier dans l'espace - que ce soit avec une coordonnée géoréférencée (ex: latitude et longitude, UTM) ou non (description textuelle d'une localité, altitude,

profondeur) - et l'heure (date, heure du jour).

En général, les données sont également liées à un nom taxonomique, mais des collections non identifiées peuvent également être incluses. Le terme a parfois été utilisé de manière interchangeable avec le terme "primary species data".

Données primaires sur les espèces

«Les données primaires sur les espèces» sont utilisées pour décrire les données brutes des collections et les données sans aucun attribut spatial. Cela comprend des données taxonomiques et nomenclaturales sans attributs spatiaux, tels que des noms, des taxons et des concepts taxonomiques sans références géographiques associées.

Exactitude et Précision

L'exactitude et la précision sont régulièrement confondues et les différences ne sont généralement pas comprises.

La précision fait référence à la proximité des valeurs mesurées, des observations ou des estimations par rapport à la valeur réelle ou vraie (ou à une valeur acceptée comme vraie - par exemple, les coordonnées d'un point de contrôle de topographie).

La précision (ou résolution) peut être divisée en deux types principaux : la précision statistique est la proximité avec laquelle les observations répétées sont conformes à elles-mêmes. Ils n'ont rien à voir avec leur relation avec la valeur réelle et peuvent avoir une grande précision, mais une faible exactitude. La précision numérique est le nombre de chiffres significatifs dans lesquels une observation est enregistrée et est devenue beaucoup plus évidente avec l'avènement des ordinateurs. Par exemple, une base de données peut afficher un enregistrement de latitude/longitude décimale à 10 décimales, c'est-à-dire ca . 1 mm quand en réalité l'enregistrement a une résolution de moins de 10-100 m (3-4 décimales). Cela donne souvent une fausse impression à la fois de la résolution et de la précision.

Ces termes - exactitude et précision - peuvent être appliqués tant aux données non spatiales que spatiales. Par exemple, une collection peut avoir une identification au niveau des sous-espèces (i.e. avoir une grande précision), mais être le mauvais taxon (c'est-à-dire avoir une faible exactitude), ou être identifié uniquement au niveau de la famille (haute exactitude, mais faible précision).

Qualité des données

La qualité des données est multidimensionnelle et implique la gestion des données, l'analyse et la modélisation, l'assurance de la qualité et le contrôle, le stockage et la présentation. Comme l'ont déclaré Chrisman (1991) et Strong et al. (1997), la qualité des données est liée à l'utilisation et ne peut être évaluée indépendamment de l'utilisateur. Dans une base de données, les données n'ont pas de qualité ni de valeur réelle (Dalcin 2004) ; ils n'ont que de la valeur potentielle, qui est aperçue lorsque quelqu'un utilise ces données pour faire quelque chose d'utile. La qualité de l'information se rapporte à sa capacité à satisfaire ses clients et à répondre aux besoins de ceux-ci (English 1999).

Redman (2001), a suggéré que pour que des données soient aptes à être utilisées, elles doivent être accessibles, exactes, opportunes, complètes, cohérentes avec d'autres sources, pertinentes, compréhensives, fournir un niveau de détail approprié, et être faciles à lire et à interpréter.

Un problème que le dépositaire des données pourrait devoir examiner est ce qu'il pourrait être nécessaire de faire avec la base de données pour augmenter sa convivialité pour un public plus large (i.e. - augmenter son utilisation potentielle ou sa pertinence) et la rendre donc plus adaptée à un large éventail de finalités. Il y aura un compromis entre une facilité d'utilisation accrue et la quantité d'efforts nécessaires pour ajouter des fonctionnalités supplémentaires et accroître cette facilité d'utilisation. Cela peut nécessiter des éléments tels que l'atomisation des champs de données, l'ajout d'informations de géo-référencement, etc.

Assurance Qualité/Contrôle de la Qualité

La différence entre le contrôle de la qualité et l'assurance de la qualité n'est pas toujours claire. Taulbee (1996) fait la distinction entre le contrôle de la qualité et l'assurance de la qualité et souligne que l'un ne peut exister sans l'autre si l'on veut atteindre les objectifs de qualité. Elle définit le contrôle de qualité comme un jugement de qualité basé sur des normes, des processus et des procédures internes établis pour contrôler et surveiller la qualité ; et l'Assurance de la qualité comme un jugement de qualité basé sur des normes externes au processus et est l'examen des activités et des processus de contrôle de la qualité pour s'assurer que les produits finis répondent à des normes de qualité prédéterminées.

Dans une approche plus orientée vers l'entreprise, Redman (2001) définit l'assurance de la qualité comme étant « les activités qui sont conçues pour produire des produits d'information sans défaut pour répondre aux besoins les plus importants des clients les plus importants au prix le plus bas possible ».

La manière dont ces conditions doivent être appliquées dans la pratique n'est pas claire, et dans la plupart des cas, les termes semblent être largement utilisés de manière synonymique pour décrire la pratique générale de la gestion de la qualité des données.

Incertitude

L'incertitude peut être considérée comme une « mesure de l'incomplétude des connaissances ou des informations sur une quantité inconnue dont la valeur réelle pourrait être établie si un dispositif de mesure parfait était disponible » (Cullen and Frey 1999). L'incertitude est une propriété de la compréhension des données par l'observateur, et elle concerne davantage l'observateur que les données en soi. Il y a toujours une incertitude dans les données; la difficulté est d'enregistrer, de comprendre et de visualiser cette incertitude afin que d'autres puissent la comprendre. L'incertitude est un terme clé pour comprendre les risques et l'évaluation des risques.

Erreur

L'erreur englobe à la fois l'imprécision des données et leurs inexactitudes, et de nombreux facteurs contribuent à l'erreur. L'erreur est généralement perçue comme étant aléatoire ou systématique. L'erreur aléatoire tend à faire référence à une déviation de l'état réel de manière aléatoire. Une erreur systématique ou un biais résulte d'un changement uniforme des valeurs et est parfois décrit comme ayant une « exactitude relative » dans le monde cartographique (Chrisman, 1991). Dans la détermination de l'« adéquation à l'usage », une erreur systématique peut être acceptable pour certaines applications et ne pas convenir pour d'autres.

Un exemple peut être l'utilisation d'un datum géodésique¹ différent - quand il est utilisé tout au long de l'analyse, il ne peut pas causer de problèmes majeurs. Des problèmes se poseront toutefois lorsqu'une analyse utilise des données de différentes sources et avec différents biais - par exemple des sources de données qui utilisent des données géodésiques différentes, ou lorsque des identifications peuvent avoir été effectuées en utilisant une version antérieure d'un code nomenclatural.

« Parce que l'erreur est inévitable, elle devrait être reconnue comme une dimension fondamentale des données » (Chrisman, 1991). Il est seulement possible de répondre aux questions sur les limitations des données, et même sur les limites des connaissances actuelles lorsque l'erreur est incluse dans la représentation des données. Les erreurs connues dans les trois dimensions de l'espace, de l'attribut et du temps

doivent être mesurées, calculées, enregistrées et documentées.

Validation et nettoyage

La validation est un processus utilisé pour déterminer si les données sont inexactes, incomplètes ou déraisonnables. Le processus peut inclure des vérifications de format, des vérifications de complétude, des vérifications du caractère raisonnable, des vérifications de limite, un examen des données pour identifier les aspects aberrants (géographiques, statistiques, temporels ou environnementaux) ou autres erreurs, et l'évaluation des données par des experts du domaine (i.e. spécialistes taxinomiques). Ces processus se traduisent généralement par le signalement, la documentation et la vérification ultérieure des enregistrements suspects. Les contrôles de validation peuvent également impliquer la vérification de la conformité aux standards, aux règles et aux conventions applicables. Une étape clé de la validation et du nettoyage des données est d'identifier les causes profondes des erreurs détectées et de se concentrer sur la prévention de la réapparition de ces erreurs (Redman 2001).

Le nettoyage des données (appelé « data cleaning » en anglais) désigne le processus de « correction » des erreurs dans les données qui ont été identifiées au cours du processus de validation. Le terme est synonyme de « purification des données » (« data cleansing » en anglais), bien que certains utilisent le terme purification des données pour englober à la fois la validation et le nettoyage des données. Dans le processus de nettoyage des données, il est important que les données ne soient pas perdues par inadvertance et que les changements apportés aux informations existantes soient effectués avec le plus grand soin. Il est souvent préférable de conserver à la fois les anciennes (données originales) et les nouvelles (données corrigées) côte à côte dans la base de données de sorte que si des erreurs sont commises dans le processus de nettoyage, les informations originales peuvent être récupérées.

Documentation



Dans cette vidéo (09:47), nous vous donnerons un aperçu de l'importance de la documentation en ce qui concerne la gestion et la publication des données. Vous en apprendrez plus sur la cartographie des données (mapping), les relations entre données et les métadonnées. Si vous ne pouvez pas regarder la vidéo intégrée, vous pouvez la [télécharger](#) localement. (MP4 - 29,2 Mo)

▶ <https://www.youtube.com/watch?v=Z5-SYImGRGc> (YouTube video)

Workflows de numérisation



Cette vidéo (07:20) sur les workflows de numérisation identifie cinq groupes (ou étapes) dans le processus de numérisation des objets de collection d'histoire naturelle à l'aide d'images numériques. Ces étapes peuvent être facilement adaptées à d'autres sources de données sur la biodiversité. Si vous ne pouvez pas regarder la vidéo intégrée, vous pouvez la [télécharger](#) localement. (MP4 - 26,8 Mo)

▶ <https://vimeo.com/120369455> (Vimeo video)



Comme le souligne la vidéo, les protocoles de numérisation varient d'une institution à l'autre, mais il est essentiel que le protocole choisi soit convenu, documenté et respecté.

Pendant cet atelier, nous n'enseignons pas la numérisation en soi, car ce sujet nécessiterait son propre cours (d'au moins une semaine). Nous nous focalisons plutôt à introduire la base de la saisie de données sur la biodiversité. Cependant, nous voulons vous fournir des ressources sur la numérisation, car nous pensons que beaucoup d'entre vous seront intéressés.

Il existe de nombreuses façons d'organiser les efforts de numérisation et la numérisation peut donc paraître intimidante au départ. Il est important de se rappeler que, dans la plupart des cas, quelqu'un d'autre a déjà essayé de numériser les mêmes types de spécimens et objets que vous. Dans cet exercice, nous vous présentons quelques ressources pratiques d'ordonnement des tâches de numérisation pour vous aider à démarrer. Celles-ci constitueront également la base du travail que nous ferons au cours de l'atelier sur la sélection, la modification et l'évaluation du déroulement des tâches.

Certaines étapes du processus peuvent inclure :

- **L'édition du contenu et les étapes avant la numérisation** : Cela comprend la préparation des données d'origine pour le processus de numérisation, y compris l'assignation d'identifiants uniques qui aideront à se référer à la source originelle sans erreur et à garder toutes les informations dérivées ensemble.
- **Saisie d'images** : Cela inclut une bonne partie de planification, pas seulement sur la saisie d'image elle-même (par ex. définition de la séquence de tâches, sélection du matériel adéquat), mais aussi comment les images seront gérées et où elles seront stockées.
- **Traitement d'image** : Cela inclut le contrôle de qualité, la conversion de fichiers, etc.
- **Saisie électronique des données** : Le cœur du processus de numérisation, comprend la saisie de l'information clés dans une base de données. La vidéo souligne que la méthode la plus commune de saisie de l'information est via un clavier, mais de plus en plus d'institutions se tournent vers des technologies avancées de saisie de données.
- **Géoréférencement** : Les informations géographiques sont très importantes pour l'analyse de la biodiversité, donc les projets de numérisation devraient chercher à extraire l'information géographique la plus précise possible.

Le Integrated Digitized Biocollections (**iDigBio**) est le centre de coordination des Ressources Nationales pour la Numérisation Avancée des Collections sur la Biodiversité des Etats Unis (**ADBC**). Il mène un effort national visant à rendre disponibles les données et les images de millions de spécimens biologiques dans un format électronique standard pour la communauté des chercheurs, les organismes gouvernementaux, les étudiants, les éducateurs et le grand public. Il a produit plusieurs vidéos qui abordent les processus de numérisation.

Il y a d'autres vidéos dans la série iDigBio qui pourraient vous intéresser, si vous souhaitez en savoir plus sur l'ordonnancement spécifique des tâches pour différents types de spécimens :

- « Numérisation des collections conservées dans des liquides » (4:34 mn) <https://vimeo.com/120369690>
- « Processus de prise de vues pour la numérisation des spécimens de vertébrés conservés à sec » (7:25 mn) <https://vimeo.com/160615629>
- « Numérisation des spécimens d'herbier » (7:34 mn) <https://vimeo.com/120369768>

Outils logiciels



Evaluation d'outils logiciels utilisés en informatique de la biodiversité

Au cours des activités du cours, nous ferons des démonstrations et travaillerons avec différents outils logiciels liés à la numérisation des données, à la qualité des données et à leur transformation. Vous en utilisez probablement déjà plusieurs dans votre travail quotidien.

Les formateurs, les mentors et les anciens participants au cours de notre communauté ont dressé une liste contenant des informations sur les outils en informatique de la biodiversité. Elle fournit des liens vers leurs sites internet, des informations clés et un résumé de leurs points forts et points faibles.

Lien de téléchargement [Software-database-EN.xlsx](#). (23 ko)

Lors de l'analyse d'un logiciel pour la biodiversité que vous n'avez pas encore utilisé, vous devez réfléchir à la manière dont vous l'adapterez à vos besoins. Vous trouverez ci-dessous une liste à partir de laquelle vous pouvez commencer votre évaluation. Elle s'inspire du chapitre "Caractéristiques d'une bonne solution de base de données" du manuel du GBIF "[Lancer un projet de numérisation](#)" :

- **Prix** : l'un des facteurs les plus déterminant pour votre projet. Prenez en compte les autres coûts au-delà du prix de la licence logicielle, comme le matériel informatique nécessaire pour l'exécuter, la maintenance, les mises à jour et l'expertise nécessaire pour l'utiliser.
- **Fonctionnalité** : Vous devez avoir une vision claire de ce que vous attendez du logiciel et vous assurer qu'il le fait efficacement. Ne soyez pas distrait par des fonctionnalités supplémentaires qui peuvent rendre le logiciel plus complexe inutilement.
- **Stabilité** : Certaines solutions qui sont sur le marché depuis longtemps et qui sont supportées par des institutions ou des entreprises solides sont plus susceptibles d'être exemptes de bogues et/ou d'avoir de bons systèmes en place pour résoudre les problèmes qui se posent. Elles sont aussi plus susceptibles d'être mises à jour et portées par des systèmes d'exploitation plus modernes.
- **Extensibilité** : Certains logiciels fonctionnent très bien lorsqu'ils sont présentés de manière prête-à-emploi, mais leurs performances se dégradent après un certain temps ou lors de leur utilisation avec une plus grande quantité de données ou lorsque plusieurs utilisateurs y accèdent simultanément. Vérifiez les opinions des autres utilisateurs en ligne.
- **Intégration** : Assurez-vous que le logiciel accepte et produit les formats de données que vous utilisez et dont vous avez besoin. La transformation de données est une tâche qui prend du temps.
- **Prise en charge de la langue** : il est essentiel que tous les utilisateurs du logiciel puissent comprendre son interface et la documentation qui rendra possible son utilisation.
- **Documentation et support technique** : assurez-vous d'explorer la documentation et les

supports existants. Vous en aurez besoin à un moment ou à un autre.

- **Courbe d'apprentissage** (« Learning curve » en anglais): Certains logiciels nécessitent des formations spécifiques, alors que d'autres sont plus intuitifs et s'apprennent à l'usage, en s'appuyant sur les systèmes d'aide en ligne.

Installer OpenRefine



Installer le logiciel requis pour les activités ultérieures du cours



OpenRefine est un outil contenant un ensemble de fonctionnalités pour travailler avec des données tabulaires, qui améliore la qualité globale d'un jeu de données. C'est une application qui s'exécute sur votre ordinateur comme mini serveur web. Pour l'utiliser, votre navigateur web devrait pointer vers ce serveur web. Donc, considérez OpenRefine comme une application web personnelle et privée.

Nous utiliserons OpenRefine pendant la partie du cours concernant la mobilisation des données, en particulier pendant les exercices pratiques. Il sera nécessaire d'installer OpenRefine sur votre ordinateur portable. Si vous êtes sûr de vos capacités, vous pouvez suivre les étapes fournies dans le lien ci-dessous, pour installer le logiciel sur votre ordinateur. Si vous n'êtes pas sûr d'y arriver, veuillez demander de l'aide. Suivez ce lien pour l'installation : [OpenRefine download page](#) pour plus de détails.



Des mots de passe administrateurs peuvent être requis pour installer ces logiciels.

Conditions requises pour l'installation

1. Seulement pour les utilisateurs de Linux : Java JRE installé.
2. Google Chrome, Microsoft Edge ou Mozilla Firefox installés. Internet Explorer n'est pas pris en charge.



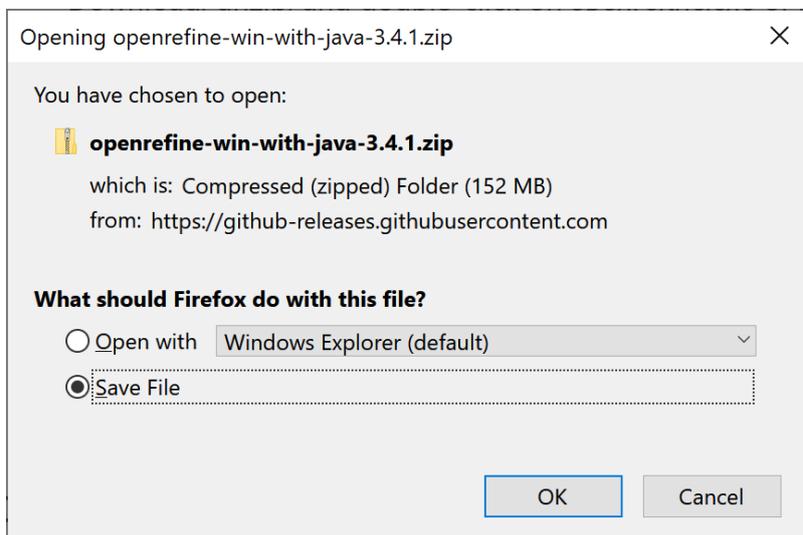
La dernière version stable est OpenRefine 3.4.1, publiée le 24 septembre 2020. Les instructions d'installation détaillées sont disponibles sur <https://docs.openrefine.org/manual/installing/>.

Installation sur MS Windows

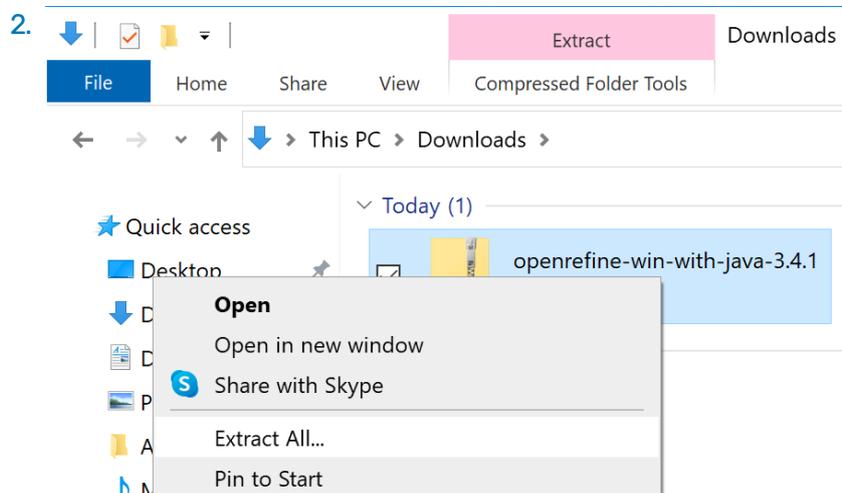
1. Téléchargez le **Kit Windows with embedded Java**. Choisissez d'enregistrer le fichier plutôt que de l'ouvrir.
2. Trouvez le fichier téléchargé. Faites un clic droit dessus et choisissez "Extract all...". Dézippez, et double-cliquez sur openrefine.exe ou refine.bat si le premier ne fonctionne pas.
3. Une fenêtre de commande apparaîtra (ne la fermez pas) et aussitôt, une nouvelle fenêtre de navigateur Web affichera l'application.

▼ *Instructions détaillées pour MS Windows (cliquez pour développer)*

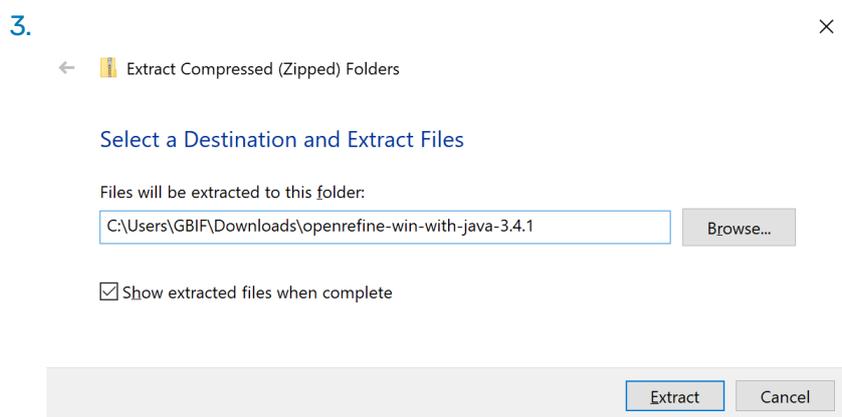
- 1.



Téléchargez le **Kit Windows with embedded Java**. Choisissez d'enregistrer le fichier plutôt que de l'ouvrir.

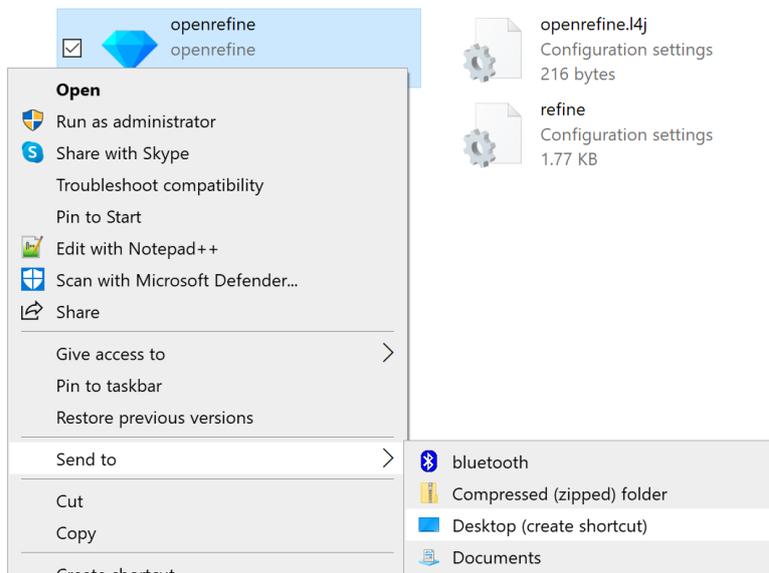


Trouvez le fichier que vous avez téléchargé. Faites un clic droit dessus et choisissez "Extract all..."

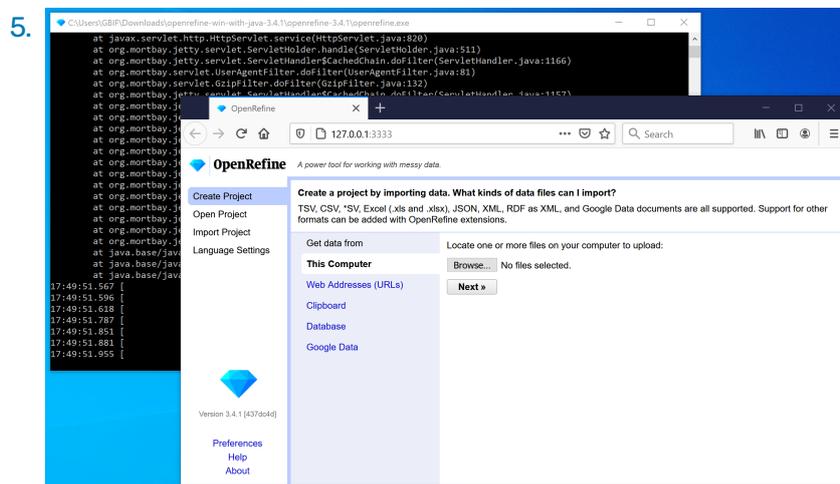


Cliquez sur "Extract"

4.



Trouvez les fichiers extraits. En option, faites un clic droit sur « openrefine» et choisissez « "Send to → Desktop (create shortcut)» pour créer un raccourci sur votre bureau. Puis double-cliquez sur "openrefine"



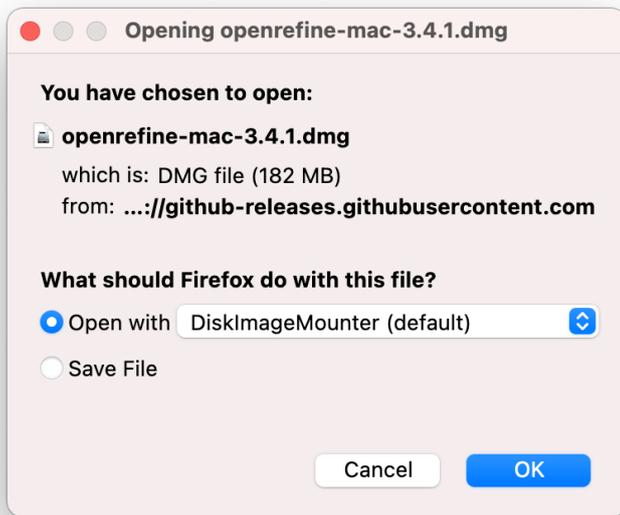
Une fenêtre de console noire s'ouvre, et peu de temps plus tard le navigateur s'ouvre. OpenRefine est maintenant prêt à l'emploi.

Installation sur Mac

1. Téléchargez le [Mac kit](#).
2. Téléchargez, ouvrez, glissez l'icône dans le dossier d'applications. Vous n'avez pas besoin d'installer Java séparément.
3. Faites un double clic dessus et une nouvelle fenêtre de navigateur Web affichera l'application.

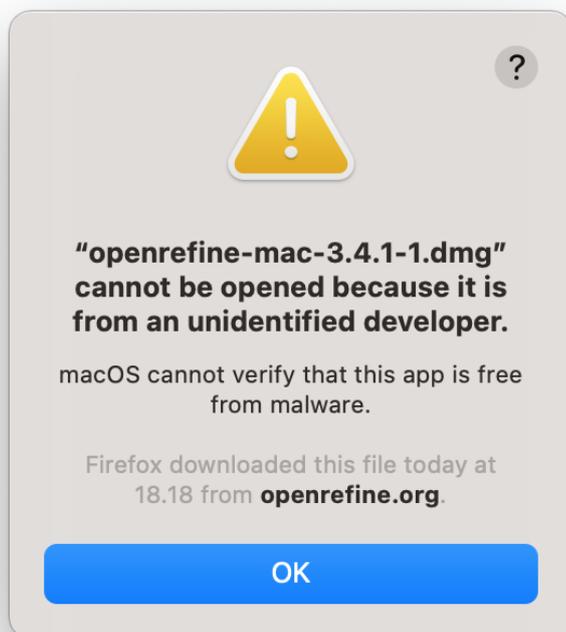
▼ Instructions détaillées pour Mac (cliquez pour développer)

- 1.



Téléchargez le [Mac kit](#).

2.



Un avertissement est affiché. Cliquez sur "OK".

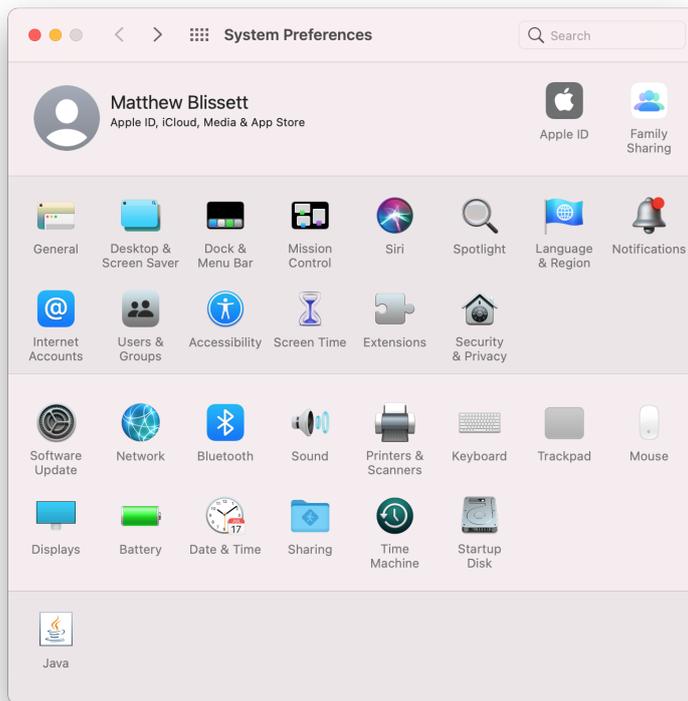
3.



System Preferences

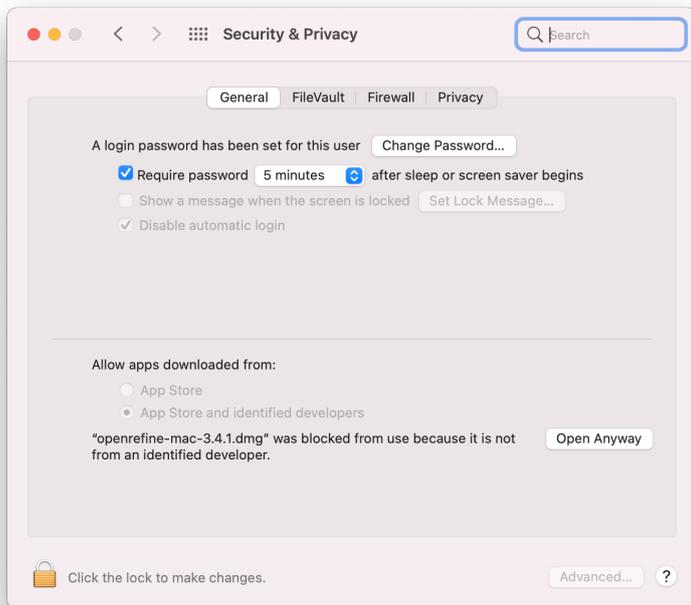
Ouvrir les Préférences Système.

4.



Ouvrir Sécurité et confidentialité

5.



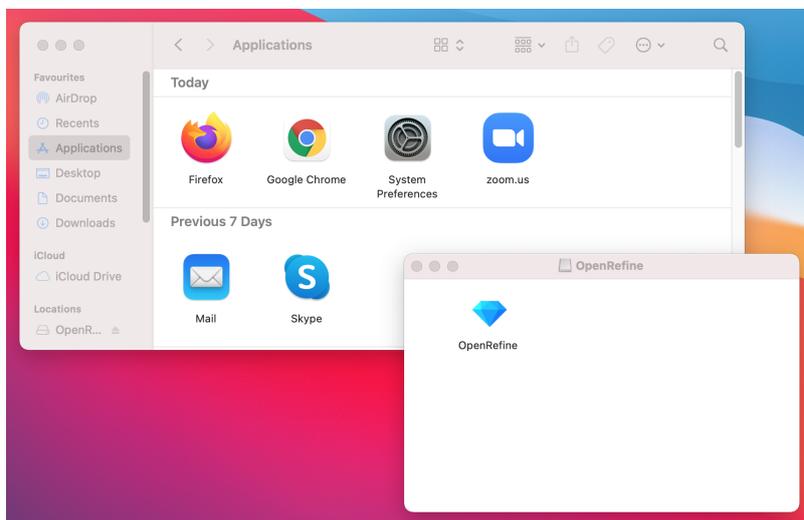
Choisissez "Open Anyway" en bas de page.

6.



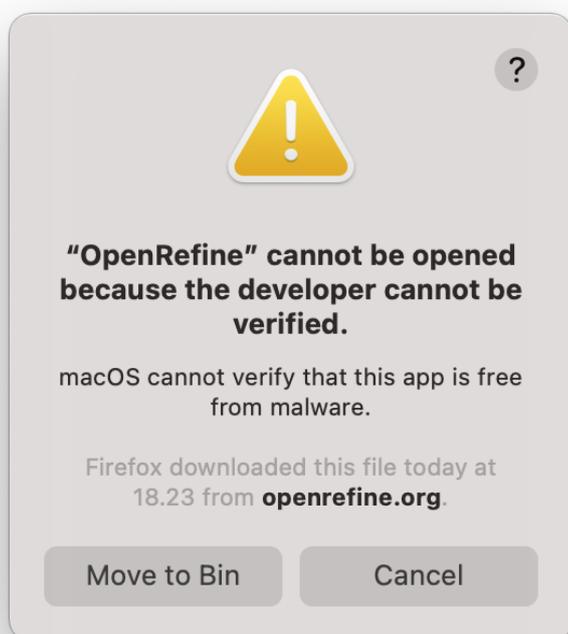
Choisissez "Open"

7.



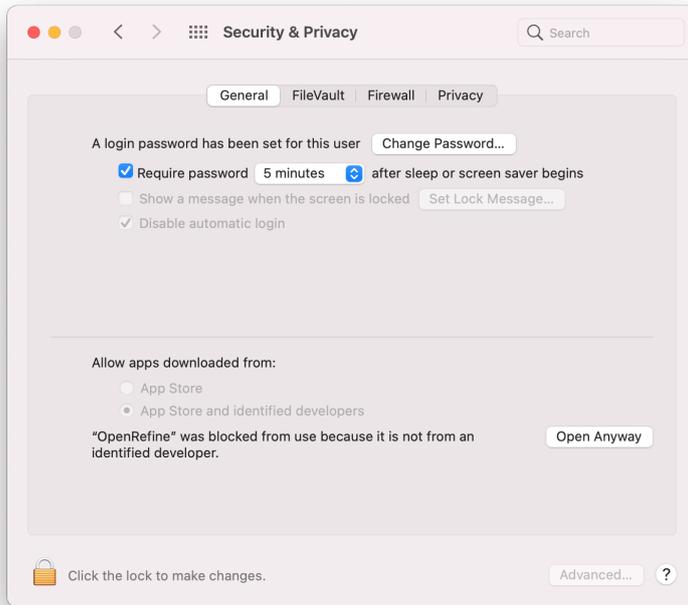
Enfin, l'archive de l'application est ouverte ! Glissez-la dans votre dossier Applications.

8.



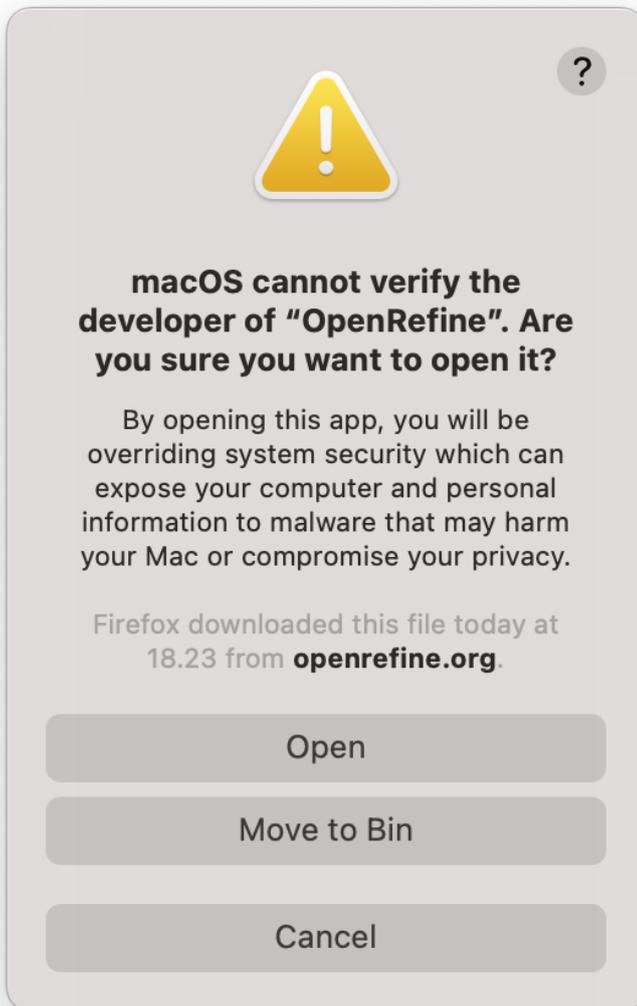
Double-cliquez sur l'icône OpenRefine. Une autre alerte de sécurité s'affiche !

9.



Retournez à "Security & Privacy" et cliquez sur "Open Anyway" – à nouveau.

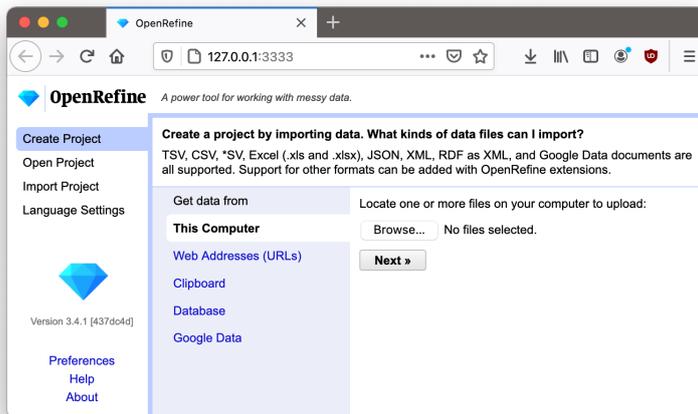
10.



(Pour éviter ces avertissements, les développeurs d'OpenRefine devaient payer Apple.)

Cliquez sur "Open".

11.



Enfin ! L'application est en cours d'exécution.

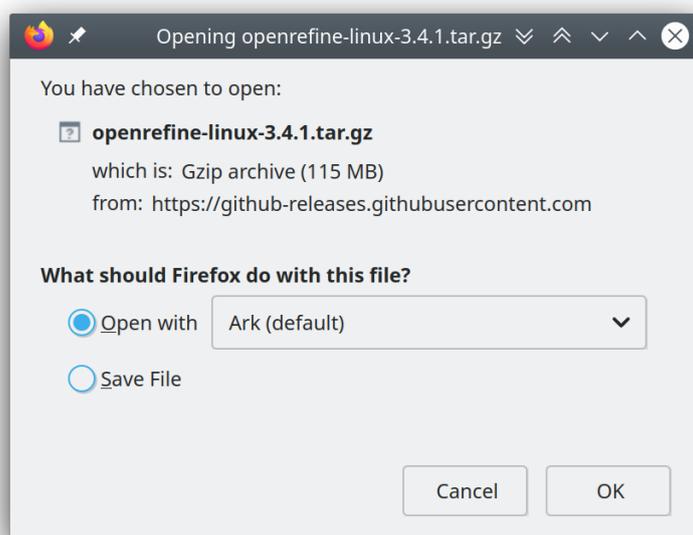
Installation sur Linux

1. Téléchargez le [Linux kit](#).
2. Téléchargez, extrayez, puis tapez `./refine` pour démarrer. Cela nécessite que Java soit installé sur votre ordinateur.

▼ Instructions détaillées pour Linux (cliquez pour développer)

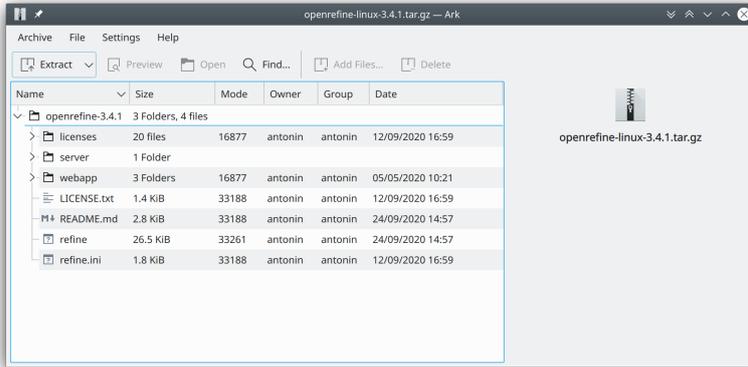
Ces instructions sont pour KDE (par exemple, Kubuntu, SuSE), mais le processus est similaire pour Gnome (par exemple, Ubuntu, Red Hat, CentOS).

1.



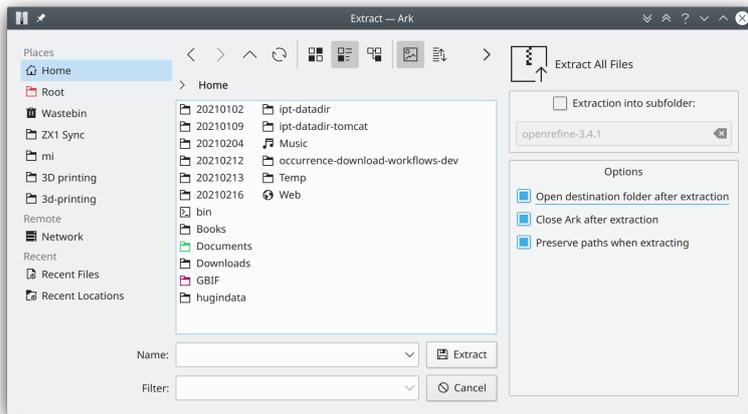
Téléchargez le [Linux kit](#).

2.



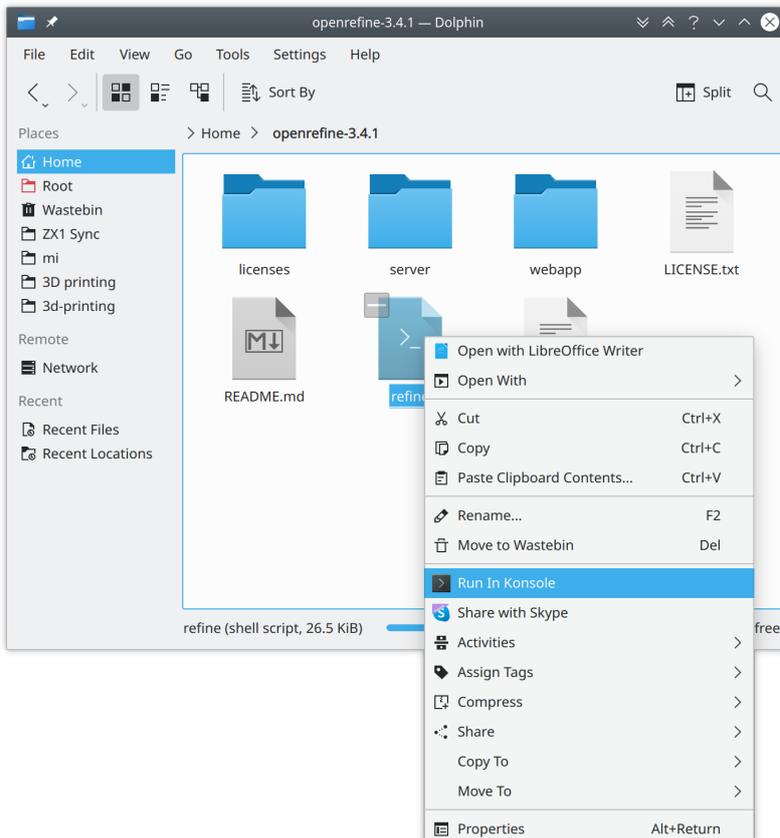
Cliquez sur « Extract » pour décompresser l'application téléchargée.

3.



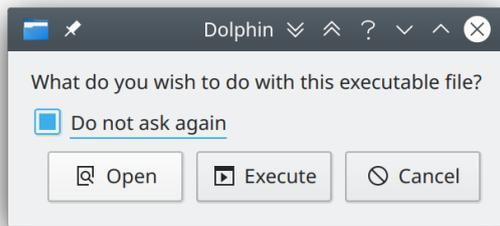
Choisissez un endroit approprié. Sélectionnez aussi "Open destination folder after extraction" et "Close Ark after extraction"

4.



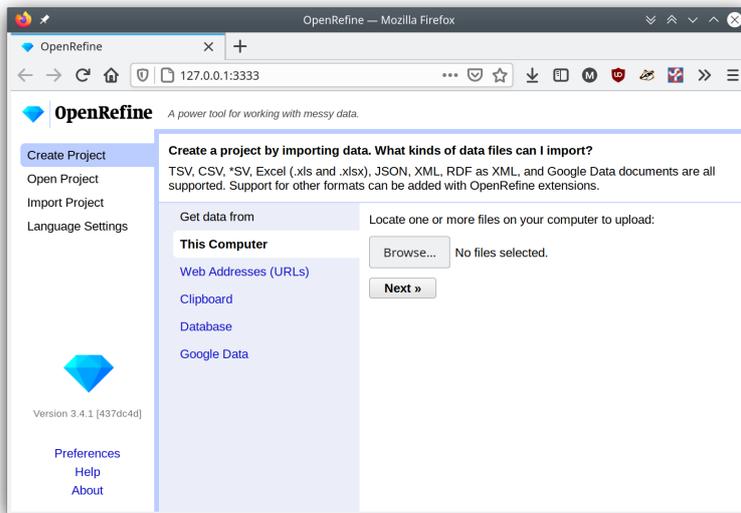
Faites un clic droit sur « refine » et choisissez « Run in Konsole », ce qui est nécessaire pour que vous puissiez quitter en toute sécurité OpenRefine plus tard, en fermant la fenêtre Konsole.

5.



Confirmez que vous souhaitez exécuter l'application téléchargée.

6.



OpenRefine est en cours d'exécution.

Compte-rendu des fondations



Passez un quiz sur les concepts que vous avez appris dans cette section.

1. Pour l'énoncé donné, entrez le terme correct (base de données, langage de base de données, programme de base de données)
 - combine et présente des fonctions et des fonctionnalités pour manipuler des données ensemble dans une interface unifiée

 - une collection structurée et organisée de données et/ou d'informations conservées sur un ordinateur

 - la façon dont un être humain communique avec un ordinateur

2. Si vous ouvrez un fichier de données et vous voyez ce qui suit, que soupçonneriez-vous d'être le problème ?

Être, ou ne pas être, c'est là la question.

- Rien
- Il est corrompu
- Le mauvais encodage a été utilisé pour ouvrir le fichier
- L'expéditeur a utilisé une police bizarre

3. Pour le logiciel donné, saisissez le type de logiciel (saisie des données, gestion des données, nettoyage des données, publication des données).

◦ Integrated Publishing Toolkit (IPT)

◦ Specify

----- ET -----

◦ iNaturalist

◦ OpenRefine

4. Pour l'exemple donné, entrez le bon type de données (binaire, booléen, flottant, entier, entier long, texte, texte non structuré)

◦ 1236975

◦ 1101111

◦ Nous avons marché 5 miles sur la route à l'ouest du bureau de poste au centre de la ville. Ensuite, nous avons parcouru 2 miles vers le nord, sur un chemin de terre menant à la rivière. Puis nous avons continué vers l'ouest le long de la rivière pendant 5 miles.

◦ 1024

◦ 29.0

◦ Oui/Non

◦ 6 lapins ont été observés

5. Lesquels de ces termes décrivent un "nom de champ/colonne" ?

- Assigné
- Descriptif
- Identifiant
- Lisible
- Unique
- Interface utilisateur

6. Lesquels de ces termes décrivent une "étiquette du champ" ?

- Assigné
- Descriptif
- Identifiant
- Lisible
- Unique
- Interface utilisateur

7. Pour chaque énoncé, entrez la structure correcte (ligne, colonne, table)

- Toutes les données se réfèrent à un concept UNIQUE.

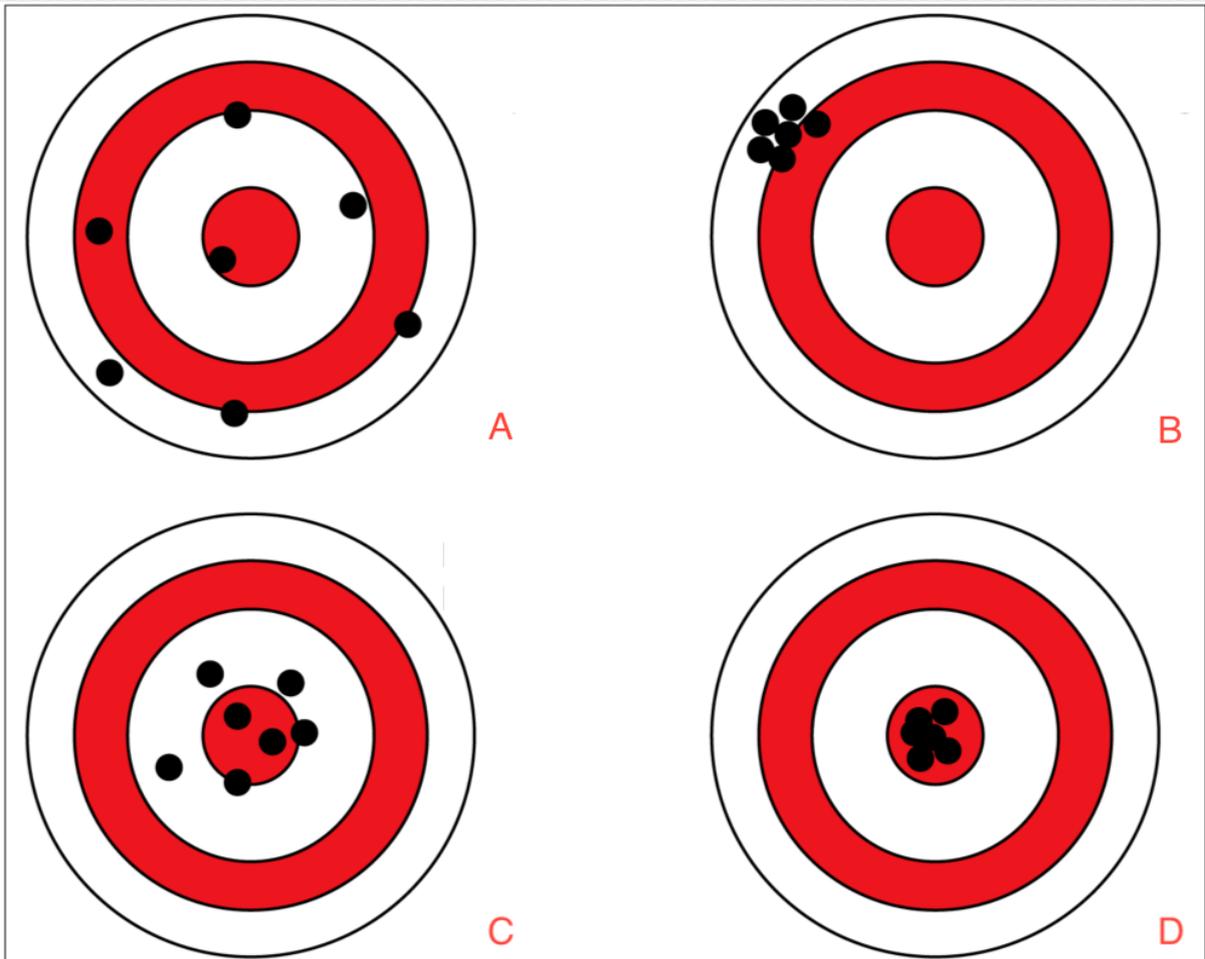
- Un attribut a le MÊME champ/type de données pour chaque enregistrement.

- Les attributs d'un enregistrement restent TOUJOURS ensemble.

8. Qui détermine l'adéquation à l'usage de vos données ?

- Le directeur du musée ou de département
- Les utilisateurs des données pour la recherche ou l'éducation
- Le collecteur des données sur le terrain
- La personne saisissant les données dans la base de données

9. Pour les instructions données, entrez l'image correspondante. (A, B, C, D)



- Haute exactitude, basse précision

- Faible exactitude, haute précision

- Haute exactitude, haute précision

- Faible exactitude, basse précision

10. Identifier les relations de données lorsque le jeu de donnée B doit être fusionné avec le jeu de donnée A (0:1, 1:0, 1:1, 1:∞, ∞:1, ∞:∞). Toutes les relations ne sont pas utilisées.

- Le champ Collecteur existe dans les deux jeux de données A et B

- Le champ Pays n'existe que dans le jeu de donnée B

- Le champ Nom existe dans le jeu de donnée A, mais le jeu de données B contient les champs Prénom et Nom

- Le champ ID existe dans le jeu de données A et B

- Élévation existe dans le jeu de données A, mais pas dans le jeu de données B

- La date existe dans le jeu de données A, mais Jour, Mois et Année sont des champs

séparés dans le jeu de données B

11. Les métadonnées sont importantes car (sélectionnez les affirmations vraies) :

- elles permettent aux utilisateurs de déterminer si un jeu de données convient à leur usage.
- elles vous permettent de partager des coordonnées exactes pour chaque occurrence.
- elles vous permettent de savoir sous quelles conditions légales la réutilisation des données est autorisée.
- cela s'applique également à tous le matériel supplémentaires et associés, y compris les images, les vidéos et autres média.
- elles vous permettent de connaître les prochaines heures d'exposition/d'ouverture de l'institution.

Cas Pratique I - Spécimens d'herbier



Familiarisez-vous avec le scénario du cas d'utilisation.

Le Cas Pratique I est utilisé pour les modules de planification, de saisie de données, de gestion de données et de publication de données. Il est recommandé de télécharger la [feuille d'exercice](#) (MS Word 345 ko) pour que vous puissiez prendre des notes tout au long des exercices. Une solution suggérée sera fournie en annexe. Le case pratique I n'est pas noté.

Scénario

Un projet de mobilisation des données dans un herbier régional



Eriocaulon bilobatum Morong collecté au Guatemala par Rapid Reference Collection (RRC) | Field Museum of Natural History - Keller Action Science Center (sous licence CC-BY-NC 4.0)

Ce récit a été développé comme base pour des exercices pratiques dans le cours de mobilisation des données sur la biodiversité. Le concept et le contenu de l'exercice ont été développés par Alberto Gonzaëlez-Talavañ, Néstor Beltrañ, Nicolas Noe' et Sharon Grant. Les données proviennent d'un véritable jeu de données, mais ont été modifiées pour ces exercices. C'est un scénario fictif qui n'est destiné qu'à des fins d'enseignement.

Description

L'Université de White Plains est un établissement d'enseignement tertiaire reconnu au Guatemala et une référence nationale pour la recherche sur la biodiversité. Le département de biologie végétale héberge un herbier de taille moyenne contenant environ 80 000 spécimens récoltés dans et autour de la région et datant du milieu du 20ème siècle à nos jours. La collection comprend des spécimens importants, y compris des spécimens types et des espèces endémiques.

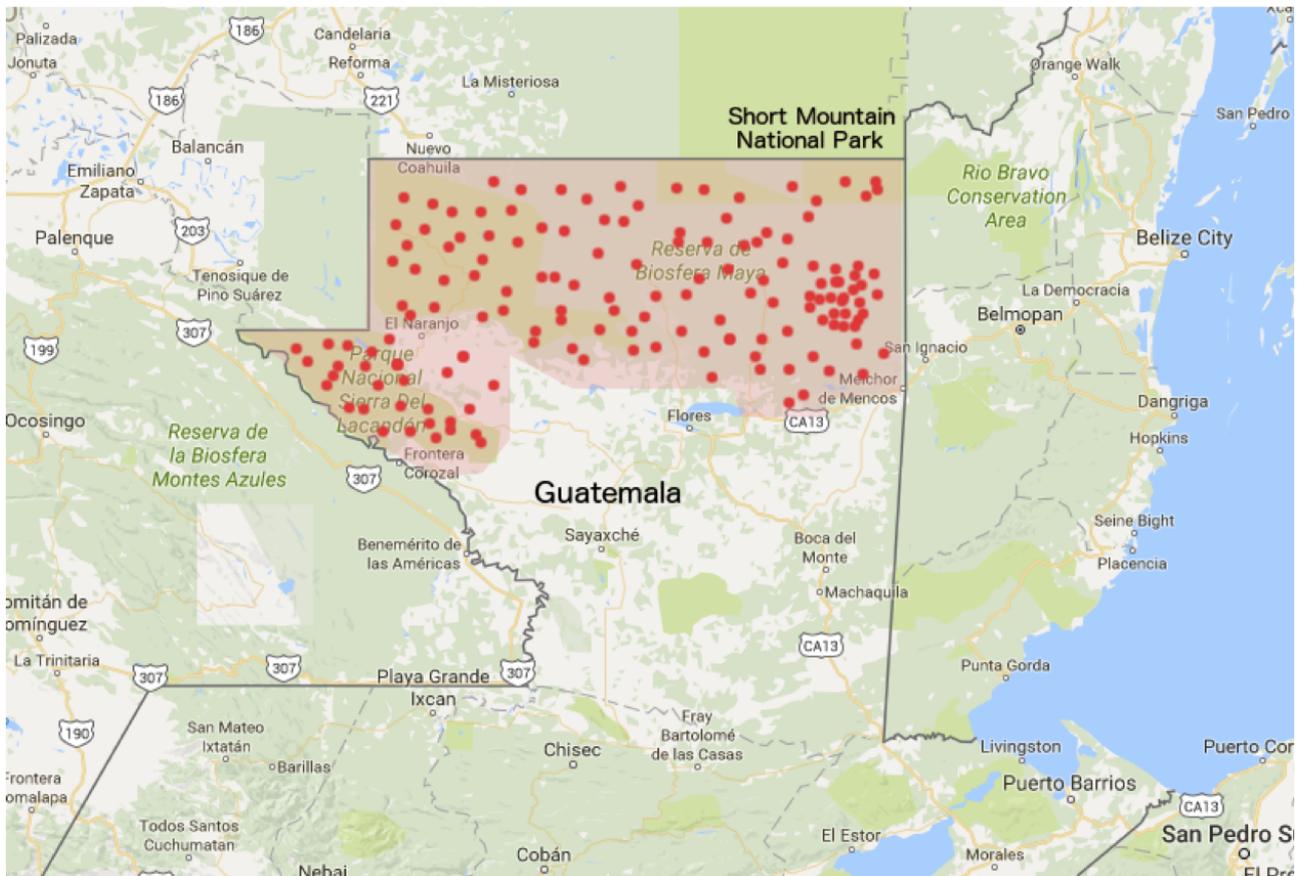
Actuellement, la conservation de la collection est la responsabilité du professeur de systématique végétale, qui accomplit les tâches de curateur ainsi que des travaux de recherche et d'enseignement réguliers. L'administrateur du département est responsable des tâches administratives quotidiennes de l'herbier, comme l'achat de consommables, incluant le papier et les étiquettes. Le personnel enseignant et les étudiants qui collaborent au sein du département y travaillent occasionnellement et mettent à jour les identifications des spécimens et il y a deux botanistes retraités qui se portent régulièrement volontaires dans la collection pour aider le professeur à préparer les prêts.

L'université a déjà une plateforme de recherche en ligne pour les collections de sa bibliothèque qui est maintenue par l'équipe web de l'université sur des serveurs hébergés en externe. Aucun spécimens d'histoire naturelle n'est actuellement disponible via cette plate-forme.

Le responsable du département de biologie végétale a récemment obtenu une subvention de 50 000

\$USD de deux ans pour photographier et publier en ligne les informations des collections botaniques. L'équipe veut profiter de cette occasion pour établir un protocole permanent de numérisation et de publication qui donnera une plus grande visibilité à l'herbier et continuera d'attirer des fonds.

Collecte de données



L'Herbier comprend environ 80 000 spécimens récoltés entre 1960 et 2015 concentrés principalement dans le hotspot de biodiversité – Parc national de Short Mountain. La collection s'agrandit encore grâce à des échanges, des dons et plusieurs projets de recherche actifs. Pour chaque spécimen, des renseignements sur le collecteur, l'heure, la date, le lieu et la taxonomie sont documentés. Jusqu'à présent, aucune mesure de contrôle de qualité n'a été effectuée sur les données et il n'y a pas eu d'imagerie systématique des spécimens.

Description du jeu de données

Le professeur a une simple base de données non relationnelle sur son ordinateur qui sert d'index pour un grand nombre de spécimens, mais pas tous. Toutes les images existantes des spécimens sont conservées localement par les chercheurs indépendants qui les ont prises.

Exercices

Les exercices individuels pour ce cas pratique sont situés dans les modules correspondants.

Planification

Exercice 1a-c

Capture de données

Exercice 2

Gestion de données

Exercice 3a-c

Publication de données

Exercice 4

Feuille d'exercice

UC1-Herbarium-Exercice_EN.docx (MS Word 345 ko)

Planification



Dans le module Planification, vous passerez en revue les étapes clés de la planification d'un projet et apprendrez comment créer un flux de travail viable. De plus, vous créerez un plan de projet ou flux de travail idéalisé basé sur le Cas d'Utilisation 1. Vous identifierez les objectifs, les tâches, les principaux intervenants et les rôles et attribuerez des tâches spécifiques aux étapes.

Ressources



Dans cette vidéo (15:09), vous apprendrez comment définir les éléments qui auront un impact sur un projet et examinerez les interactions entre les éléments. Si vous ne pouvez pas regarder la vidéo en ligne sur, vous pouvez la **télécharger** localement. (MP4 - 25,1 Mo)

▶ <https://www.youtube.com/watch?v=VRvUdMjd93c> (YouTube video)

Organiser



Cette vidéo (15:52) se concentre sur les tâches et les moyens de les organiser afin de préparer des plans pratiques réalisables et une documentation claire. Si vous ne pouvez pas regarder la vidéo en ligne sur, vous pouvez la **télécharger** localement. (MP4 - 26,2 Mo)

▶ <https://www.youtube.com/watch?v=uhhK6B2Vwls> (YouTube video)

Exercice 1a-c



Pour cette activité, vous utiliserez une plateforme d'exercices virtuelle. Consultez les instructions de l'exercice ci-dessous.

Références

- Définitions de rôles
- Définitions d'étape

Pour vous aider à réaliser les exercices de planification pour toutes les études de cas, nous avons mis à disposition cinq tables de jeu de cartes virtuelles.

Parce qu'il s'agit d'espaces partagés, nous vous demandons de respecter ces règles afin que tous ceux qui veulent les utiliser puissent le faire :

1. Nettoyez après vous. Si vous utilisez une table de cartes, assurez-vous de la retourner à son état propre et original lorsque vous avez terminé. Cela signifie que toutes les cartes sont retirées et enlever/supprimer toutes les nouvelles cartes que vous avez ajoutées au paquet. Vous pouvez le faire en cliquant sur les boutons de rappel sous chaque paquet de cartes.
2. Si la table est occupée, essayez une autre table. Si vous voyez qu'une table de cartes est utilisée, veuillez utiliser l'un des autres liens de table de cartes. Si vous ne trouvez pas de table de cartes vide, avvertissez-nous à training@gbif.org pour que nous puissions vérifier les tables pour être sûr que toutes les tables sont réellement utilisées (ou si nous avons besoin de créer une table supplémentaire pour vous).
3. Ne laissez pas de table vide. Ne laissez pas votre travail sur une table lorsque vous vous absentez pour une longue période de temps. Si vous devez vous absenter pendant une heure ou deux, c'est acceptable, mais s'il vous plaît ne laissez pas trainer les cartes pendant plus de 4 heures. Si vous avez besoin de quitter votre poste de façon prolongée, veuillez utiliser la plateforme eLearning pour informer le groupe et les formateurs du lien que vous utilisez, pourquoi vous devez laisser vos cartes et quand vous reviendrez pour terminer votre travail.
4. Prenez souvent des captures d'écran. Nous ne pouvons pas garantir que vos cartes ne seront pas réinitialisées accidentellement par quelqu'un de l'atelier ou par quelqu'un de votre propre projet. Vous ne pouvez pas réinitialiser les cartes une fois qu'elles ont été rappelées et réinitialisées. Si vous voulez soumettre une capture d'écran avec votre exercice, n'hésitez pas à le faire.
5. Les formateurs se réservent le droit de réinitialiser les tables de cartes. Si nous observons qu'une ou plusieurs tables contiennent des cartes qui n'ont pas été remises et qui n'ont pas été utilisées depuis plus de quatre heures sans aucune notification sur le forum de la plateforme eLearning, nous effacerons et réinitialiserons la table pour que les autres puissent l'utiliser.

Liens vers la salle de jeux (chaque table s'ouvrira dans une nouvelle fenêtre/onglet) :

BLEU : <https://playingcards.io/ndqppx>

VERT : <https://playingcards.io/zpwe9j>

ORANGE : <https://playingcards.io/868jgh>

VIOLET : <https://playingcards.io/g6khh7>

ROUGE : <https://playingcards.io/3pad4w>

JAUNE : <https://playingcards.io/w488nb>

Un grand merci à Jwalant Patel et Eric Ma pour avoir trouvé et aidé à créer les tables de jeu en ligne pour cet exercice et à Kate Webbink pour son expertise artistique. La plateforme de table de jeu en ligne est fournie par <https://playingcards.io/>.

Exercice 1a

Lire le **Cas Pratique I** (si vous ne l'avez pas déjà fait).

En utilisant les cartes, sélectionnez les objectifs qui correspondent le plus aux grandes lignes du projet et choisissez ensuite les tâches qui doivent être réalisées pour les accomplir. Ensuite, identifiez les personnes/ressources dont le projet dispose et assignez-leur des cartes de rôles, le cas échéant. Enfin, attribuez-les à leurs groupes d'intervenants et à leurs affiliations.

1. Examinez les cartes GOALS avec le groupe, choisissez et disposez celles qui correspondent au cas d'utilisation.
2. Examinez les cartes TASK avec le groupe.
3. Attribuez les cartes TASK à chacune des cartes GOAL sélectionnées à partir de la description du cas d'utilisation.
4. Identifiez les institutions et les personnes mentionnées dans le texte du cas d'utilisation et en prendre note.
5. Déposez les cartes AFFILIATIONS sur la table.
6. Examinez les cartes STAKEHOLDER, identifiez celles mentionnées dans le cas d'utilisation et décidez ensuite à quelle AFFILIATION elles appartiennent.
7. Examinez les cartes ROLE, identifiez celles mentionnées dans le cas d'utilisation et décidez à quel groupe de STAKEHOLDER elles appartiennent.
8. Prenez des notes sur les cartes si nécessaire.
9. Une fois que les cartes sont disposées, prenez des photos/captures d'écran.
10. Utilisez le **fichier d'exercice** pour donner vos réponses.

Question

Y a-t-il des ressources ou des objectifs manquants dans les cartes que vous estimez essentiels à la réussite du projet ? Notez-les sur la feuille de réponses.

Exercice 1b

En utilisant l'analyse des parties prenantes (stakeholders) et des objectifs (goals) de l'exercice 1a, créez un processus en utilisant les cartes STAGE.

1. Lisez à nouveau l'introduction du cas d'utilisation, si nécessaire.
2. Décidez de quelles TASKS sera responsable chaque ROLE.
3. Passez en revue les cartes STAGE et ordonnez les TASKS de façon appropriée.
4. Utilisez la feuille d'exercice téléchargée précédemment pour donner vos réponses.

Questions

- Existe-t-il des goulets d'étranglement évidents dans le processus ? Par exemple : y a-t-il trop de tâches pour un rôle ou une ressource en particulier ?
- Selon vous, quelles questions sont importantes pour réussir la mobilisation des données du point de vue de chaque intervenant/rôle ? Par exemple : Quels sont les livrables ? Sont-ils réalistes, en prenant en considération le cadre temporel du projet ?

- Reprenez vos notes et priorisez par ordre d'importance.
- Si vous avez le temps, vous pouvez explorer différentes combinaisons car différents scénarios sont possibles dans différents contextes, ou même essayez de décrire la situation pour votre propre projet.

Exercice 1c

Cet exercice devrait être utilisé lorsque le cours est enseigné en ligne, ou sur place en tant qu'activité de groupe.

Après les exercices, le présentateur de chaque groupe devra :

1. Présenter les parties prenantes (STAKEHOLDERS) ou les tâches (TASKS) manquantes, en expliquant pourquoi elles ont été ajoutées.
2. Mettre en évidence les deux questions/sujets les plus critiques identifiés par le groupe.

Points de discussion potentiels entre les groupes :

- Quelles similitudes et différences peuvent être observées dans les processus ?
- Y avait-il des problèmes communs qui se sont présentés au sein des groupes ?

Compte-rendu



Passez un quiz sur les concepts que vous avez appris dans cette section.

1. Quel est l'ordre des cinq regroupements de processus PMBoK ?

- Planification, Initiation, Surveillance et Contrôle, Exécution, Clôture
- Initiation, Planification, Exécution, Surveillance et Contrôle, Clôture
- Initiation, Planification, Exécution, Clôture, Surveillance et Contrôle
- Initiation, Planification, Surveillance et Contrôle, Exécution, Clôture

2. Quels sont les types d'éléments à livrer ? (plusieurs réponses correctes)

- Déterminé
- Implicite
- Estimé
- Direct
- Indirect
- Supposé

3. Qu'est-ce qu'un goulot d'étranglement ?

- un blocage qui retarde le développement ou le progrès
- un espace où quelque chose ou quelqu'un manque
- un problème, ou une situation qui empêche quelqu'un de faire quelque chose, ou qui rend quelque chose impossible.

4. Quels sont les exemples de tâches de mobilisation ? (Plusieurs réponses correctes)

- Affiliation
- Publication
- Imagerie
- Géoréférencement
- Augmentation de la sensibilisation du public

Capture de données



Dans ce module, vous en apprendrez sur le concept de standards, en particulier le standard Darwin Core et ses composantes. Vous découvrirez également les types de données primaires sur la biodiversité et comment partager au mieux ces informations au sein du GBIF. Enfin, vous examinerez les principes de qualité des données dans le contexte de la saisie des données et vous en apprendrez davantage sur la qualité et la cohérence des données (en particulier sur des sujets tels que le géoréférencement, les dates, noms et vérification croisée des taxons).

Standards et Darwin Core



Dans cette vidéo (15:37), vous apprendrez comment vous interagissez avec les standards tous les jours. Puis vous serez introduits au **Standards de l'information sur la biodiversité**, y compris le **Standard Darwin Core** que vous continuerez à utiliser tout au long de ce cours. Si vous ne parvenez pas à regarder la vidéo intégrée, vous pouvez la **télécharger** localement (MP4 - 27 Mo)

▶ <https://www.youtube.com/watch?v=S02PJHPsRAs> (YouTube video)

Origines et types de données



Dans cette vidéo (10:45), vous allez passer en revue **les données primaires sur la biodiversité** qui peuvent être partagées au sein du GBIF. Si vous ne pouvez pas regarder la vidéo intégrée, vous pouvez la **télécharger** localement. (MP4 - 19 Mo)

▶ <https://www.youtube.com/watch?v=wKeOveydjsw> (YouTube video)

Questions

- Votre type de données est-il différent de ce que vous pensiez initialement ?
- Avec quel type de données travaillez-vous ?
- Comment publieriez-vous vos données sur le GBIF (en utilisant quel noyau et/ou extension) ?

Saisie, traitement et qualité des données



Dans cette vidéo (09:11), vous allez explorer les principes de la qualité des données appliqués à la saisie des données, spécifiquement lors de la capture de données à partir des étiquettes de collecte, des cahiers de terrain, des feuilles de calcul, etc. Si vous ne pouvez pas regarder la vidéo intégrée, vous pouvez la [télécharger](#) localement. (MP4 - 19 Mo)

▶ <https://www.youtube.com/watch?v=QkDJlkmwBMA> (YouTube video)

Exercice 2



Pour cette activité, vous allez effectuer un exercice simulant une saisie de données. Vous allez commencer à travailler avec [les termes Darwin Core](#) et choisirez quelles données sont nécessaires pour votre organisation/projet, puis vous examinerez lesquelles de ces données seront partagées plus tard lors de la publication.

Lire le [Cas Pratique I](#) (si vous ne l'avez pas déjà fait).

Imaginez que vous êtes la personne chargée de transcrire les données qui se trouvent sur les planches d'herbier.

1. Téléchargez [UC1-2-base-material.zip](#). (34,4 Mo). Il y a 10 images. Deux images par spécimen pour un total de cinq spécimens. Les planches d'herbier sont en espagnol (les données peuvent vous être transmises par divers moyens et dans d'autres langues que la vôtre), mais vous devriez être en mesure de reconnaître les informations comprises sur les étiquettes. N'oubliez pas d'utiliser les deux images par spécimen pour compiler les informations.
2. Téléchargez le modèle de feuille de calcul : [UC1-2-occurrence-template.xlsx](#) (57,3 ko) pour transcrire les informations trouvées sur chaque image pour les cinq spécimens.
3. Utilisez la feuille d'exercice téléchargée précédemment pour donner vos réponses.



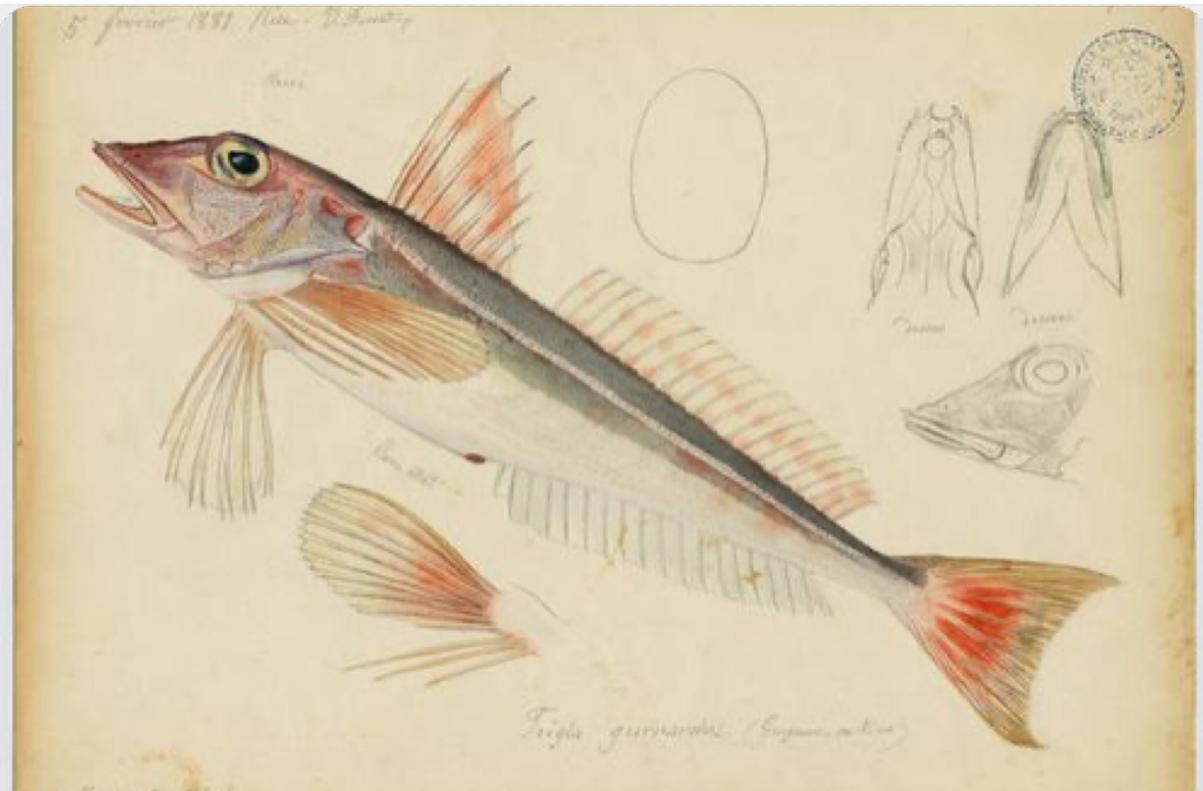
vous devrez peut-être ajouter des champs à la feuille de calcul, car vous pourrez peut-être capturer plus d'informations à partir des étiquettes que ce qui était prévu dans le modèle.

Compte-rendu



Passez un quiz sur les concepts que vous avez appris dans cette section.

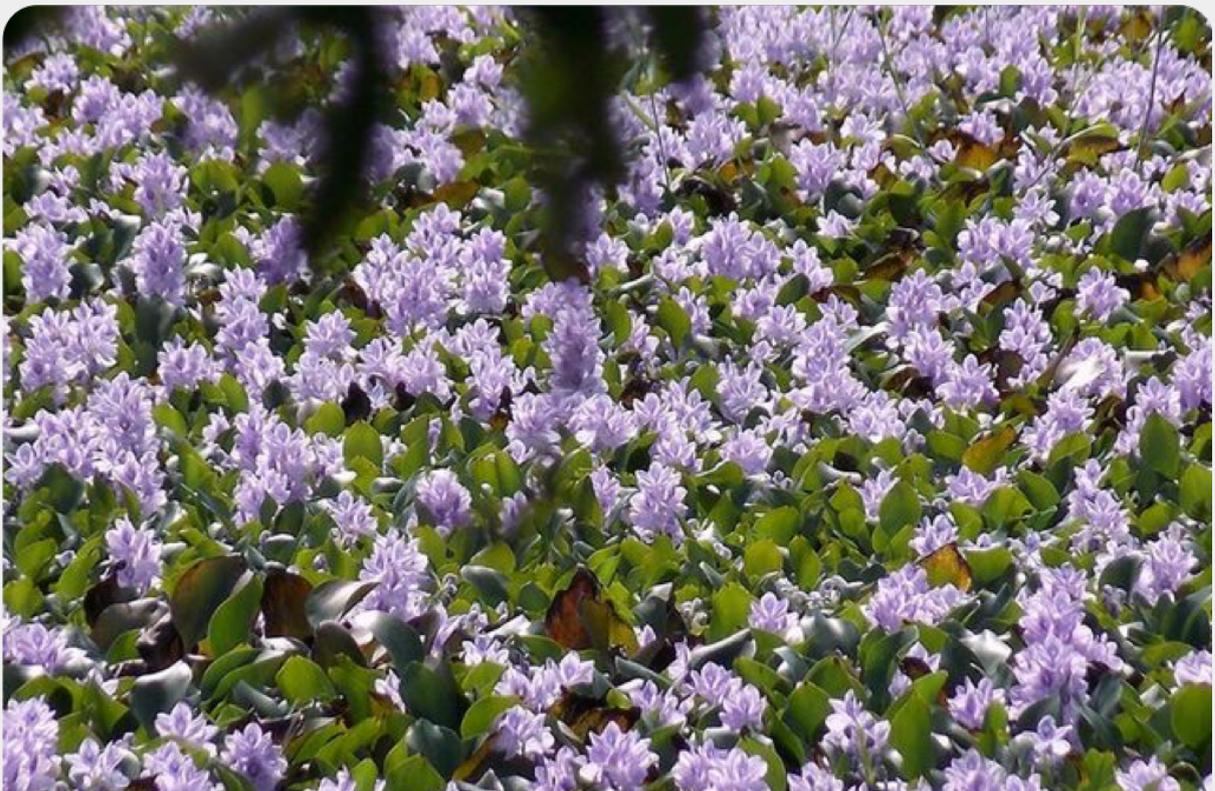
1. Quel(s) type(s) de jeu de données choisiriez-vous pour une collection ichthyologique ?



Eutrigla gurnardus (Linnaeus, 1758) / Muséum d'histoire naturelle de Nice

- occurrence
- liste taxonomique
- événement d'échantillonnage

2. Quel(s) type(s) de jeu de données choisiriez-vous pour une liste d'espèces envahissantes ?



Jacinthe d'eau (*Eichhornia crassipes*) observée à Bourail, Nouvelle-Calédonie, où il s'agit d'une espèce introduite et envahissante par GRIIS. Photo par Gérard (2016) sous licence CC BY-SA 2.0

- occurrence
- liste taxonomique
- événement d'échantillonnage

3. Quel(s) type(s) de jeu de données choisiriez-vous pour la flore et la faune d'une étude d'impact environnementale ?

Les études d'impact sur l'environnement sont réalisées par des experts afin d'évaluer la biodiversité et les biotopes d'une surface donnée, avant, pendant et après qu'elle soit affectée par des activités anthropiques (travaux routiers, éoliennes, mines, construction de bâtiments, etc.).



Entomologiste capturant des papillons par Matthieu Gauvain (CC-BY-SA)

- occurrence
- liste taxonomique
- événement d'échantillonnage

4. Quel(s) type(s) de jeu de données choisiriez-vous pour des données de suivi des oiseaux ?

Les données de suivi des oiseaux sont enregistrées à l'aide d'appareils spécifiques, tels que les balises GPS montées sur des oiseaux vivants, permettant ainsi aux scientifiques de suivre leurs routes migratoires ou leurs sites de reproduction.



Vautour Fauve observé à la Réserve Naturelle de Gamla par גיוזיג - MinoZig (CC0)

- occurrence
- liste taxonomique
- événement d'échantillonnage

5. Quel(s) type(s) de jeu de données choisiriez-vous pour des données de pièges à insectes ?



Piège à insectes par miheco (CC-BY-SA)

- occurrence

- liste taxonomique
- événement d'échantillonnage

6. Quel(s) type(s) de jeu de données choisiriez-vous pour des données de gestion d'un parc national ?

Les données acquises dans le cadre de la gestion des aires protégées (comme les parcs nationaux mais aussi les réserves naturelles plus petites) peuvent être diverses et avoir des origines différentes : relevés botaniques, suivi des animaux marqués, observations effectuées par les gestionnaires et les gardes, et même des données de «science citoyenne» déduites des images partagées sur les réseaux sociaux.



éléphants sri-lankais observés par pen_ash.

- occurrence
- liste taxonomique
- événement d'échantillonnage

7. Quel(s) type(s) de jeu de données choisiriez-vous pour un inventaire éclair (bioblitz) dans le cadre d'un programme de science participative ?

Les données de sciences participatives sont souvent recueillies par le biais de journées thématiques de travail sur le terrain, connues sous le nom de « bioblitz », ou inventaire éclair. Les bénévoles se rassemblent généralement dans une zone donnée et passent la journée à essayer d'observer et d'identifier autant d'espèces qu'ils le peuvent dans cette zone.

Les données de chaque participant sont saisies et agrégées dans l'outil de saisie ou de gestion des données du programme de science participative.



À la recherche d'oiseaux avec le personnel du parc par le US National Park Service (réutilisation autorisée sur Google Image search)

- occurrence
- liste taxonomique
- événement d'échantillonnage

8. Quel(s) type(s) de jeu de données choisiriez-vous pour une liste d'espèces régionale ?



Rhinocéros noir observé au Zoo de Magdebourg en Allemagne par Mani300

- occurrence
- liste taxonomique
- événement d'échantillonnage

Gestion de données



Dans ce module, vous passerez en revue les concepts principaux, les outils connexes et les meilleures pratiques pour la gestion des données, en particulier, le nettoyage et la normalisation des données.

Principes de gestion des données



Dans cette vidéo (09:49), vous allez passer en revue un ensemble important de principes nécessaires pour améliorer les données à travers les processus de nettoyage des données. Si vous ne pouvez pas regarder la vidéo intégrée, vous pouvez la [télécharger](#) localement. (MP4 - 16,6 Mo)

▶ <https://www.youtube.com/watch?v=4ijm1cJeVHE> (YouTube video)

Outils de gestion des données



Dans cette vidéo (06:42), vous en apprendrez plus une variété d'outils que vous pouvez utiliser pour améliorer la qualité de vos données. Si vous ne pouvez pas regarder la vidéo intégrée, vous pouvez la [télécharger](#) localement. (MP4 - 10.3 Mo)

▶ <https://www.youtube.com/watch?v=Ru3vWiYU3gw> (YouTube video)

Exercice 3a-c



Pour ces exercices, vous effectuerez des vérifications techniques et cohérentes, améliorerez les données avec différents outils et apprendrez à utiliser [OpenRefine](#).

Lire le [Cas Pratique I](#) (si vous ne l'avez pas déjà fait).

Votre institution fait partie de la « Global Poales Association (GPA) ». Cette association a obtenu un financement pour publier une flore à jour sur le groupe et a demandé à votre herbier de participer et de fournir des enregistrements de haute qualité que vous pourriez avoir sur cet ordre de plantes. Cet ordre est bien représenté dans votre collection et vous pensez pouvoir contribuer de manière substantielle à cet effort.

Exercice 3a

Contrôles de validation

Dans cet exercice, nous nous concentrerons sur les erreurs techniques et effectuerons un contrôle de validation de base pour identifier les erreurs techniques. Reportez-vous à [Contrôles de validation](#) pour plus d'informations sur les types d'erreurs.

1. Télécharger [UC1-3ab-data-cleaning.csv](#). (207,5 ko)
2. Importez le fichier CSV dans Excel en utilisant l'assistant Excel. Voir [Excel-tips-EN.pdf](#) (PDF, 7 Mo) pour les instructions d'importation de votre système d'exploitation (Windows, Mac, Linux).
3. Trouvez et corrigez les erreurs manuellement.
4. Utilisez la feuille d'exercice téléchargée précédemment pour donner vos réponses.

Exercice 3b

Autres outils de gestion des données

L'association GPA vous a donné une liste d'éléments de contrôle de qualité des données à vérifier :

- Tous les noms de plantes (nom complet) sont correctement orthographiés
- Tous les noms de plantes appartiennent à cet ordre
- Tous les enregistrements ont des coordonnées
- Toutes les coordonnées sont à l'intérieur du pays indiqué et converties au format décimal
- Toutes les dates sont dans la colonne appropriée et au format AAAA-MM-JJ

Les trois catégories d'erreurs sont :

- Erreurs de nomenclature
 - Erreurs de format
 - Erreurs géographiques/coordonnées aberrantes
1. Reportez-vous à [Outils utiles](#) pour terminer l'exercice. Vous n'êtes pas limité à ces outils, vous pouvez utiliser les outils que vous voulez.
 2. Utilisez le même fichier que pour l'exercice précédent.
 3. Effectuez la correction SEULEMENT pour la famille Eriocaulaceae (vous devrez sans doute filtrer les données)
 4. Corrigez les erreurs trouvées dans le jeu de données utilisé dans l'exercice 3a (exercice précédent), en utilisant les outils de votre choix et documentez les changements que vous effectuez dans la feuille d'exercice.
 5. Corrigez le fichier entier si vous avez le temps.
 6. Utilisez la feuille d'exercice téléchargée précédemment pour donner vos réponses.

Exercice 3c



Dans cette vidéo (03:27), vous en apprendrez plus sur [OpenRefine](#). Vous pouvez utiliser OpenRefine pour standardiser et améliorer la qualité de vos données. Si vous ne pouvez pas regarder la vidéo intégrée, vous pouvez la [télécharger](#) localement. (MP4 - 37,1 Mo)

▶ https://www.youtube.com/watch?v=_YFw_bfwc3Y (YouTube video)

OpenRefine

Dans cet exercice, nous utilisons OpenRefine pour améliorer la qualité d'un jeu de données grâce aux fonctionnalités par défaut du logiciel, des services web existants et des expressions régulières.

1. Lien de téléchargement [UC1-3c-open-refine.csv](#). (207,5 ko)
2. Téléchargez et complétez les exercices suivants [OpenRefine-Exercice3c-EN.pdf](#). (PDF, 1,1 Mo)
Aussi disponibles en [Français](#) et [Espagnol](#).
3. Utilisez la feuille d'exercice téléchargée précédemment pour donner vos réponses.

Conseils pour l'exercice

Contrôles de validation

Erreurs techniques Relativement simples, il est souvent possible d'automatiser les **vérifications liées à l'intégrité des données**. Celles-ci peuvent indiquer des problèmes d'exportations de fichier, de cartographie de données, de glissement de champ (par exemple le déplacement involontaire d'une colonne vers la droite) ou des données manquantes à la source.

- **Complétude** : Vérifiez si toutes les données et métadonnées sont disponibles – tous les champs sont-ils présents, tous les champs sont-ils remplis ?
- **Limites** : Par exemple, les jours sont-ils renseignés entre les bornes 1 et 31 (selon le mois) ?
- **Type de données** : Par exemple, le champ Date contient-il une date ou un nombre ?
- **Format des données** : Par exemple, les Dates sont-elles fournies sous la forme 01/01/2010 ou 01/Jan/10 ?

Erreurs de cohérence

Application de règles réelles aux données. Ces règles peuvent indiquer une saisie incorrecte de données provenant d'anciens enregistrements, d'erreurs de transcription ou de post-traitement. Certaines sont complexes à implémenter et **requièrent que les jeux de données de référence soient vérifiés**, par exemple une liste des collectionneurs connus et des habitudes de collecte. Ces règles peuvent être recueillies auprès des utilisateurs de données et des analystes.

- **Erreurs taxonomiques** : Par exemple, si une occurrence est identifiée au niveau de l'espèce, un nom scientifique et des entrées dans les champs du genre et de l'espèce ont-ils été fournis ?
- **Erreurs de dates** : Les dates de collecte, d'identification, de mise à jour et de numérisation sont-elles cohérentes ?
- **Valeurs aberrantes** : Détectez les valeurs aberrantes, mais n'oubliez pas que toutes les aberrations ne sont pas nécessairement des erreurs. Par exemple, comparez avec une répartition d'espèce connue, ou une aire environnementale connue (mais n'oubliez pas que les valeurs aberrantes peuvent être de fausses identifications, plutôt que des coordonnées incorrectes).
- **Erreurs géographiques** : Les coordonnées sont-elles dans la localité ou la région identifiée ? Par exemple, y a-t-il des occurrences terrestres dans la mer ou les occurrences marines sur terre ?
- **Protocoles de collecte** : Est-ce que le détail de l'occurrence correspond aux protocoles connus de collecte de l'organisation ou du collecteur ? Des enregistrements semblent-ils avoir été créés après la mort d'un collecteur (cela pourrait-il être un collecteur différent avec un nom similaire) ? Par exemple, des enregistrements de mammifères sont-ils attribués à une association d'observation d'oiseaux ?
- **Exactitude et précision** : Par exemple, y a-t-il des enregistrements géoréférencés indiquant une très grande exactitude ou précision alors qu'ils ont été recueillis avant l'utilisation des GPS ? (ou des GPS très précis)
- **Méthodes de collecte** : Les différentes méthodes d'arpentage (ex: transects et exploration de parcelles) ont des caractéristiques particulières. Les enregistrements sont-ils cohérents avec la

méthode fournie ?

Outils utiles

- **Outil GBIF de séquençage des noms scientifiques** : <https://www.gbif.org/tools/name-parser>
- **Global Names Resolver** : <http://resolver.globalnames.org>
- **Correspondance de nom de Catalogue of Life** : <https://data.catalogueoflife.org/tools/name-match>
- **TNRS** : <https://tnrs.biendata.org/>
- **WoRMS** : <https://www.marinespecies.org/aphia.php?p=match>
- **InfoXY** : <http://splink.cria.org.br/infoxy?criaLANG=en>
- **Calculateur de coordonnées géographiques** : <http://georeferencing.org/georefcalculator/gc.html>
- **Outil Canadensys de conversion des coordonnées** : <https://data.canadensys.net/tools/coordinates?lang=fr>
- **Outil Canadensys de séquençage des dates** : <https://data.canadensys.net/tools/dates?lang=fr>
- **Google Maps** : <https://maps.google.com/>

Compte-rendu



Passez un quiz sur les concepts que vous avez appris dans cette section.

1. Pourquoi est-il préférable de nettoyer vos données ?
 - pour les rendre aussi réutilisables que possible
 - pour atteindre vos objectifs de qualité des données
 - les données doivent être nettoyées par les utilisateurs, pas les fournisseurs
2. Comment devriez-vous organiser votre processus de nettoyage des données ?
 - travaillez seul, vous connaissez le mieux vos données
 - sollicitez vos collègues pour obtenir l'expertise qui vous manque
 - travaillez au niveau institutionnel pour harmoniser les processus concernant la qualité des données
3. Qu'est-ce qui est le mieux :
 - Empêcher les erreurs de se produire
 - Corriger les erreurs dès que vous les trouvez dans votre base de données ou dans votre feuille de calcul
 - Non pas nettoyer les erreurs mais les documenter au fur et à mesure que vous les repérez, afin que les personnes qui réutilisent vos données sachent où elles se trouvent
4. Qui est responsable de la qualité des données ?
 - La ou les personnes qui enregistrent les données sur le terrain

- Les transcripteurs de données
- Le gestionnaire de base de données
- Toutes les personnes impliquées dans la gestion des données
- Les personnes qui utilisent vos données
- GBIF

5. Quels outils peuvent être utilisés pour nettoyer vos données ?

- Excel & autres outils de gestion de feuilles de calcul
- OpenRefine
- Votre logiciel de base de données
- Des outils en ligne tels que Scientific Names Resolver ou Google Maps

Publication de données



Dans ce module, vous apprendrez des concepts sur la publication de données, incluant l'IPT, les cœurs (cores) et les extensions, et l'importance des licences, des métadonnées, des champs obligatoires et de l'hébergement des jeux de données.

Concepts de publication de données



Dans cette vidéo (11:45), vous vous familiariserez avec les concepts de publication de données et vous recevrez une introduction à l'IPT (IPT). Si vous ne parvenez pas à regarder la vidéo intégrée de, vous pouvez [la télécharger](#) sur votre ordinateur. (MP4 - 20 Mo)

► <https://www.youtube.com/watch?v=b900d9ukjSQ> (YouTube video)

Présentation générale de l'IPT



Dans cette vidéo (06:56), vous recevrez une vue d'ensemble de l'interface de publication des données IPT. Si vous ne pouvez pas regarder la vidéo intégrée, vous pouvez la [download](#) localement. (MP4 - 8,7 Mo)

► https://www.youtube.com/watch?v=gHXsaN_JWeI (YouTube video)

Installations IPT pour la formation

Si vous n'avez pas encore reçu d'identifiant, veuillez contacter training@gbif.org et vous recevrez un identifiant et un mot de passe pour vous connecter sur l'IPT de la formation.

<https://training-ipt-a.gbif.org/>

<https://training-ipt-b.gbif.org/>

<https://training-ipt-c.gbif.org/>

Démonstration de l'IPT



Dans cette vidéo (24:16), vous allez apprendre comment publier un jeu de donnée en utilisant un IPT. Si vous ne parvenez pas à regarder la vidéo intégrée Vimeo, vous pouvez [la télécharger](#) localement. (MP4 - 52,6 Mo)

▶ <https://www.youtube.com/watch?v=eDH9loTrMVE> (YouTube video)

Exercice 4



Dans cet exercice, vous publierez un jeu de données d'occurrence en utilisant l'IPT.

Lire le [Cas Pratique I](#) (si vous ne l'avez pas déjà fait).

Publication de données

Après le nettoyage des données dans le jeu de données des Poales, l'équipe considère que la publication des données en ligne via le réseau GBIF pourrait être un bon moyen de rendre cet effort visible. Il vous a été demandé de mener à bien ce travail de publication, basé sur ce jeu de données.

1. À ce stade, vous avez besoin d'un compte sur l'une des [course IPTs](#). Si vous n'avez pas encore reçu d'identifiant de connexion, veuillez contacter training@gbif.org et vous recevrez un identifiant et un mot de passe pour l'un des IPTs du cours.
2. Télécharger [UC1-4-poales-publishing.csv](#). (233,5 ko)
3. Utilisez l'installation IPT assignée et publiez ce fichier.
4. Utilisez la feuille d'exercice téléchargée précédemment pour donner vos réponses.

Compte-rendu



Passez un quiz sur les concepts que vous avez appris dans cette section.

1. Que signifie la publication des données, dans le contexte du GBIF ?
 - Exporter un fichier csv de vos données nettoyées que vous pouvez partager avec vos collègues
 - Écrire un article décrivant vos données, et le(s) protocole(s) que vous avez utilisé(s) pour les collecter, les retranscrire et les nettoyer
 - Rendre votre ou vos jeu(x) de données sur la biodiversité accessibles au public et visibles dans un format standardisé
2. Qu'est-ce qu'un IPT ?
 - un outil qui vous aide à gérer & corriger vos données
 - un outil qui vous aide à publier vos données sur le GBIF
 - un outil qui vous aide à produire un data paper (article de données)
3. Quelles licences ou dérogations Creative Commons sont recommandées par le GBIF pour la publication de données ?

CC-BY, CC-BY-SA et CC-BY-ND

CC0, CC-BY et CC-BY-NC

CC0, CC-BY et CC-BY-SA

4. Quels sont les trois noyaux (cores) que vous pouvez choisir pour une ressource sur l'IPT ?

Métadonnées, Occurrence, Multimédia

Taxon, Collection, MeasurementOrFact

Occurrence, Taxon, Événement d'échantillonnage

5. Combien de fichiers d'extensions un jeu de données peut-il avoir ?

zéro

un

autant que nécessaire

Evaluation et certification



Dans ce module, vous examinerez les critères à utiliser pour évaluer les devoirs et obtenir une certification.

Une fois le cours terminé avec succès et une évaluation réussie des devoirs (par des formateurs et des mentors), les participants ont la possibilité de recevoir un certificat officiel sous la forme d'un badge **Open Badge**.



Un score global de 2.5-2.9 permet de gagner un badge BASIC de mobilisation des données sur la biodiversité



Un score global de 3.0-4.0 permet d'obtenir un badge de mobilisation des données sur la biodiversité ADVANCED

Les participants doivent soumettre (en anglais) le Cas Pratique II et le Cas Pratique III (choix entre deux options). Chaque Cas Pratique est évalué en fonction des rubriques pédagogiques du cours. Ces rubriques définissent les compétences et les niveaux de performance sur lesquels reposent les objectifs d'apprentissage du cours.

Examinez les rubriques pour vous assurer de la bonne compréhension des compétences qui seront évaluées en vue de l'obtention de la certification.

Rubrique de planification

Planification

Compétences	Performance de départ 1	Performance en développement 2	Performance atteinte 3	Performance exceptionnelle 4
A. Connaissance des différents éléments qui font partie d'un plan de mobilisation de données de biodiversité	Comprend quelques rôles et tâches nécessaires à un plan de mobilisation solide, mais n'est pas en mesure de différencier qui effectue quelle tâche dans le plan ou mélange les rôles.	Comprend un grand nombre des rôles et tâches nécessaires à un plan de mobilisation, mais ne parvient pas à établir les liens et les interactions entre eux.	Comprend les tâches et les rôles clés nécessaires dans un plan de mobilisation et leurs interactions.	Peut identifier les rôles et les tâches supplémentaires qui peuvent être nécessaires dans des situations spécifiques.
B. Capacité à appliquer les différents éléments d'un plan de mobilisation des données de biodiversité à un contexte institutionnel donné (par exemple, le sien).	Peut voir comment quelques tâches et rôles se rapportent à un contexte institutionnel donné, mais éprouve des difficultés lorsque le contexte change même légèrement par rapport aux références utilisées.	Peut voir comment la plupart des tâches et des rôles peuvent être appliqués à un contexte institutionnel donné, mais a encore des lacunes lorsque la situation est différente des références utilisées.	Peut identifier toutes les tâches et tous les rôles pertinents et voir comment ils peuvent être appliqués à un contexte institutionnel donné, même si la situation n'est pas identique à celle des références utilisées.	Capable d'appliquer les différents éléments d'un plan de mobilisation de manière créative, en incorporant de nouveaux éléments qui ne se trouvent pas dans les références utilisées.
C. Capacité à rédiger/composer un document de planification clair.	Peut rassembler des éléments bruts qui pourraient servir de base à un plan (par exemple sous forme de points), mais il a des difficultés à les traduire en texte.	Peut construire un plan de travail de base, mais manque de cohérence entre ses éléments.	Peut écrire un document cohérent et bien articulé qui comprend tous les éléments pertinents d'un plan de travail.	Peut écrire des plans concis, avec des résumés efficaces et des niveaux de détail croissants.

Compétences	Performance de départ 1	Performance en développement 2	Performance atteinte 3	Performance exceptionnelle 4
D. Capacité à évaluer un plan de mobilisation de données de biodiversité	Peut identifier quelques éléments d'un plan de mobilisation dans un exemple donné, mais a des difficultés à les associer aux tâches et aux rôles.	Peut identifier la plupart des éléments d'un plan de mobilisation et les associer aux tâches et aux rôles lorsqu'ils sont identiques aux références utilisées. A des difficultés à évaluer la faisabilité de certains éléments du plan.	Peut reconnaître tous les éléments existants dans un plan de mobilisation, et peut identifier les lacunes, les efforts redondants et les incohérences. Peut évaluer la qualité des différents éléments. Peut évaluer les points de succès et les points faibles potentiels du plan dans son ensemble.	Peut suggérer des solutions pour résoudre les problèmes identifiés dans le plan évalué.

Rubrique de saisie de données

Saisie de données

Compétences	Performance de départ 1	Performance en développement 2	Performance atteinte 3	Performance exceptionnelle 4
A. Capacité à identifier les types de données numériques qui peuvent être extraites d'une source de données sur la biodiversité (c'est-à-dire qui peuvent être publiées en utilisant le réseau GBIF)	Peut identifier uniquement les types de données les plus évidents des sources de données de biodiversité les plus communes (p. ex. les occurrences des spécimens de collections d'histoire naturelle). Montre une faible compréhension du potentiel de publication en ligne à l'aide du GBIF.	Peut fréquemment identifier correctement au moins un type de données numériques pouvant être extraites de sources de données communes. Il a des difficultés à identifier celles qui peuvent être publiées à l'aide du GBIF.	Peut toujours identifier un (ou plusieurs) types de données numériques qui peuvent être extraites de sources de données communes. Peut identifier ceux qui peuvent être publiés à l'aide du GBIF.	Peut toujours identifier un ou plusieurs types de données numériques qui peuvent être extraites de sources de données communes et inhabituelles. Il peut identifier parmi ces types ceux qui peuvent être publiés à l'aide du GBIF et lesquels sont en cours de discussion. Peut identifier les cœurs de données et les extensions utilisés pour la publication de ces types de données.

Compétences	Performance de départ 1	Performance en développement 2	Performance atteinte 3	Performance exceptionnelle 4
B. Capacité à extraire les informations pertinentes d'une source de données sur la biodiversité vers des structures de données simples (par exemple, des feuilles de calcul) qui respectent les standards internationaux	Peut extraire uniquement des grands éléments d'information évidents (par exemple, toutes les informations géographiques en tant qu'unité unique) qui sont évidents dans la source de données. Montre une faible connaissance des standards courants pour l'enregistrement de données de biodiversité.	Peut extraire plusieurs éléments d'information de la source de données (mais pas tous) et peut les désagréger en éléments significatifs. Montre une certaine connaissance de base des standards les plus courants (par exemple DwC) et des champs de données les plus utilisés dans ces standards.	Peut identifier toutes les informations utiles dans une source de données et extraire les éléments obligatoires dans une structure de données standard (par exemple, une feuille de calcul basée sur DwC simple). Peut identifier les informations manquantes et en déduire depuis les informations existantes (par exemple, déduire le nom d'un pays à partir d'une province).	Peut identifier toutes les informations utiles dans une source de données complexe, et les diviser en éléments significatifs directement traduits en standards internationaux. Peut identifier les informations capitales manquantes dans la source et les déduire à partir des données existantes ou d'informations supplémentaires sur la source (métadonnées).

Compétences	Performance de départ 1	Performance en développement 2	Performance atteinte 3	Performance exceptionnelle 4
C. Capacité à comprendre et à appliquer les principes de base de la qualité des données au processus de saisie des données	Ne comprend pas complètement comment l'application de principes simples de qualité des données peut avoir un impact important sur le produit final, en évitant un nettoyage supplémentaire nécessaire par la suite.	Connaît certains des principes les plus généraux de la qualité des données (par exemple, éviter les fautes de frappe), mais ne sait pas vraiment comment appliquer des principes plus spécifiques au processus de saisie des données.	Connaît tous les principes de base de la qualité des données et sait comment les appliquer de manière simple au processus de capture des données. Utilise les formats de manière cohérente au cours du processus de capture des données (par exemple, dans les dates, les noms de pays). Documente de manière simple toutes les procédures et modifications liées à la qualité des données.	Fait preuve d'une bonne connaissance de tous les principes courants de la qualité des données et de la manière de les utiliser pour améliorer le processus de capture des données. Utilise les formats de données de manière cohérente et peut utiliser des répertoires géographiques, des listes de référence ou des fonctions spécifiques du logiciel pour améliorer la qualité par rapport à l'original. Documente clairement tous les changements et toutes les décisions prises en rapport avec la qualité des données.

Rubrique de gestion des données

Gestion des données

Compétences	Performance de début 1	Développement de la performance 2	Performance accomplie 3	Performance exceptionnelle 4
A. Capacité à évaluer la qualité (c.-à-d. Identifier les problèmes et leurs types) d'un jeu de données sur la biodiversité.	Utilise seulement des contrôles visuels pour analyser la qualité. Incapable de différencier les types d'erreurs. Peut détecter les valeurs manquantes dans les champs obligatoires et les incohérences graves des données.	Ne peut utiliser que des techniques très basiques (par exemple le tri) pour analyser la qualité des données. Peut détecter les incohérences entre les noms de champs et le contenu. Peut identifier de manière cohérente les erreurs techniques, mais uniquement les erreurs de cohérence les plus courantes dans un jeu de données.	Peut utiliser des outils et des techniques spécifiques pour évaluer la qualité. Reconnaît le niveau minimum de désagrégation/normalisation nécessaire pour une utilisation et une publication courantes. Peut identifier de manière cohérente les erreurs techniques et la plupart des erreurs de cohérence dans un jeu de données.	Utilise une approche systématique de l'analyse des jeux de données couvrant tous les principaux domaines de données. Peut identifier de manière cohérente les erreurs techniques et de cohérence dans un jeu de données. Peut utiliser d'autres sources de données (par exemple des métadonnées ou d'autres jeux de données) pour identifier ou déduire des erreurs de cohérence dans un jeu de données.
B. Capacité à effectuer la correction du format des données.	Ne peut apporter des corrections que manuellement dans les tableaux. Montre des connaissances génériques sur l'utilisation des types de format dans les données numériques (par exemple, dates, chaînes, nombres)	Peut identifier au moins un outil spécifique pour corriger automatiquement les erreurs de format, mais ne peut l'utiliser que dans des cas spécifiques. Sinon, utilise des mécanismes simples (par exemple, «rechercher et remplacer») pour résoudre les problèmes.	Peut utiliser au moins un outil pour corriger automatiquement les erreurs de format.	Peut utiliser les fonctionnalités avancées de plusieurs outils pour corriger les erreurs de format.

Compétences	Performance de début 1	Développement de la performance 2	Performance accomplie 3	Performance exceptionnelle 4
C. Capacité à effectuer des corrections de données nomenclaturales.	Ne peut apporter des corrections que manuellement dans les tableaux. Utilise uniquement les connaissances personnelles des groupes taxonomiques connus.	Peut identifier au moins un outil spécifique pour corriger automatiquement les erreurs de nomenclature, mais ne peut l'utiliser que dans des cas spécifiques. Sinon, utilise des mécanismes simples (par exemple, «rechercher et remplacer») pour résoudre les problèmes.	Peut utiliser au moins un outil pour corriger automatiquement les erreurs de nomenclature. Peut trouver et utiliser des informations nomenclaturales de référence appropriées pour les groupes taxonomiques avec lesquels il travaille habituellement.	Peut utiliser plusieurs outils pour corriger les erreurs de nomenclature. Peut trouver et utiliser des informations nomenclaturales de référence appropriées pour des groupes taxonomiques en dehors de ses domaines d'expertise.
D. Capacité à effectuer la correction des données géographiques.	Ne peut apporter des corrections que manuellement dans les tableaux. N'utilise que la connaissance personnelle des zones géographiques connues.	Peut identifier au moins un outil spécifique pour cartographier et/ou corriger automatiquement les erreurs dans les informations géographiques, mais ne peut l'utiliser que dans des cas spécifiques. Sinon, utilise des mécanismes simples (par exemple, «rechercher et remplacer») pour résoudre les problèmes.	Peut utiliser au moins un outil pour cartographier et/ou corriger automatiquement les erreurs dans les informations géographiques. Peut trouver et utiliser des informations géographiques de référence appropriées dans un format adapté aux zones avec lesquelles il travaille habituellement.	Peut utiliser plusieurs outils pour cartographier et/ou corriger automatiquement les erreurs dans les informations géographiques. Peut trouver et utiliser des informations géographiques de référence dans un format approprié pour des domaines en dehors de ses domaines d'expertise.

Compétences	Performance de début 1	Développement de la performance 2	Performance accomplie 3	Performance exceptionnelle 4
E. Capacité à utiliser des logiciels spécifiques (par exemple OpenRefine) comme outils de nettoyage des données.	Peut identifier au moins un outil de nettoyage de données. Peut identifier les principales fonctionnalités d'un outil de nettoyage de données (par exemple OpenRefine).	Peut identifier plusieurs outils de nettoyage de données. Peut utiliser une ou quelques-unes des fonctionnalités de base d'un logiciel de nettoyage de données pour nettoyer un ensemble de données (par exemple, créer un projet OpenRefine, utiliser la facette, le filtrage, le regroupement ou la réconciliation).	Peut utiliser toutes les fonctionnalités de base d'un logiciel de nettoyage de données pour nettoyer un ensemble de données (par exemple dans OpenRefine: facetage, filtrage, regroupement, réconciliation).	Peut utiliser les fonctionnalités avancées d'un ou plusieurs logiciels de nettoyage de données pour nettoyer les ensembles de données (par exemple dans OpenRefine: utiliser l'API, les expressions régulières, Langue d'Expression de Google Refine).
F. Capacité à documenter les procédures de transformation des données.	Décrit rarement les modifications apportées lors de la conservation, du formatage ou de la transformation des données.	Décrit les modifications effectuées la plupart du temps. Ne décrit pas les changements de manière cohérente ou complète (par exemple, décrit le changement, mais pas l'auteur).	Se souvient toujours de décrire les modifications apportées. Décrit toujours les modifications de manière cohérente, de sorte que toutes les modifications du même type puissent être facilement identifiées.	Peut décrire avec précision et cohérence les modifications apportées de manière répétable.

Rubrique de publication des données

Publication de données

Compétences	Performance de début 1	Développement de la Performance 2	Performance accomplie 3	Performance exceptionnelle 4
A. Connaissance des standards de données d'information sur la biodiversité (BDI).	Montre une connaissance limitée ou inexistante des standards de données BDI et des standards de données acceptées par le GBIF.	Peut identifier les standards BDI et sait lesquelles sont acceptées par le GBIF, mais ne sait pas où trouver des informations sur la façon de les utiliser. Impossible d'identifier les termes obligatoires.	Connaît les normes BDI acceptées par le GBIF. Peut trouver une liste des cœurs de données et des extensions acceptés. Publie des jeux de données selon les standards GBIF requis et/ou recommandés pour les termes de données et de métadonnées et sait comment trouver les définitions des termes.	Montre une compréhension des caractéristiques et des limites des différents standards BDI.
B. Capacité d'analyser la pertinence d'un jeu de données sur la biodiversité pour la publication via le GBIF.	Montre une connaissance limitée ou inexistante des critères formels auxquels un jeu de données doit répondre pour être publiable via le GBIF.	Connaît les critères formels auxquels un jeu de données doit répondre pour être publiable via le GBIF, mais ne peut pas évaluer si un jeu de données donné les satisfait.	Peut évaluer correctement si un ensemble de données peut être actuellement publié via le GBIF. Peut attribuer au moins un type de données valide (= cœur) à un ensemble de données en fonction de la description fournie par le détenteur des données et après avoir analysé l'ensemble de données.	Peut identifier plusieurs options de publication pour un ensemble de données (si possible).

Compétences	Performance de début 1	Développement de la Performance 2	Performance accomplie 3	Performance exceptionnelle 4
C. Utilisation de l'IPT: capacité à produire / analyser des métadonnées de haute qualité.	Montre une connaissance limitée ou inexistante des caractéristiques des bonnes métadonnées.	Connaît les caractéristiques de bonnes métadonnées, mais a du mal à les reconnaître.	Connaît les caractéristiques des bonnes métadonnées et comment les reconnaître. Peut produire des recommandations sur la façon d'améliorer les métadonnées existantes.	Connaît les caractéristiques des métadonnées de haute qualité et comment les produire.
D. IPT use: capacity to upload/connect data and map it to existing cores & extensions.	Utilisation de l'IPT: capacité à télécharger / connecter des données et à les mapper aux cœurs et extensions existants.	Peut uniquement télécharger des ensembles de données à un seul fichier dans IPT et mapper vers un seul type de cœur sans extensions.	Peut télécharger plusieurs fichiers dans un IPT dans le cadre d'un seul jeu de données et les mapper correctement à un cœur et au moins une extension. Peut utiliser la fonction de valeur constante IPT.	Peut télécharger plusieurs fichiers dans un IPT dans le cadre d'un seul jeu de données et les mapper correctement à un cœur et à plusieurs extensions. Peut utiliser la fonction de traduction de données IPT.
E. Utilisation de l'IPT: capacité à utiliser l'outil pour publier et enregistrer des jeux de données.	Peut afficher un ensemble de données publié et les métadonnées associées sur un IPT. Peut télécharger un fichier DwC-A à partir d'un IPT. Peut naviguer dans un ensemble de données enregistré de l'IPT vers le portail GBIF.	Peut mettre à jour un ensemble de données publié existant en téléchargeant un nouveau fichier source. Peut republier le fichier, sans erreur.	Peut publier et enregistrer avec succès un nouveau jeu de données. Peut comprendre et agir sur des messages d'erreur dans l'IPT.	Montre une compréhension de la gestion des versions des jeux de données dans l'IPT.

Cas Pratique II - Espèces envahissantes



Familiarisez-vous avec le scénario du cas d'utilisation.

Vous avez le choix entre deux scénarios pour le cas pratique II :

- Liste de vérification des espèces envahissantes
- Échantillonnage de Lepidoptera

Votre choix pour le Cas Pratique II sera noté.

Scénario

Suivi des espèces envahissantes



Leucaena leucocephala (Lam) de Wit observé à Hawaii par Sharon Grant (sous licence CC-BY-NC 4.0)

Ce récit a été élaboré comme base pour des exercices pratiques du cours de mobilisation des données sur la biodiversité. Le concept et le contenu de l'exercice ont été développés par Sharon Grant, John Wieczorek, David Bloom et Laura Anne Russell.

Il s'agit d'un scénario fictif, basé sur un jeu de données réel, dont le but est de servir de support pédagogique. Le jeu de données original est attribué à Simpson A (2016). Big Island Invasive Species Committee - Pest Reports - 2005-2010. Version 4.1. United States Geological Survey. [Jeu de données d'occurrences](#) accessibles via GBIF.org le 2017-07-13.

Description

Le Conseil sur les espèces invasives de Hawaï (Hawaii Invasive Species Council : HISC) a reçu une subvention fédérale pour collaborer avec des écoles secondaires dans le cadre du programme national d'Hawaï sur les espèces invasives. Le but du projet est d'accroître les connaissances locales sur les espèces invasives, accroître la collecte de données et générer des listes d'espèces annotées pour les zones signalées. Un gestionnaire est employé à temps plein pour superviser le projet. Tous les fonds et les allocations sont gérés par l'Associé Fiscal du HISC.

Le directeur du Comité des espèces envahissantes (CSI) de chaque île a reçu un sous-prix pour la mise en place de programmes éducatifs locaux et la collecte de données. Les programmes ont formé des élèves du secondaire à devenir des mentors d'élèves et à faciliter la collecte d'images et de données par les membres de la communauté locale. Le Comité des espèces envahissantes de la Grande Île (CIIS) a reçu un nouveau sous-prix pour étendre sa base de données centrale afin de tenir

compte des données de chaque ISC, de fournir aux écoles participantes leurs propres sites Internet et de mettre à jour un portail de données consultable unique pour servir le gouvernement, le public et les efforts de recherche scientifique.

Deux écoles de chaque île ont été sélectionnées parce qu'elles étaient situées dans des zones où les connaissances et la documentation relatives à l'évaluation des espèces envahissantes étaient médiocres ou inexistantes. Les enseignants ont travaillé avec le responsable à la sensibilisation du Conseil des espèces envahissantes (CSI) local pour créer du matériel pédagogique détaillant 21 espèces de plantes envahissantes, y compris comment identifier les étapes de vie de chaque espèce et les méthodes de contrôle les plus efficaces.

L'école supérieure de l'Université d'Hawaï à Maui (UHM) organise un cours de sensibilisation auprès du grand public. Quatre étudiants de l'Université, dans le cadre de leur évaluation finale, sont en train de valider les identifications sur base photographique, et les descriptions soumises par chaque école secondaire à leur SIC local.

Collecte de données

Des élèves de chaque école secondaire ont organisé une série d'enquêtes sur une journée dans leur quartier. Les participants, guidés par les techniciens locaux de l'ISC et les mentors des étudiants, ont visité différents endroits où des guides photo leur ont été donnés et où un itinéraire à suivre pour la collecte de données leur a été attribué. Le long de chaque itinéraire, ils ont été chargés d'identifier les espèces cibles et de prendre 1 à 3 photos à l'aide de téléphones portables équipés d'un GPS.

Les informations, décrivant les observations de chacune des 21 espèces invasives ciblées, ont été enregistrés à l'aide d'un formulaire numérique de collecte de données au cours de chaque collecte. Les participants ont mis en ligne les images capturées via leurs téléphones portables et ont été encouragés à cliquer sur leur localisation en utilisant une carte Google, incorporée dans le formulaire, pour attribuer une latitude et une longitude à chaque observation. La conception du formulaire était basée sur le formulaire de déclaration des ravageurs du HISC.

REPORTER INFORMATION

Report Number:	<input type="text" value="T07222017172139"/>		
First Name:	<input type="text"/>	Last Name:	<input type="text"/>
Email:	<input type="text" value="e.g. pestreport@643pest.com"/>	Phone:	<input type="text" value="e.g. 808-123-4567 or 8081234567"/>

PEST SIGHTING INFORMATION

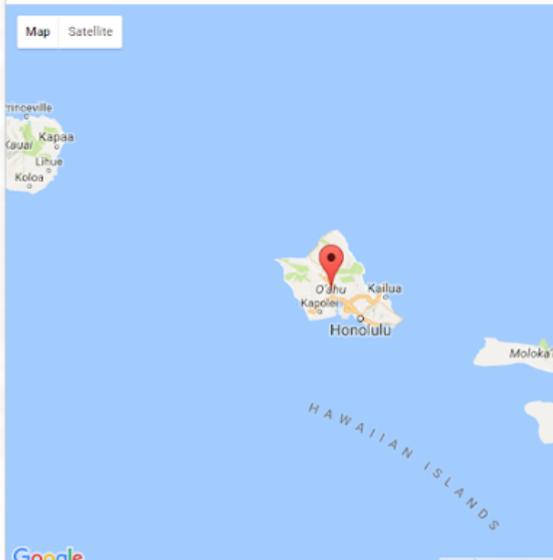
NOTE: Asterisk [*] and red label color indicate a required field.

*Date of Pest Sighting: <input type="text"/>	*Island of Pest Sighting: (Please choose an island before entering information into the Location field). <input type="text" value="--Select Island--"/>
*Pest Name: <input type="text"/>	Location of Pest Sighting: (Street address, cross streets, city, mile marker, place name or general area) <input type="text"/>
*Pest Description: (Plant: size; flower, foliage, or fruit color / scent / orientation; habitat) (Insect / animal: size; color; plant / host found on or nearby; habitat) <input type="text"/>	Additional Comments: <input type="text"/>

Image Upload (Allowed File Types: .jpg, .png, .gif | Upload up to 3 images):

Drag & drop files here ...

PEST SIGHTING LOCATION DETAILS

	Directions: <ol style="list-style-type: none">The map will have automatically moved to the island chosen from the Island drop-down list above. If an island was not chosen or the incorrect island was chosen, please go back and make a new selection. Choosing an island is a mandatory step.Use the map tools at left (Zoom In, Zoom Out, Pan) for help finding and zooming in to a desired area or location. Use the Geo-location Search box below to search for and zoom in to a specific geo-location (address, city, place name) in Hawaii.To pinpoint the exact location of the pest sighting: a) click on and drag the red map marker to a particular location; or b) click on any particular location on the map to move the red map marker to that location. The coordinates (Latitude, Longitude) submitted with the pest report are shown in the boxes below and reflect the final position of the red map marker.				
	Geo-location Search: <input type="text" value="Enter an address, city, place name..."/>				
	Marker Coordinates: <table><tr><td><input type="text" value="Latitude"/></td><td><input type="text"/></td><td><input type="text" value="Longitude"/></td><td><input type="text"/></td></tr></table>	<input type="text" value="Latitude"/>	<input type="text"/>	<input type="text" value="Longitude"/>	<input type="text"/>
<input type="text" value="Latitude"/>	<input type="text"/>	<input type="text" value="Longitude"/>	<input type="text"/>		

Description des données numériques

Une base de données, créée et hébergée par le service informatique de UHM, contient les images et les données du formulaire en ligne, mais ces données ne sont pas accessibles publiquement. Les données ont été exportées en tant que fichiers aux valeurs séparées par des virgules (.csv), et envoyées aux quatre étudiants diplômés de l'UHM pour validation taxonomique à l'aide des images et des descriptions soumises. L'analyste SIG du BIISC a utilisé les coordonnées de Google Maps et les données d'image EXIF pour vérifier la qualité des observations et ajouter les coordonnées géographiques manquantes. Les mentors étudiants ont renommé tous les fichiers image pour qu'ils correspondent au numéro d'observation, afin de faciliter le référencement croisé ultérieur par le BIISC.

Feuille d'exercice des espèces envahissantes

Télécharger le [fichier d'exercice](#). (MS Word, 342 ko)

Exercice 1

Planification

Vous êtes le gestionnaire local de l'ISC et, en raison du succès des cours et des enquêtes auprès du grand public, 10 écoles de plus sur votre île aimeraient mettre en place leurs propres projets l'année suivante. Vous aimeriez les y accompagner, mais le financement de votre ISC expirera à la fin de cette année. Le HISC a indiqué qu'il se montrera favorable à une demande de subvention réduite pour étendre vos programmes au cours de l'année suivante, et le BIISC a offert de l'aide.

Exercice 1a

Analyser les implications financières de l'augmentation du nombre d'écoles

1. Évaluez les options suivantes pour augmenter le nombre d'écoles participantes. Vous ne pouvez sélectionner que DEUX de ces options, donc vous devez choisir judicieusement.
2. Utilisez le fichier d'exercices pour fournir vos réponses.

Options

1. Payer des stagiaires d'été supplémentaires pour travailler au comité ISC local afin de coordonner les enquêtes .
2. Offrir un soutien financier au BIISC pour mettre en place des sites Web pour chaque nouvelle école.
3. Offrir une rémunération financière aux étudiants diplômés. Vous ne serez pas en mesure de leur verser à tous les quatre l'équivalent d'un salaire régulier, mais pourrez couvrir les coûts de postes à temps partiel pour deux d'entre eux.
4. Etablir un contrat avec une société d'informatique pour construire une base de données pouvant ingérer les données automatiquement à partir du formulaire en ligne. Le système inclura une interface d'administration pour permettre la manipulation de données et les exportations de fichiers csv.
5. Financer quatre activités de sensibilisation du public (par exemple un inventaire éclair ou BioBlitz) pour promouvoir la sensibilisation du grand public et accroître la participation des bénévoles.
6. Préparer et effectuer une formation réutilisable pour les enseignants des écoles afin de leur enseigner comment préparer les données à soumettre au BIISC.

Exercice 1b

Assigner des rôles

Les personnes suivantes sont disponibles dans le nouveau projet pour le traitement et la mobilisation des données.

1. Attribuer des rôles pour maximiser l'efficacité du traitement et de la transformation des données, afin de produire des données de la meilleure qualité aussi efficacement que possible ?
2. Utilisez le fichier d'exercices pour fournir vos réponses.

Rôles

- Analyste SIG du BIISC : capacité avancée d'utilisation des ordinateurs, SIG et des outils d'analyse des données.
- Gestionnaire ISC : bonnes compétences en informatique.
- Responsable de la communication de l'ISC : bonnes compétences d'identification sur le terrain, compétences de base en informatique. Expert en réseaux sociaux.
- Mentor étudiant 1 : connaissances taxonomiques de base. Compétences de base en informatique.
- Mentor étudiant 2 : connaissances taxonomiques de base. Compétences de base en informatique.
- Etudiant en botanique 1 : connaissances taxonomiques avancées, compétences en programmation.
- Étudiant en botanique 2 : connaissances taxonomiques avancées.
- Étudiant en botanique 3 : connaissances taxonomiques avancées.
- Étudiant en botanique 4 : connaissances taxonomiques avancées.

Exercice 2

Saisie de données

Le BIISC prévoit maintenant de rendre accessibles et publiques toutes les données du projet en publiant des jeux de données sur le GBIF. En tant qu'assistant de sensibilisation du BIISC, vous devez identifier les champs Darwin Core pertinents correspondant le mieux aux données du formulaire en ligne. Vous avez remarqué que des données supplémentaires décrivant les espèces et les lieux ont été ajoutées au formulaire de données par les étudiants diplômés qui ont procédé à leur validation. Vous devez étendre la structure de données afin d'intégrer ces informations à partir du formulaire en ligne, avec les données taxonomiques et géographiques supplémentaires.

1. Télécharger [UC2-IS-2-ForCapture.csv](#). (7,1 Mo)
2. En utilisant le jeu de données téléchargé, créez une feuille de calcul comme exemple de la structure de données étendue et des champs que vous avez identifiés comme étant pertinents pour la correspondance avec le Darwin Core.
3. Utilisez la feuille d'exercice pour fournir vos réponses et soumettre la feuille de calcul.

Exercice 3

Gestion des données

Au cours de l'été, les stagiaires du bureau principal du HISC ont créé des listes de référence

taxonomique à partir des données d'occurrence originales recueillies et complétées à partir du formulaire en ligne. En assumant le rôle de chef de projet HSC, vous devez maintenant effectuer les contrôles de qualité finaux avant la publication.

1. Téléchargez [UC2-IS-3-ForCleaning.xlsx](#). (156 ko)
2. Évaluez le jeu de données et identifiez quels types d'erreurs il contient.
3. Identifiez différentes façons de corriger ces problèmes et effectuez ces corrections pour autant d'erreurs que possible.
4. Utilisez la feuille d'exercice pour fournir vos réponses et soumettre la feuille de calcul.

Exercice 4

Publication de données

Le HISC est maintenant prêt à publier les données de la liste de référence taxonomique et les occurrences associées sur le GBIF. Pour cet exercice, vous assumerez le rôle de responsable du projet. Votre responsabilité : publier les données de la liste de référence nettoyée et les occurrences associées en ligne via le réseau GBIF.

1. Téléchargez [UC2-IS-4-ForPublication.xlsx](#). (99 ko)
2. Utilisez l'installation IPT indiquée précédemment pour publier le jeu de données fourni.
3. Utilisez la feuille d'exercice pour fournir vos réponses, ainsi qu'un lien vers le jeu de données publié.

Cas Pratique II - Observations de Lépidoptères



Familiarisez-vous avec le scénario du cas d'utilisation.

Vous avez le choix entre deux scénarios pour le cas pratique II :

- Liste de vérification des espèces envahissantes
- Échantillonnage de Lepidoptera

Votre choix pour le Cas Pratique II sera noté.

Scénario

Échantillonnage de Lepidoptera à travers différents pays



Papilio machaon Linnaeus, 1758 observé en Israël par רניו רמוע (Licence sous CC-BY-NC 4.0)

Ce scénario a été élaboré comme base pour les exercices pratiques du cours de mobilisation des données de biodiversité. Le concept et le contenu de l'exercice ont été développés par Alberto González-Talaván, sur la base des travaux antérieurs d'Alberto González-Talaván, Danny Vélez, Larissa Smirnova, Laura Russell, Mélianie Raymond et Nicolas Noé. Il s'agit d'un scénario fictif qui n'est destiné qu'à des fins pédagogiques.

Description

L'International Butterfly Amateur Network (IBAN) fournit un cadre pour les groupes d'observation amateurs nationaux pour capturer des données sur l'apparition de papillons (Lepidoptera) depuis 2009. Un vaste réseau d'observateurs amateurs utilise un protocole standard basé sur la méthode "Pollard walk" pour saisir ces renseignements sur les feuilles de papier qu'ils transmettent à leur bureau national. Certains de ces bureaux numérisent ces informations dans des feuilles de calcul, mais d'autres n'ont pas les ressources humaines pour le faire et ils envoient les enregistrements papier à l'IBAN pour traitement. L'IBAN produit un rapport annuel basé sur les observations fournies par ces membres nationaux, avec la mise à jour des cartes de distribution et l'analyse des tendances démographiques de certaines espèces clés.

Le siège de l'IBAN est principalement composé de volontaires. Avec la popularité croissante de la science citoyenne et l'intérêt général pour les papillons en tant que groupe d'organismes charismatique, de plus en plus de données sont reçues chaque année et les fiches de données non numérisées s'accumulent rapidement. Le comité directeur de l'IBAN tente d'identifier un flux de travail plus efficace et plus agile pour la création de données numériques, car il souhaite commencer à publier ces données en ligne régulièrement. Ils aimeraient également commencer à traiter des images numériques que leurs bénévoles sont déjà en train de capturer avec des appareils mobiles comme des téléphones et des tablettes. Leur objectif ultime est de faire connaître le réseau et de

renforcer les collaborations avec les gouvernements locaux et régionaux pour influencer les politiques de conservation de Lepidoptera dans les pays concernés.

Il n'existe actuellement aucun accord formel entre l'IBAN et les amateurs encodant des données, comme par exemple pour couvrir la manière dont les données peuvent être utilisées. Le comité de pilotage a quelques craintes sur le fait qu'ils devront formaliser cet accord lorsqu'ils commenceront à publier les données en ligne.

Collecte de données

Le protocole recommandé – les marches Pollard – est basé sur des transects de 300 à 600 m de longueur, divisées en sections de 50 m. Chaque transect doit couvrir un seul type d'habitat.

Lors de chaque visite, les opérateurs doivent compter toutes les espèces de lépidoptères qui peuvent être observées à moins de 5 m du transect. Les comportements particuliers (ponte d'œufs ou butinage), ainsi que le stade de développement (par exemple, des larves ou des œufs), devraient également être enregistrés.

Dans la plupart des pays, ces échantillonnages ont lieu une fois toutes les deux semaines, de début octobre à fin juin.

Il y a des mesures de contrôle de la qualité en place : chaque enregistrement signalé est marqué "En attente d'approbation". Le statut de l'enregistrement n'est changé en « Approuvé » qu'après vérification par un expert taxonomique désigné. Les espèces repérées hors de leur saison régulière ou de leur zone de distribution sont signalées pour vérification supplémentaire.

L'heure de la journée et les conditions météorologiques sont enregistrées au début du transect. Le long de ce transect, le nombre d'individus de chaque espèce observée est comptabilisé. Les espèces non identifiées sont comptabilisées et enregistrées, soit au niveau de la famille, soit comme un complexe prédéfini de deux ou trois espèces similaires. Les papillons vus en dehors de la plage de 5 mètres sont enregistrés comme "Extra+le numéro de la section la plus proche" (ex. 5-extra). L'heure de fin du transect est également enregistrée.

Exemple de capture de données analogiques

DATA CAPTURE SHEET					
Recorder:	Hadas Lebruder				
Recorder ID:	IBAN 1002				
Date:	19/10/2012				
Data sheet nr.	0129				
Transect nr.	tr 029				
Transect length:	175m				
Start time:	11:45 am				
End time:	12:17 pm				
Location description	Eilatot forest				
Latitude (start)	32,29309				
Longitude (start)	34,89637				
Latitude (end)	-				
Longitude (end)	-				
Temperature	28 °C				
Weather	Sunny, clear sky. No wind.				
Section	Lat	Long	Start Time	Length	Notes
1	32,29309	34,89637	11:45	25	
2	-	-	11:49	25	
3	-	-	11:54	25	
4	-	-	11:57	25	
5	-	-	12:02	25	
6	-	-	12:08	25	
7	-	-	12:12	25	



Species	Nr.	Section	Time	Distance	Notes
<i>Lampides boeticus</i>	1	1	11:47	2	
<i>Gegenes pumilio</i>	1	3	11:55	1	
<i>L. boeticus</i>	1	3	11:58	2.5	
<i>Pieris brassicae</i>	1	5	12:02	0.5	Nectaring
<i>Colias croceus</i>	1	7-ext	12:13	10	
<i>G. pumilio</i>	1	7	12:13	4	
<i>Lycæna thesarmon</i>	2	7	12:14	2.5	

Description des données numériques

Certains bureaux nationaux utilisent des groupes de bénévoles pour numériser les archives papier et produire des tableaux numériques. Les tableaux sont très simples et comprennent trois fiches de données. La première capture les informations liées aux efforts d'échantillonnage, la seconde les

conditions météorologiques et la troisième les espèces rencontrées et le nombre d'individus observés par l'amateur.

#	Date	Transect	Time From	To	Species Name	1	1e	2	2e	3	3e	4	4e	5	5e	6	6e	7	7e	8	8e	9	9e	10	10e	11	11e	12	12e		
17	Mar,11 2015	TR027	11:20	12:00	Pieris brassicae																	1									
18	Mar,11 2015	TR027	11:20	12:00	Pieris rapae							3	2	1								1	2	2							
19	Mar,11 2015	TR027	11:20	12:00	Complex 13-14							2											2								
Gilad yaar habanim						Abundance/Richness: 17 / 3																									
20	Mar,9 2015	TR006	09:45	10:30	Archon apollinus	1																	1						3		
21	Mar,9 2015	TR006	09:45	10:30	Pieris rapae	2																									
22	Mar,9 2015	TR006	09:45	10:30	Pontia daplidice																							1			
23	Mar,9 2015	TR006	09:45	10:30	Anthocharis cardamines																	1									
24	Mar,9 2015	TR006	09:45	10:30	Gonepteryx cleopatra																								4		
Carmel Hurshan haarbaim						Abundance/Richness: 13 / 5																									
25	Mar,8 2015	TR007	10:13	10:35	Archon apollinus																				1						
26	Mar,8 2015	TR007	10:13	10:35	Pieris brassicae								1						1												
27	Mar,8 2015	TR007	10:13	10:35	Gonepteryx cleopatra																	2				2					
28	Mar,8 2015	TR007	10:13	10:35	Lasiommata megera emilyssa																							1			
Kibutz Sasa						Abundance/Richness: 8 / 4																									
29	Mar,7 2015	TR054	12:00	12:30	Papilio machaon									3								1		1							
30	Mar,7 2015	TR054	12:00	12:30	Anthocharis cardamines							2	1															1	1		
31	Mar,7 2015	TR054	12:00	12:30	Vanessa atalanta																	1									
32	Mar,7 2015	TR054	12:00	12:30	Complex 10-12	3																	4	3		1		1			
Nachshonim Kakal forest						Abundance/Richness: 43 / 4																									

eventId	scientificName	individualCount	quantity	quantityType	recordedBy	Approved
1000-tr010-s00	Lepidoptera	0	0	individuals	Zvika Avni	Approved
1001-tr011-s1	Carcharodus alceae	1	0.004	individuals	Viki Soroker	forApproval
1001-tr011-s1	Lycaenidae	3	0.012	individuals	Viki Soroker	Approved
1001-tr011-s11	Pieridae	2	0.008	individuals	Viki Soroker	Approved
1001-tr011-s12	Leptotes pirithous	2	0.008	individuals	Viki Soroker	Approved
1001-tr011-s2	Carcharodus alceae	1	0.004	individuals	Viki Soroker	Approved
1001-tr011-s4	Pieris rapae	3	0.012	individuals	Viki Soroker	Approved
1001-tr011-s6	Azanius jesus	1	0.004	individuals	Viki Soroker	Approved
1001-tr011-s7	Pieridae	1	0.004	individuals	Viki Soroker	Approved
1001-tr011-s7	Pieris rapae	1	0.004	individuals	Viki Soroker	Approved
1001-tr011-s8	Leptotes pirithous	1	0.004	individuals	Viki Soroker	Approved
1002-tr029-s1	Lampides boeticus	1	0.004	individuals	Hadas Lebrider	Approved
1002-tr029-s3	Gegenes pumilio	1	0.004	individuals	Hadas Lebrider	Approved
1002-tr029-s3	Lampides boeticus	1	0.004	individuals	Hadas Lebrider	Approved
1002-tr029-s5	Pieris brassicae	1	0.004	individuals	Hadas Lebrider	forApproval
1002-tr029-s7	Colias croceus	1	0.004	individuals	Hadas Lebrider	Approved
1002-tr029-s7	Gegenes pumilio	1	0.004	individuals	Hadas Lebrider	Approved
1002-tr029-s7	Lycaena thersamon	2	0.008	individuals	Hadas Lebrider	Approved

Feuille d'exercice Lepidoptera

Télécharger le fichier d'exercice. (MS Word, 342 ko)

Exercice 1

Planification

Le volume de données analogiques (registre papier) arrivant au siège de l'IBAN dépassera bientôt leur capacité de numérisation, le comité directeur a donc décidé de reconsidérer l'approche actuelle dans ce domaine de leur travail qui s'est développé sans gestion au cours des dernières années. À ce jour, c'est ainsi que le travail a été organisé :

- Les registres papier arrive par courrier. Le secrétaire ouvre les paquets et rassemble les registres.
- Il y a cinq bénévoles avec des compétences informatiques de base utilisant deux ordinateurs partagés pour numériser les registres papier. Ces bénévoles sont aussi des citoyens en science participative eux-mêmes, donc ils sont familiers avec la taxonomie de l'ordre Lepidoptera, et avec les espèces présentes dans le pays où se trouve le siège de l'IBAN.
- Les numériseurs arrivent et partent chaque fois qu'ils ont le temps de sorte qu'ils vérifient habituellement la disponibilité de l'ordinateur par téléphone. Parfois, il y a des conflits de temps et certains doivent rentrer chez eux car les deux ordinateurs sont occupés, et parfois les deux ordinateurs sont inutilisés.
- Lorsqu'ils numérisent, ils choisissent généralement un registre en papier à la fois à partir d'une pile et le numérisent (s'ils le peuvent). Les problèmes courants qui se posent sont les suivants :
 - le numériseur ne connaît pas l'espèce observée (des fautes d'orthographe arrivent),
 - le numériseur ne connaît pas la zone où l'échantillonnage a eu lieu,
 - le numériseur ne peut pas lire l'écriture manuscrite ou la langue dans laquelle certains des commentaires sont écrits.
- Une seule experte taxonomique obtient tous les tableaux numérisés et produit les rapports et les cartes de distribution à partir d'eux. Normalement, elle doit se débarrasser d'environ 15 % des données numérisées en raison d'incohérences, de fautes d'orthographe ou d'autres erreurs qu'elle n'a pas le temps de vérifier.

Exercice 1a

Analyser la composante financière de leur nouveau plan de numérisation

Le comité directeur analyse les options suivantes pour leur nouveau plan de numérisation, qui ont toutes des répercussions financières sur leur budget déjà réduit. Ils savent qu'ils ne peuvent mettre en œuvre que DEUX de ces options, donc ils doivent choisir judicieusement. Utilisez la feuille d'exercice pour recommander les options qu'ils doivent sélectionner et expliquer pourquoi vous les avez choisis.

1. Option 1 : Achetez trois ordinateurs supplémentaires pour que tous les numériseurs puissent travailler simultanément.
2. Option 2 : Offrez un soutien financier aux bureaux nationaux pour acheter des scanners à plat et envoyer/partager les registres électroniquement plutôt que par courrier.
3. Offrir une rémunération financière aux numériseurs. Vous ne serez pas en mesure de leur verser à tous les cinq l'équivalent d'un salaire de base, mais pourrez couvrir les coûts de postes à temps partiel pour trois d'entre eux.
4. Option 4 : Achetez le logiciel existant de numérisation de la biodiversité en anglais, qui est fourni avec vérification d'entrée taxonomique et aides intégrées pour corriger les informations géographiques.
5. Option 5 : Contratez une société de développement de logiciels pour développer des logiciels de numérisation personnalisés. Pour le même prix que le logiciel commercial, les développeurs fourniront une solution dans la langue locale, qui correspondra parfaitement au schéma de données d'origine et fournira également un portail web de données pour exposer les résultats de

l'effort de numérisation.

6. Option 6 : Organiser un cours pour les cinq numérisateurs pour améliorer leurs compétences en taxonomie, en utilisation informatique et en standard informatique de la biodiversité.

Exercice 1b

Assigner des rôles

Ceci sont les ressources humaines disponibles pour cet effort de numérisation. Comment assigneriez-vous les rôles pour maximiser l'efficacité du processus de numérisation et produire des données de la plus haute qualité possible ? Utilisez la feuille d'exercice pour fournir vos réponses.

1. Assistant administratif. Aucune connaissance taxonomique. Utilisation de base de l'ordinateur. Peut lire 3 langues.
2. Bénévole 1. Connaissance taxonomique de base. Utilisation de base de l'ordinateur.
3. Bénévole 2. Connaissance taxonomique de base. Utilisation de base de l'ordinateur.
4. Bénévole 3. Connaissance taxonomique de base. Utilisation de base de l'ordinateur. Peut lire 3 langues.
5. Bénévole 4. Connaissance taxonomique de base. Utilisation de base de l'ordinateur. Peut lire 3 langues.
6. Bénévole 5. Connaissance taxonomique de base. Utilisation avancée de l'ordinateur (y compris les SIG et les outils d'analyse de données).
7. Expert en taxonomie. Connaissance taxonomique avancée. Utilisation avancée de l'ordinateur (y compris les SIG et les outils d'analyse de données).

Exercice 2

Saisie de données

Imaginez que vous êtes un des bénévoles numérisant les archives papier reçues au siège de l'IBAN. Vous avez reçu deux archives papier.

1. Télécharger les journaux 1 et 2 [UC2-LS-2-ForCapture.zip](#). (943 ko)
2. Quelle structure de données utiliseriez-vous pour refléter les données dans ces journaux d'enregistrements ?
3. Créez une feuille de calcul en utilisant cette structure et les données des journaux.
4. Use the exercise sheet to provide your answers and submit the spreadsheet created in the previous step.

Exercice 3

Gestion des données

En assumant le rôle de l'un des bénévoles possédant des compétences informatiques avancées, imaginez qu'il vous a été assigné la responsabilité des questions de qualité des données. Votre tâche principale est de réduire la quantité de données actuellement rejetées (environ 15 %) avant traitement en raison d'erreurs et d'incohérences. Vous avez reçu un jeu de données en tant que produit brut de l'effort de numérisation.

1. Télécharger [UC2-LS-3-ForCleaning.xlsx](#). (44 ko)
2. Évaluez le jeu de données et identifiez quels types d'erreurs il contient.
3. Identifiez différentes façons de corriger ces problèmes, et effectuez ces corrections pour autant d'erreurs présentes que possible.
4. Utilisez la feuille d'exercice pour fournir vos réponses et soumettre la feuille de calcul.

Exercice 4

Publication de données

Pour cet exercice, vous prendrez le rôle de l'expert taxonomique qui collabore avec l'IBAN à son siège. Certaines de vos responsabilités précédentes (rédaction du rapport annuel, et production des cartes de distribution de base) ont été transmises aux volontaires, et une nouvelle responsabilité vous a été confiée : publier les données nettoyées en ligne via le réseau GBIF. Le volontaire responsable de la qualité des données a fourni un jeu de données à publier.

1. Télécharger [UC2-LS-4-ForPublication.xlsx](#). (58 ko)
2. Utilisez l'installation IPT indiquée précédemment pour publier le jeu de données fourni.
3. Utilisez la feuille d'exercice pour fournir vos réponses, ainsi qu'un lien vers le jeu de données publié.

Cas pratique III - Oiseaux dans la littérature

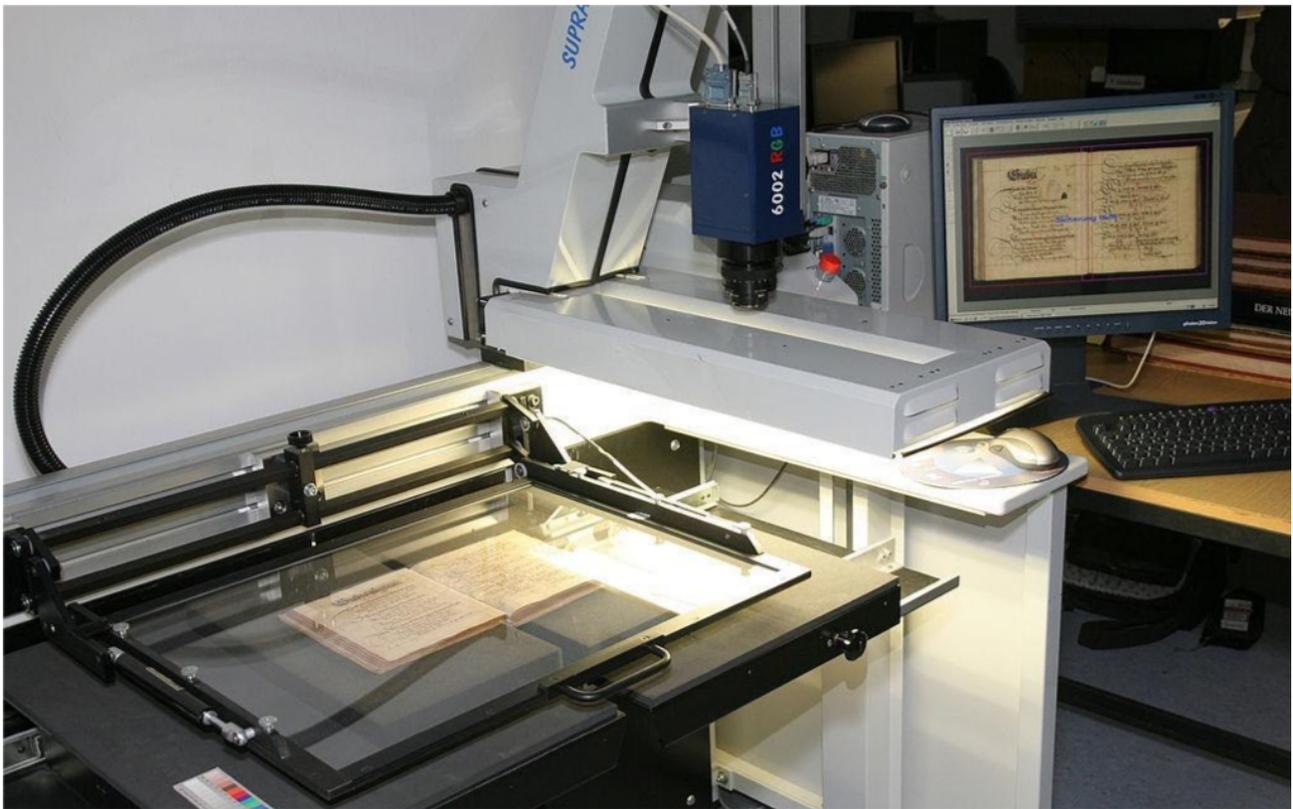


Familiarisez-vous avec le scénario du cas d'utilisation.

Le cas pratique III sera noté.

Scénario

Projet de mobilisation de données de littérature « Oiseaux découverts dans les phares danois, 1883-1939 »



Scanner haute résolution pour le projet de numérisation de livres par Heiko Hornig (sous licence CC BY-SA 2.5)

Ce scénario a été élaboré comme base pour les exercices pratiques du cours de mobilisation des données de biodiversité. Le concept et le contenu de l'exercice ont été développés par Alberto González-Talaván, sur la base des travaux antérieurs d'Alberto González-Talaván, Danny Vélez, Larissa Smirnova, Laura Russell, Mélianie Raymond et Nicolas Noé. Il s'agit d'un scénario fictif qui n'est destiné qu'à des fins pédagogiques.

Il s'agit d'un scénario fictif basé sur un projet et un jeu de données véritables et qui est destiné uniquement à des fins d'instruction. Le [projet original](#) et le [jeu de données original](#) sont attribués au nœud GBIF danois, [DanBIF](#).

Description

Le Musée d'Histoire Naturelle du Danemark (NHM-DK) est un centre de recherche associé à l'Université de Copenhague. Leur bibliothèque est membre de l'association nationale des bibliothèques qui a récemment reçu des fonds de l'État pour mettre en ligne les ressources détenues par ses membres. Le NHM-DK aimerait commencer à numériser les cahiers de terrain, des publications et des livres contenus dans leur bibliothèque, dont certaines ont une valeur historique significative.

Après une courte consultation avec leurs partenaires réguliers, NHM-DK a reçu une suggestion du chef du bureau de gestion du Parc national du Nordjylland. Ils aimeraient que le contenu d'une compilation littéraire classique particulière soit numérisé pour un projet qu'ils exécutent : "Oiseaux aux phares danois, 1883-1939 (En danois, « Fuglene ved de danske Fyr, 1883-1939 »). Ils veulent utiliser toutes les données d'occurrence enregistrées dans ces livres à partir de deux phares (Lodbjerg Fyr et Hanstholm Fyr) pour un projet d'exposition sur place.

Le NHM-DK a entamé des discussions avec leur nœud national GBIF, DanBIF, au sujet de la mobilisation des informations contenues dans ces volumes, à savoir préserver leur contenu pour le futur et fournir un accès en ligne à tout le monde. Avec la participation de DanBIF, il y a l'intention de

publier et d'enregistrer les données extraites avec le GBIF. Comme le GBIF nécessite une licence pour toutes les données publiées, le musée a décidé de publier les données avec une licence Creative Commons permettant l'utilisation des données avec attribution (CC-BY).

Les services informatiques requis sont fournis par l'Unité Technologique de l'Université de Copenhague, comme pour tous les projets numériques de musées.

Le directeur adjoint de HHM-DK, qui coordonne ce travail a développé un plan général pour le travail :

1. Le musée effectuera la numérisation de la littérature à l'aide de deux membres du personnel de la bibliothèque formés à l'utilisation du scanner de bibliothèque pour scanner des volumes délicats. Ils extraient également du texte des scans à l'aide d'un logiciel OCR (Optical Character Recognition).
2. Trois bénévoles de la Société Ornithologique de Copenhague (COS) qui collaborent régulièrement avec le musée et connaissent les oiseaux de la région ont été engagés pour aider et compléteront le transfert des données des PDF scannés au format tableur. Ils devront se rendre au musée et utiliser les ordinateurs disponibles dans la bibliothèque pour accéder aux fichiers stockés dans l'intranet du musée (réseau privé).
3. Le curateur Ornithologique dans le Département des Oiseaux du NHM-DK dirigera l'équipe responsable de la vérification taxonomique, de la conservation des données, du nettoyage, formatage et de la transformation. Il supervisera également l'entrée des métadonnées pour le jeu de données publié. L'équipe comprend un chercheur collaborateur originaire de Suède et deux étudiants postdoctoraux. Ils ont été sélectionnés pour cette tâche parce qu'ils sont habitués à travailler avec des données numériques sur la biodiversité. Ils utiliseront tous leurs propres ordinateurs de travail.
4. Le gestionnaire du Nœud DanBIF s'assurera que l'institution est correctement enregistrée dans GBIF en tant que fournisseur de données et que le directeur adjoint et le conservateur de la collection ornithologique disposent des références appropriées et d'un accès à l'instance IPT de DanBIF pour télécharger et publier les données.

Collection de données originale

Dans la période 1883-1939, 45 phares et bateaux-phares étaient actifs au Danemark. Ces phares ont été utilisés par plusieurs espèces d'oiseaux durant la période de migration des oiseaux entre 1886 et 1939. La présence et les activités de ces oiseaux ont été enregistrées, surtout par les gardiens de ces phares qui ont aussi recueilli des spécimens qui ont été envoyés au musée à Copenhague. Ces oiseaux ont été soigneusement préservés et catalogués par les responsables de collections du musée et les spécimens y sont encore présents aujourd'hui. Des observations sur les conditions météorologiques durant les nuits où les oiseaux ont été observés par les gardiens ont également été documentées.

Description des données analogiques

Ceci est un exemple de description d'une série d'observations d'espèces provenant d'un livre (en allemand, à l'exception du nom commun de l'espèce qui est fournie en danois).

2. *Cerchneis tinnunculus*, Linn. (*Taarnfalk*).

nistet in mehreren Kirchthürmen in der Gegend von **Viborg**. An den Kirchen von **Mönsted** und **Daubjerg** gab es im Jahre 1880 mehrere Sippen, jede von 7 Stücken, was auf einen besonders reichlichen Futtervorrath auf den umliegenden grossen Haiden (Eidechsen) schliessen lässt. In der Gegend von **Thisted**, wo der Thurmfalke sehr häufig brütet, waren die Jungen stets zu 6 vorhanden, nur ein einziges Mal fand ich auf **Egebäksande** eine Sippschaft von 7. (H.)

Bei **Horsens** habe ich den Thurmfalken an den Kirchen von **Vähr**, **Hansted**, **Hundslund**, **Thyrsted** und **Oelsted** brütend gefunden. In dem Kirchthurme von **Vähr** fand ich ein Nest mit 6 Eiern 6. Mai 1875, 6 Eier 30. April 1876, 4 Eier 28. April 1878; 1877 und 1879 waren keine da, 1880 hatten sie 5 Eier am 2. Mai, 1882 ebenfalls 5 Eier und 1883 sieben.

In **Oelsted** 5 Eier am 3. Mai 1879, 1882 ebenfalls 5 und 1883 6 Eier.

Auf den Kirchen (Frauen- und Petri-) von **Kopenhagen** habe ich ihn oft im Herbst und Winter 1879 gesehen; im Frühjahr 1880 brütete er auf der Frauenkirche, dem »runden Thurm« und dem Kanzleigebäude (in einem Ventil) und gewiss auch auf der Petrikerche. (F.)

Description des données analysées et traduites

Ceci est un exemple de la sortie numérisée et traduite à partir de l'exemple analogique ci-dessus.

Output of the OCR software	Translation into English
<p>2. <i>Cerchneis tinnunculus</i>, Linn. («Tarnfalk») nistet in mehreren Kirchthürmen in der Gegend von Viborg.</p> <p>An den Kirchen von Mönsted und Daubjerg gab es im Jahre 1880 mehrere Sippen, jede von 7 Stücken, was auf einen besonders reichlichen Futtermorrath auf den umliegenden grossen Haiden (Eidechsen) schliessen lässt. In der Gegend von Thisted, wo der Thurmfalke sehr häufig brütet, waren die Jungen stets zu 6 vorhanden, nur ein einziges Mal fand ich auf Egebäksande eine Sippschaft von 7. (H.)</p> <p>Bei Horsens habe ich den Thurmfalken an den Kirchen von Vähr, Hansted, Hundslund, Thyrsted und Oelsted brütend gefunden. In dem Kirchthurme von Vähr fand ich ein Nest mit 6 Eiern 6. Mai 1875, 6 Eier 30. April 1876, 4 Eier 28. April 1878; 1877 und 1879 waren keine da, 1880 hatten sie 5 Eier am 2. Mai, 1882 ebenfalls 5 Eier und 1883 sieben.</p> <p>In Oelsted 5 Eier am 3. Mai 1879, 1882 ebenfalls 5 und 1883 6 Eier.</p> <p>Auf den Kirchen (Frauen- und Petri-) von Kopenhagen habe ich ihn oft im Herbst und Winter 1879 gesehen; im Frühjahr 1880 brütete er auf der Frauenkirche, dem »runden Thurm« und dem Kanzleigebäude (in einem Ventil) und gewiss auch auf der Petrikerche. (F.)</p>	<p>2. <i>Cerchneis tinnunculus</i>, Linn. ("Tarnfalk") nests in several steeples around Viborg.</p> <p>At the churches of Mönsted and Daubjerg there were several family groups in 1880, each of 7 individuals, suggesting a particularly abundant source of food on the surrounding heather (lizards). In the area of Thisted, where the tower falcon broods very often, young were always present in broods of 6, only once did I find a group of 7 on Egebäksande (H.)</p> <p>In Horsens I found kestrels brooding on the churches of Vähr, Hansted, Hundslund, Thyrsted and Oelsted. In the steeple of Vähr I found a nest with 6 eggs on 6 May 1875, 6 eggs on 30 April 1876, 4 eggs on 28 April 1878; in 1877 and 1879 there were none, on 2 May 1880 they had 5 eggs, in 1882 also 5 eggs, and in 1883 seven.</p> <p>In Oelsted 5 eggs on 3 May 1879, in 1882 also 5, and in 1883 6 eggs.</p> <p>On the churches (Our Lady's and St. Peter's) of Copenhagen, I have often seen it in the autumn and winter of 1879; in spring 1880, it brooded on Our Lady's church, the "round tower" and the law firm building (in a valve) and certainly also on St. Peter's Church. (F.)</p>

Description des données numériques

En étudiant l'extrait du livre, les volontaires de la Société d'Ornithologie de Copenhague suggèrent d'extraire les données suivantes du texte numérisé et traduit :

- Nom scientifique tel qu'il apparaît dans le livre
- Nom(s) commun(s) en danois comme apparaissant dans le livre
- Localité
- Année/mois/jour
- Nombre d'individus observés
- Sexe
- Etape du cycle de vie
- Remarques
- URL de la page numérisée du livre dont l'occurrence provient

Feuille d'exercices sur les oiseaux de la littérature

Télécharger la [fiche d'exercices](#). (MS Word, 342 ko)

Exercice 1

Planification

L'équipe doit développer un flux de travail durable pour numériser les ressources littéraires, extraire toute information précieuse sur la biodiversité et la publier en ligne via GBIF. Ils doivent élaborer un

plan qui pourra être maintenu à l'avenir une fois que le financement de l'association nationale des bibliothèques sera terminé.

La section **scénario** de ce cas pratique comprend une brève description du flux de travail conçu par le directeur adjoint. Sur la base du flux de travail et du texte d'accompagnement, complétez ce qui suit :

1. Identifier les différentes parties prenantes participant à ce projet
2. Identifiez leur affiliation et attribuez chacun d'eux à un groupe d'intervenants
3. Identifier les rôles qui leur sont associés et assigner les tâches pour lesquelles ils sont actuellement responsables
4. Effectuer une analyse critique du flux de travail, identifier les risques et les lacunes potentiels, et proposer des moyens d'améliorer le flux de travail, de maximiser l'efficacité du projet de numérisation et de produire des données de la plus haute qualité possible.
5. Utilisez le fichier d'exercices pour fournir vos réponses.

Exercice 2

Saisie de données

Les scans et la reconnaissance de caractères (OCR) des livres sont terminés. Les données sur les occurrences doivent maintenant être extraites de ces sources et compilées dans un format de feuille de calcul.

Les données originales étaient en allemand et, afin de les rendre plus largement utilisables lors de leur publication en ligne, le responsable du projet souhaite les rendre disponibles en anglais.

2. Cerchneis tinnunculus, Linn. ("Taarnfalk") nests in several steeples around Viborg.

At the churches of Mönsted and Daubjerg there were several family groups in 1880, each of 7 individuals, suggesting a particularly abundant source of food on the surrounding heather (lizards). In the area of Thisted, where the tower falcon broods very often, young were always present in broods of 6, only once did I find a group of 7 on Egebåksande (H.)

In Horsens I found kestrels brooding on the churches of Vähr, Hansted, Hundslund, Thyrted and Oelsted. In the steeple of Vähr I found a nest with 6 eggs on 6 May 1875, 6 eggs on 30 April 1876, 4 eggs on 28 April 1878; in 1877 and 1879 there were none, on 2 May 1880 they had 5 eggs, in 1882 also 5 eggs, and in 1883 seven.

In Oelsted 5 eggs on 3 May 1879, in 1882 also 5, and in 1883 6 eggs.

On the churches (Our Lady's and St. Peter's) of Copenhagen, I have often seen it in the autumn and winter of 1879; in spring 1880, it brooded on Our Lady's church, the "round tower" and the law firm building (in a valve) and certainly also on St. Peter's Church. (F.)

1. Jouez le rôle d'un volontaire chargé de transformer le texte traduit en occurrences individuelles dans une feuille de calcul. Des numéros uniques devront être attribués à ces occurrences.
2. Créez une feuille de calcul en utilisant les champs de données énumérés dans la **Description des données numériques** en utilisant les données trouvées dans l'exemple ci-dessus et enregistrées par : Chr. Fr. Lütken.
3. Use the exercise sheet to provide your answers and submit the spreadsheet created in the

previous step.



Dans les exemples utilisés, les occurrences individuelles ne contiennent pas toujours les données pour compléter toutes les colonnes de la feuille de calcul.

Exercice 3

Gestion des données

Les données ont maintenant été compilées dans un format de feuille de calcul par les bénévoles de la Société ornithologique de Copenhague. En tant que conservateur de la collection ornithologique au département des oiseaux, vous êtes responsable de la qualité des données du jeu de données.

Grâce au géoréférencement rétrospectif, des coordonnées ont été ajoutées au jeu de données ainsi que la localité, mais aucune autre information géographique à une échelle supérieure. Étant donné que toutes les observations ont été faites au Danemark, le continent et le pays peuvent facilement être ajoutés. De plus, seul le nom scientifique a été fourni. Une taxonomie supérieure peut être dérivée en utilisant des outils logiciels tels qu'OpenRefine. Vous savez aussi qu'il y a des erreurs typographiques qui ont été faites par les numériseurs.

1. Télécharger [UC3-DL-3-ForCleaning.zip](#). (45 ko)
2. Identifier et corriger toutes les années non valides.
3. Vérifier et corriger la taxonomie.
4. Vérifier que les coordonnées sont correctes pour les deux localités données. Corrigez celles qui ne le sont pas. Les coordonnées doivent être au format décimal.
5. Ajouter des données pour les éléments manquants qui peuvent être dérivés en utilisant les données disponibles
6. N'oubliez pas de conserver les informations originales fournies et de documenter vos modifications et assomptions au niveau des enregistrements individuels et des métadonnées.
7. Use the exercise sheet to provide your answers and submit the cleaned text file extracted from the step 1.



Le jeu de données ne doit contenir que les années 1883-1939

Exercice 4

Publication de données

Pour cet exercice, vous assumerez le rôle de la personne responsable de la publication des données nettoyées en ligne via le réseau GBIF. Vous avez reçu un fichier multimédia et un fichier d'historique d'identification qui devrait être publié en même temps que les observations. Le membre du personnel responsable de la qualité des données vous a fourni des jeux de données nettoyés pour que vous puissiez les publier.

1. Télécharger [UC3-DL-4-ForPublication.zip](#). (65 ko)
2. Utilisez l'installation IPT indiquée précédemment pour publier le jeu de données fourni.
3. Utilisez la feuille d'exercice pour fournir vos réponses, ainsi qu'un lien vers le jeu de données publié.

Tâches finales



Pour vos activités finales, vous devrez compléter et soumettre le Cas Pratique II et le Cas Pratique III pour évaluation.

CAS PRATIQUE II

Il y a deux options pour le CAS PRATIQUE II ([Espèces envahissantes](#) OU [Observations de Lepidoptères](#)). Vous n'avez qu'à sélectionner l'un d'eux pour votre travail.

Fichiers requis pour la soumission :

- feuille d'exercice remplie (Document MS Word ou similaire est acceptable)
- feuille de calcul de capture de données (MS Excel, csv, txt ou similaire est acceptable)
- jeu de données nettoyé/standardisé (MS Excel, csv, txt ou similaire est acceptable)

CAS PRATIQUE III

Il n'y a qu'une seule option pour le cas pratique III ([Oiseaux de la littérature](#)).

Fichiers requis pour la soumission :

- feuille d'exercice remplie (Document MS Word ou similaire est acceptable)
- feuille de calcul de capture de données (MS Excel, csv, txt ou similaire est acceptable)
- jeu de données nettoyé/standardisé (MS Excel, csv, txt ou similaire est acceptable)



Inclure le numéro du cas d'utilisation et de l'exercice avec votre nom sur tous les fichiers soumis, par exemple, Russell-UC2-IS-feuille-d-exercice.docx, Russell-UC2-IS-2.xlsx, Russell-UC2-IS-3.xlsx. **Tous les fichiers doivent être soumis en anglais.** Contactez training@gbif.org si vous avez des questions.

Soumission du travail

Les travaux peuvent être soumis à partir de la version en ligne (HTML) du cours.

Évaluation du cours



Complétez l'évaluation du cours

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.

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
- Devenir un éditeur de données avec GBIF
- Best Practices for Publishing Biodiversity Data from Environmental Impact Assessments
GBIF Secretariat & IAIA: International Association for Impact Assessment (2020).
- Guidance for private companies to become data publishers through GBIF: Template document to support the internal authorization process to become a GBIF publisher
Rui Figueira, Pedro Beja, Cristina Villaverde, Miguel Vega, Katia Cezón, Tainan Messina, Anne-Sophie Archambeau, Rukaya Johaadien, Dag Endresen & Dairo Escobar (2020).
- Publishing DNA-derived data through biodiversity data platforms
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
- 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 IPT ou ne pas installer IPT
- centres d'hébergement de données IPT
- Installation de l'IPT / Configuration du webinaire
- Installation de l' IPT en vidéo
- Demonstration pratique d'usage d'IPT en video

La numérisation

- [Resources de numérisation iDigBio](#)
- [Flux de travail 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](#)
- [Establishing an Effective GBIF Participant Node: Concepts and general considerations](#)
GBIF Secretariat (2019).

Géoréférencement

- [Georeferencing Best Practices](#)
Arthur D. Chapman & John R. Wieczorek (2020).
- [Georeferencing Quick Reference Guide](#)
Paula F. Zermoglio, Arthur D. Chapman, John R. Wieczorek, Maria Celeste Luna & David A. Bloom (2020).
- [Georeferencing Calculator Manual](#)
David A. Bloom, John R. Wieczorek & Paula F. Zermoglio (2020).
- [Ressources de Géoréférencement](#)

Espèces invasives

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

Living Atlases

- [Atlas des vivants](#)
- [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](#)
- [Guía para la limpieza de datos sobre biodiversidad con OpenRefine](#)
Paula F. Zermoglio, Camila A. Plata Corredor, John R. Wieczorek, Ricardo Ortiz Gallego & Leonardo Buitrago (2021).
- [Using Google Refine and taxonomic databases \(EOL, NCBI, uBio, WORMS\) to clean messy data](#)
iPhylo blog post. Rod Page 2012.
- [Reconciling author names using Open Refine and VIAF](#)
iPhylo blog post. Rod Page 2013.
- [Validating scientific names with the GBIF Portal web service API](#)
Guest post was written by Gaurav Vaidya, Victoria Tersigni and Robert Guralnick 2013.
- [iDigBio Cleaning data with OpenRefine](#)
iDigBio 2013.
- [Have We Got the Names "Right"?](#)
Canadensys 2014.
- [Cleaning data with OpenRefine](#)
Desmet and Brosens 2016 TDWG.
- [EasyOpen Redlist](#)
Querying the IUCN Red List, using a species list, OpenRefine, and some pre-written code. Olly Griffin July 2019.

Planification/Collaboration

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

Qualité

- [Principles of Data Quality](#)
Arthur Chapman 2005.
- [Principles and Methods of Data Cleaning: Primary Species and Species-Occurrence Data](#)
Arthur Chapman 2005.
- [Be careful with dates in Excel](#)
DataOne 2014.
- [Encodage de caractères pour les débutants](#)
- [Guide MVZ d'enregistrement des localités dans les notes de terrain](#)

Espèces sensibles

- [Current Best Practices for Generalizing Sensitive Species Occurrence Data](#)
Arthur D. Chapman 2020.

Taxonomie

- [Jeux de données des checklist du GBIF et lacunes des données](#)
- [Laboratoires GBIF - Analyseur de noms](#)
- [Laboratoires GBIF - Correspondance d'espèces](#)
- [Global Names Resolver](#)
- [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 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 par 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 `basisOfRecord` DwC 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, maintenue par TDWG et largement

utilisée 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'objets numériques. Un identifiant ou une poignée persistante utilisée 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, et les publications officielles.

DwC-A

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

1. les données réelles, 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 Unique Global

IPT

Integrated Publishing Toolkit est une application web libre et open source (logiciels) 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 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 collectif 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

Nombre de Ressources Uniformes

UUID

Identifiant universel unique

Annexe : Articles sur les données (data papers)



Un data paper est un document évalué par les pairs décrivant un jeu de données, publié dans un journal révisé par les pairs. Cela demande des efforts pour préparer, gérer/conserver et décrire les données. Les data papers permettent une reconnaissance de ces efforts via une publication savante. Nous ne couvrons pas la manière de créer des data papers dans ce cours, néanmoins, comme activité optionnelle, vous pouvez regarder cette vidéo (51:51) présentée par Lizanne Roxburgh. Dans cette vidéo, vous en apprendrez plus sur la publication de data papers. Si vous ne pouvez pas regarder la vidéo intégrée, vous pouvez la [télécharger](#) localement (MP4 - 99,2 Mo)

► <https://vimeo.com/265350948> (Vimeo video)

Vous pouvez en savoir plus sur les data papers sur GBIF.org.

What it is: A scholarly publication of searchable metadata – a document describing a dataset, or a group of datasets

DOI: indexation and citation

Promote and publicize existence of data

Provide scholarly credit to data publishers through citable journal publications

Describe the data in a structured human-readable form



Integrated Publishing Toolkit (IPT) facilitates authoring of metadata and auto-generation of Data Paper manuscripts

Harvestmen_of_French_Guiana

This dataset provides information on specimens of harvestmen (Arthropoda, Arachnida, Opiliones) collected in French Guiana. Field collections have been initiated in 2012 within the framework of the Center for the Study of Biodiversity in Amazonia (CEBA: www.labex-ceba.fr/en/). This dataset is a work in progress. Occurrences are recorded in an online database stored at the EDB laboratory after each collecting trip and the dataset is updated on a monthly basis. Voucher specimens and associated DNA are also stored at the EDB laboratory until deposition in natural history Museums. The latest version of the dataset is publicly and freely accessible through our Integrated Publication Toolkit at http://130.120.204.55:8080/ipt/resource.do?r=harvestmen_of_french_guiana or through the Global Biodiversity Information Facility data portal at <http://www.gbif.org/dataset/3c9e2297-bf20-4827-928e-7c7eefd9432c>.

Summary

Date Published	May 20, 2015
Version	23 (Latest)
Update Frequency	Monthly (Next publication: Jun 19, 2015)
Darwin Core	download (47 KB) 1474 records
Archive	
EML	download (24 KB)
RTF	download (23 KB)
GBIF Registration	3c9e2297-bf20-4827-928e-7c7eefd9432c
Organisation	Laboratoire EDB "Evolution et Diversité Biologique"
Endorsing Node	GBIF France

Keywords

Occurrence; French Guiana; Neotropics; Opiliones

Language

Metadata Language English
Resource Language English

External Links

Resource <http://www.gbif.org/dataset/3c9e2297-bf20-4827-928e-7c7eefd9432c>
Homepage

Resource Contact

Name Sébastien Cally

Consultez les liens ci-dessous pour voir un data paper tel qu'il apparaît sur l'IPT, sur GBIF.org et dans le Journal. Ils sont inter-connectés.

- Journal : <https://doi.org/10.3897/BDJ.2.e4244>
- GBIF : <https://www.gbif.org/dataset/3c9e2297-bf20-4827-928e-7c7eefd9432c>
- IPT : http://130.120.204.55:8080/ipt/resource.do?r=harvestmen_of_french_guiana

Annexe : Solutions



Cette annexe contient les réponses et des informations complémentaires à tous les questionnaires de révision. Cette section contient également une solution suggérée pour le CAS PRATIQUE I.

Solutions de révision des fondations

Pour l'énoncé donné, entrez le terme correct (base de données, langage de base de données, programme de base de données)

- regroupe et propose des fonctions et des caractéristiques permettant de manipuler des données, le tout dans une interface unifiée + **programme de base de données**
- collection structurée et organisée de données et/ou d'informations conservées sur un ordinateur + **base de données**

- la façon dont un humain communique avec un ordinateur
langage de base de données

Si vous ouvrez un fichier de données et voyez ce qui suit, que soupçonnez-vous d'être le problème

Être, ou ne pas être, c'est là la question.

- Un mauvais encodage a été utilisé pour ouvrir le fichier

Pour le logiciel donné, saisissez le type de logiciel (saisie des données, gestion des données, nettoyage des données, publication des données).

- Integrated Publishing Toolkit (IPT)
publication de données
- Specify + **saisie des données** ET **gestion des données**
- iNaturalist + **saisie des données**
- OpenRefine + **nettoyage des données**

Pour l'exemple donné, entrez le type de données correct (binaire, booléen, flottant, entier, entier long, texte, texte non structuré)

- 1236975 + **entier long**
- 01101111 + **binaire**
- Nous avons marché à 5 milles sur la route vers l'ouest depuis le bureau de poste au centre de la ville. Nous sommes ensuite allés à 2 milles au nord sur un chemin de terre vers la rivière. Puis nous avons continué vers l'ouest le long de la rivière pendant encore 5 miles. + **texte non structuré**
- 1024 + **entier**
- 29.0 + **flottant**
- Oui/Non + **booléen**
- 6 lapins ont été observés + **texte**

Lesquels de ces termes décrivent un "nom de champ/colonne" ?

- Assigné(s)
- Identifiant
- Unique

Lequel de ces termes décrit une "étiquette de champ" ?

- Descriptif
- Lisible
- Interface d'utilisateur

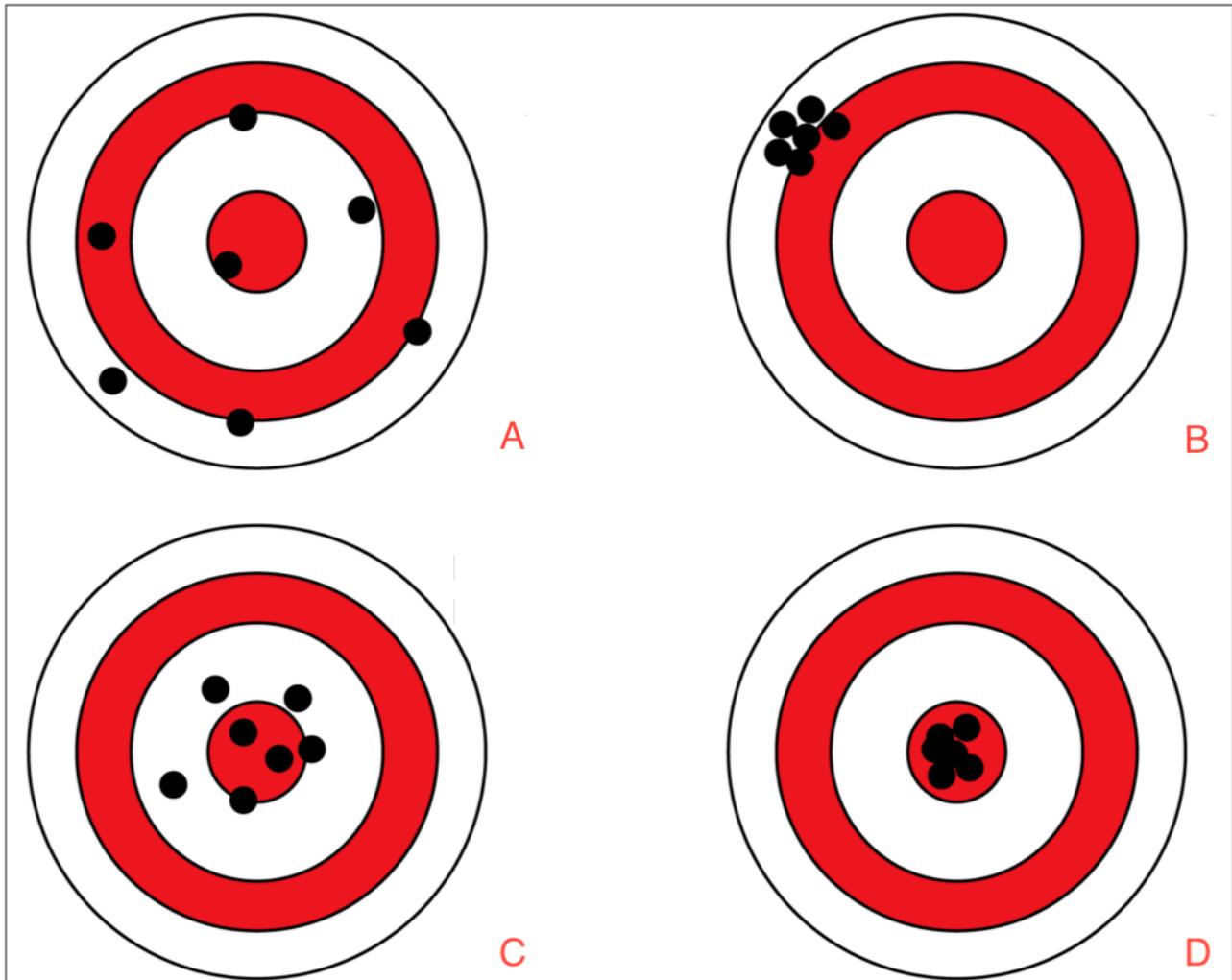
Pour chaque relevé, entrez la structure correcte (ligne, colonne, table)

- Toutes les données font référence à un concept SINGULIER. + **table**
- Un attribut a le MÊME type de données/champ pour chaque enregistrement. + **colonne**
- Les attributs d'un enregistrement de données restent TOUJOURS ensemble. + **ligne**

Qui détermine l'aptitude à l'usage de vos données ?

- Les utilisateurs des données pour la recherche ou l'éducation

Pour les assertions données, entrez l'image correspondante. (A, B, C, D)



- Haute exactitude, faible précision + **C**
- Faible exactitude, haute précision + **B**
- Haute exactitude, haute précision + **D**
- Faible exactitude, faible précision + **A**

Identifier les relations de données où l'ensemble de données B doit être fusionné avec l'ensemble de données A (0:1, 1:0, 1:1, 1:∞, ∞:1, ∞:∞). Toutes les relations ne sont pas utilisées.

- Le champ Collector existe dans l'ensemble de données A et B + **1:1**
- Le champ pays n'existe que dans l'ensemble de données B + **0:1**
- Le champ nom existe dans l'ensemble de données A, mais l'ensemble de données B contient les champs Prénom et Nom + **1:∞**
- Le champ ID existe dans l'ensemble de données A et B + **1:1**
- L'élévation existe dans l'ensemble de données A, mais pas dans l'ensemble de données B + **1:0**
- La Date existe dans l'ensemble de données A, mais le Jour, le Mois et l'Année sont des champs séparés dans l'ensemble de données B + **1:∞**

Les métadonnées sont importantes car (sélectionnez les affirmations qui sont VRAIES)

- elles permettent aux utilisateurs de déterminer si un ensemble de données convient à leur utilisation.
- elles vous permettent de savoir sous quelles conditions légales la réutilisation des données est autorisée.

Solutions de la révision sur la planification

Quel est l'ordre des cinq groupes de processus PMBoK ?

- Initier, planifier, exécuter, surveiller et contrôler, clôturer

En savoir plus : <https://quizlet.com/306742513/1-introduction-pmbok-guide-6th-edition-flash-cards/>

Quels sont les types de livrables ?

- Déclaré - OUI
- Implicite - OUI
- Estimé - NON
- Direct - OUI
- Indirect - OUI
- Deviné - NON

Qu'est-ce qu'un goulot d'étranglement ?

- un blocage qui retarde le développement ou le progrès - OUI un espace où quelque chose ou quelqu'un manque - NON, CECI EST UNE LACUNE
- un problème, ou une situation qui empêche quelqu'un de faire quelque chose, ou qui rend quelque chose impossible. NON, CECI EST UNE BARRIÈRE

Quels sont les exemples de tâches de mobilisation ?

- Affiliation - NON, Ceci est un type lié à la ressource
- Publication - OUI
- Imagerie - OUI
- Géoréférencement - OUI
- Sensibilisation accrue du public - NON, il s'agit d'un objectif implicite.

Solutions de vérification de la saisie des données

Quel(s) type(s) de jeu de données choisiriez-vous pour une collection d'ichtyologie?

- occurrence

La plupart du temps, les spécimens de bases de données de collection sont partagés en tant que données d'occurrence. Chaque occurrence (spécimen ou groupe de spécimens) a son propre identifiant unique (parfois dérivé de son numéro de catalogue dans la collection source) et les champs Darwin Core utilisés pour les partager au sein du GBIF décrivent chaque spécimen : nom scientifique, la date à laquelle il a été recueilli sur le terrain, qui l'a collecté et/ou identifié, où, etc. Chaque collection peut avoir plus d'un spécimen d'une même espèce, à condition que chaque

spécimen soit identifié par une identification unique.

- liste de contrôle

Il est également possible de créer et de partager une liste de contrôle taxonomique dérivée d'une base de données de collection ; dans ce cas, il est recommandé de partager la liste de contrôle en tant qu'ensemble de données taxonomiques, avec la liste des occurrences (spécimens) associées à elle en utilisant le noyau Occurrence comme extension du noyau taxon sur l'IPT du GBIF.

Quel(s) type(s) de jeu de données choisiriez-vous pour une liste d'espèces envahissantes?

- occurrence

Certains éditeurs de données partageront des ensembles de données d'occurrence provenant d'études ou de programmes de suivi de spécimens provenant de certaines espèces envahissantes spécifiques ; lorsque les données se concentrent sur les individus au lieu de l'espèce envahissante, en général, elles peuvent être partagées en tant que données d'occurrence.

- la liste de contrôle

Les espèces invasives peuvent être suivies et surveillées à différentes échelles (régionales, nationales, thématiques...); car ce type de jeu de données se concentre davantage sur les espèces et leur distribution sur une portée géographique donnée, ils sont principalement partagés en tant que jeux de données taxonomiques au sein du GBIF (https://www.gbif.org/dataset/search?project_id=GRIIS [voir les résultats de recherche GRIIS]).

Quel(s) type(s) de jeu de données choisiriez-vous pour la flore et la faune d'une étude d'impact environnementale?

- occurrences

Les données sont enregistrées par des naturalistes sur le terrain et peuvent être partagées en tant que simples jeux de données d'Occurrence.

- événement d'échantillonnage

Ils peuvent également être partagés en tant que jeux de données d'événements si des protocoles normalisés (comme des plots de végétation, les transectes, les pièges...) sont utilisés pour collecter les données.

Quel(s) type(s) de jeu de données choisiriez-vous pour des données de suivi des oiseaux?

- occurrence

Ces données sont partagées comme des jeux de données d'occurrences : idéalement, chaque oiseau est identifié avec son organisme et chaque occurrence (GPS ping) a son propre ID d'occurrence, qui est utile pour suivre les différents emplacements GPS du même oiseau dans le cadre du programme ou du projet de suivi. (Voir [exemple](#))

Quel(s) type(s) de jeu de données choisiriez-vous pour des données de pièges à insectes?

- occurrence

Bien que ces données puissent être partagées en tant que jeux de données d'occurrence simples, il est préférable qu'ils soient partagés comme des jeux de données d'événement, où l'emplacement, l'identifiant et le contenu de chaque piège peuvent être mieux détaillés.

- échantillonnage d'événement

Pièges à insectes (ainsi que d'autres pièges tels que pièges à trappes, les pièges à malaise...) sont généralement utilisés dans les programmes de surveillance pour vérifier la présence (ou l'absence) de certaines espèces et/ou évaluer leur abondance spécifique. L'utilisation du champ « eventID » pour identifier chaque piège permet aux utilisateurs de récupérer tous les spécimens dans chaque piège. La même logique s'applique aux autres protocoles de terrain tels que les transectes, les tracés, les caméras à distance, etc. en utilisant le Noyau d'Événement au lieu du

Noyau d'Occurrence, vous serez en mesure de partager beaucoup plus d'informations sur le contexte de la collecte de données, et permettre aux utilisateurs de mieux comprendre (et même de répliquer) votre travail.

Quel(s) type(s) de jeu de données choisiriez-vous pour des données de gestion d'un parc national ?

- occurrence
enregistrer des individus d'espèces
- liste de contrôle
Il est important de savoir combien d'espèces sont présentes dans le périmètre du parc ou de la réserve et leur état de conservation.
- échantillonnage d'événement
vérifiez et suivez les populations

Quel(s) type(s) de jeu de données choisiriez-vous pour un inventaire éclair (bioblitz) dans le cadre d'un programme de science participative?

- occurrence
Les jeux de données Bioblitz sont principalement partagés en tant qu'ensembles de données d'occurrence.
- événement d'échantillonnage
Selon le programme scientifique citoyen, des protocoles d'échantillonnage spécifiques peuvent être utilisés par les volontaires, auquel cas les données peuvent être partagées en tant que jeu de données d'événement.

Quel(s) type(s) de jeu de données choisiriez-vous pour une liste d'espèces régionale?

- liste de contrôle
Les listes d'espèces géographiques ou thématiques sont souvent utilisées pour partager des informations sur les espèces présentes dans une région donnée ; la plupart du temps, ces listes mentionnent également la répartition de chaque espèce ainsi que son état de conservation dans ce domaine. Les listes d'espèces régionales peuvent donner un aperçu utile de la biodiversité et des habitats d'une région, et doivent être partagés en tant qu'ensembles de données taxonomiques, avec ou sans occurrences associées.

Solutions des révisions sur la gestion des données

Pourquoi est-il préférable de nettoyer vos données ?

- pour les rendre aussi utilisables que possible
- pour atteindre vos objectifs de qualité des données

Vous devriez toujours vous efforcer de gérer et de publier des données de la meilleure qualité possible. Cela améliorera votre travail quotidien (il est plus facile de travailler avec des données organisées et bien nettoyées), ainsi que le travail des réutilisateurs potentiels de vos données, qui ont besoin de les comprendre et de faire confiance à leur source avant de les utiliser.

Comment devriez-vous organiser votre processus de nettoyage des données ?

- sollicitez vos collègues pour obtenir l'expertise qui vous manque
- travaillez au niveau institutionnel pour harmoniser les processus concernant la qualité des données

Personne n'est censé tout savoir sur les données sur la biodiversité ; vous pouvez demander de l'aide et des conseils à vos collègues ou à d'autres personnes expertes dans leur domaine, et vous assurez que vous appliquez les bonnes pratiques recommandées par votre institution lorsque vous nettoyez vos données.

Qu'est-ce qui est le mieux :

- empêcher les erreurs de se produire
- corriger les erreurs dès que vous les trouvez dans votre base de données ou dans votre feuille de calcul

La meilleure façon d'éviter de propager des erreurs dans vos données est de les empêcher de se produire au début du processus de collecte/enregistrement des données.

Bien sûr, les erreurs sont inévitables. Vous devriez donc les nettoyer dès que vous les trouvez et documenter le processus de nettoyage.

Si vous n'avez pas le temps ou les ressources pour nettoyer correctement vos données, Il est préférable d'attendre de pouvoir le faire au lieu de publier des données erronées, qui pourraient perturber le travail des personnes souhaitant les réutiliser.

Qui est responsable de la qualité des données ?

- Toutes les personnes impliquées dans la gestion des données

Chaque personne impliquée dans votre processus de gestion des données est au moins en partie responsable de leur qualité, des techniciens de terrain au(x) gestionnaire(s) de la base de données.

Les personnes qui pourraient ultérieurement utiliser vos données peuvent vous informer de toute erreur restante dans vos données, et devraient les utiliser de manière appropriée pour leurs propres recherches, mais la qualité initiale des données n'est pas leur responsabilité.

Le GBIF peut effectuer des vérifications automatiques sur vos données (par ex. la détection des valeurs manquantes, des aberrations géographiques, des noms scientifiques inconnus) mais ne peut pas être tenu responsable des erreurs qui se sont produites plus tôt dans le processus de gestion des données.

Quels outils peuvent être utilisés pour nettoyer vos données ?

- Excel et autres outils de gestion des feuilles de calcul
- OpenRefine
- Votre logiciel de base de données
- Des outils en ligne tels que le Scientific Names Resolver ou Google Maps

Tous types d'outils peuvent être utilisés pour nettoyer vos données, mais vous devez identifier ceux qui répondront à vos besoins en termes de réconciliation taxonomique, de géo-référencement, de suppression des doublons, etc. Vous pouvez trouver [la liste des outils utiles](#) dans la section dédiée à la gestion des données.

Solutions des révisions sur la publication des données

Que signifie la publication des données, dans le contexte du GBIF ?

- Rendre votre ou vos jeu(x) de données sur la biodiversité accessible(s) au public et visible(s) dans

un format standardisé

La publication de données dans le GBIF signifie rendre votre ou vos jeu(x) de données sur la biodiversité accessible(s) publiquement dans un format standardisé (la plupart du temps, le Darwin Core), afin qu'il(s) puisse(nt) être découvert(s) et réutilisé(s) par d'autres personnes.

Qu'est-ce qu'un IPT ?

- un outil qui vous aide à publier vos données sur le GBIF
- un outil qui vous aide à produire un data paper (article de données)

L'IPT (Integrated Publishing Toolkit) est un logiciel codé en Java qui vous permet de charger et de publier des données sur le GBIF. Il ne doit pas être utilisé comme outil de gestion des données ou de nettoyage des données.

L'IPT peut également vous aider dans le processus d'écriture et de soumission d'un data paper (article de données), grâce au fichier EML qu'il génère automatiquement lorsque vous remplissez les métadonnées de votre ressource de données.

Quelles licences ou dérogations Creative Commons sont recommandées par le GBIF pour la publication de données ?

- CC0, CC-BY et CC-BY-NC

Les licences et dérogations Creative Commons recommandées pour publier vos jeux de données sur le GBIF sont CC0, CC-BY et CC-BY-NC. Ce sont des licences et/ou des dérogations largement reconnues qui s'alignent sur les exigences internationales de l'open-data en matière de partage et de réutilisation des données.

Veuillez noter que vous ne devez choisir que la dérogation/licence CC0 ou CC-BY pour votre (vos) jeu(x) de données publié(s) dans le cadre du programme BID.

Quels sont les trois noyaux ("Cores") que vous pouvez choisir pour une ressource sur l'IPT ?

- Occurrence Core, Taxon Core, Event Core

Vous pouvez choisir l'un des trois noyaux ("Cores") suivants pour chacune de vos ressources IPT : Occurrence, Taxon ou Event.

Le standard Darwin Core vous permet également de lier des extensions au noyau de votre choix, comme SimpleMultimedia ou MeasurementOrFact.

Les métadonnées sont remplies dans une section séparée de l'IPT et sont partagées en utilisant le standard EML, pas le Darwin Core (qui n'est utilisé que pour les données).

Combien de fichiers d'extensions un jeu de données peut-il avoir ?

- autant que nécessaire

Une fois que vous avez choisi un noyau pour votre ressource IPT, vous pouvez y ajouter des extensions Darwin Core. Vous pouvez n'ajouter qu'une ou plusieurs extensions, selon le type de Core que vous avez choisi, et les extensions qui sont compatibles avec celui-ci.

Les extensions ne sont pas obligatoires (vous pouvez publier un jeu de données sans aucune extension) mais peuvent être utiles si vous voulez partager des informations supplémentaires que vous ne pouvez pas associer avec le noyau de votre choix.

Solution suggérée pour le Cas pratique I

solution suggérée (PDF 144 ko)

Remerciements

Conception du cours et instruction

Le succès de ce cours dépend fortement du soutien apporté aux participants par le réseau des mentors GBIF pour le renforcement des capacités. Visitez la page du GBIF sur [lehttps://www.gbif.org/article/5SExsCfj7UaUkMCsuc60ec/mentors-and-trainers](https://www.gbif.org/article/5SExsCfj7UaUkMCsuc60ec/mentors-and-trainers)[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, aux matériaux, et à l'enseignement du cours:

- Nestor Beltran*
- David Bloom
- Katia Cezón*
- Dag Endresen
- Alberto González-Talaván*
- Sharon Grant*
- Marie-Elise Lecoq
- Sophie Pamerlon*
- Nicolas Noe*
- Mélianie Raymond*
- Laura Anne Russell*
- John Wieczorek
- Paula Zermoglio

*Initiateurs du programme

Reconnaissance spéciale à Arthur Chapman pour la réutilisation de ses documents sur la qualité des données.

Traducteurs

Français

- Maxime Coupremagne
- Jaures Gbètoho
- Marie Grosjean
- Patricia Mergen
- Sophie Pamerlon
- Andry Jean Marc Rakotomanjaka

- Y. Sabastian Wirsiy

Portugais

- Rui Figueira
- Clara Baringo Fonseca
- Keila Elizabeth Macfadem Juarez
- Tainan Messina

Espagnol

- Leonardo Buitrago
- Victor Chocho
- Camila Plata
- Anabela Plos
- William Ulate
- Paula Zermoglio

Ressources

- La plateforme de tables de jeu en ligne est fournie par [PlayingCards.io](https://playingcards.io). Nous tenons à remercier Jwalant Patel et Eric Ma pour avoir trouvé et aidé à créer les tables de jeu en ligne, ainsi que Kate Webbink pour son expertise artistique.
- Les icônes utilisées dans ce cours sont faites par Freepik à partir de www.flaticon.com
- [OpenRefine](#)
- [Kit de publication intégré](#)

Support des ressources

- [Plateforme Belge pour la Biodiversité](#)
- [GBIF France](#)
- [GBIF Norvège](#)
- [GBIF Espagne](#)
- [SiB Colombie](#)
- [The Field Museum](#)
- [VertNet](#)

Colophon

Suggestion de citation

GBIF Secretariat (2021) Cours de mobilisation des données sur la biodiversité du GBIF. 12^e édition. Secrétariat du GBIF: Copenhague. <https://doi.org/10.35035/ce-c6cr-6w42>. [Date du cours.]

Contributeurs

Le *Cours de mobilisation des données sur la biodiversité du GBIF* a été développé à l'origine dans le cadre du [Biodiversity Information for Development](#), un programme financé par l'[Union européenne](#). Le programme initial a été créé par Nestor Beltran, Sharon Grant, Nicolas Noee, Sophie Pamerlon, Alberto Gozaëlez-Talavañ, Melianie Raymond, Laura Anne Russell et Katia Cezón, avec des contributions supplémentaires des formateurs, mentors et étudiants du GBIF.

Licence

Le *cours de mobilisation des données sur la biodiversité du GBIF* est sous licence [Licence Creative Commons Attribution 4.0](#).

URI permanent

<https://doi.org/10.35035/ce-c6cr-6w42>

Contrôle de document

12ème édition, Mai 2021