

Projet Living Data : Proposition d'héberger un projet de sauvetage de données

Section A : Informations professionnelles

(1) Votre nom : _____

Votre organisation : _____

(2) Qui est le ou la propriétaire initial-e des données et à quelle organisation cette personne est-elle affiliée ? Si vous n'êtes pas propriétaire des données, veuillez expliquer pourquoi le ou la propriétaire initial-e des données ne postule pas (par exemple, décès), ainsi que votre intérêt et votre engagement quant aux données.

(3) Expliquez pourquoi vous ou votre organisation ne pouvez pas archiver ces données sans la participation du Projet Living Data. Par exemple, vous pourriez manquer de formation spécialisée en gestion de données et en recherche reproductible, vous pourriez être proche de la retraite, ou la personne propriétaire initiale des données pourrait être décédée.

Section B : Les données

(4) Années au cours desquelles les données ont été collectées : de _____ à _____.

(5) Couverture spatiale de l'ensemble de données (s'il s'agit d'un seul emplacement, indiquez-le ; sinon, veuillez expliquer la zone sur laquelle les données ont été collectées).

(6) En termes généraux, quels types de variables ont été collectés dans cet ensemble de données (géolocalisation, température, abondance, diversité, croissance, etc.) ?

(7) Quel était le principal objectif de la collecte de cet ensemble de données ? Veuillez indiquer dans votre réponse s'il s'agit de données d'échantillonnage ou de données expérimentales.

(8) Pourquoi pensez-vous qu'il est important de préserver ces données pour les générations futures de scientifiques ? Les raisons pourraient par exemple inclure : des données historiques appropriées pour établir des bases de référence pour les changements actuels ou futurs ; des séries chronologiques utiles à la compréhension des processus dynamiques en écologie ; une expérience classique qui a eu de profondes répercussions sur la théorie ; des données sur des espèces ou des écosystèmes de grande importance pour des raisons biologiques, culturelles ou commerciales.

(9) Les données à sauver sont actuellement dans ce format (cochez tout ce qui s'applique) :

- Feuilles de calcul Excel
- Feuilles de calcul dans un autre format (précisez : _____)
- MS Access
- Base de données SQL
- Fiches techniques papier
- Rapports publiés
- Autres : (précisez : _____)

(9a) Je certifie que toutes les données à récupérer sont en format numérique et peuvent être transmises en ligne au Projet Living Data. Si ces données se trouvent sur des supports numériques obsolètes (bandes magnétiques, disquettes, etc.), je me chargerai de leur extraction.

Oui Non

Section C : Compétences requises

(10) Tous les étudiants et étudiantes aux cycles supérieurs seront formés aux outils R de validation et de compilation des données, ce qui, selon nous, suffira pour la plupart des projets de sauvetage de données. Si vous pensez que votre jeu de données nécessite des compétences informatiques supplémentaires, veuillez cocher les cases correspondantes.

- Programmation (énumérez les langages : _____)
- Système d'information géographique (ex. ArcGIS, QGIS, progiciels R géospatiaux)
- Langage de requête structuré (SQL, utilisé dans les bases de données relationnelles)
- Autres : _____

(11) Des connaissances dans ces domaines sont nécessaires ou souhaitables pour le projet (cochez tout ce qui s'applique) :

- Entomologie / écologie des insectes
- Ornithologie / écologie aviaire
- Mammalogie / écologie de la faune
- Foresterie / écologie forestière
- Pêche / écologie marine
- Agriculture / agroécologie
- Botanique / écologie végétale
- Limnologie / écologie aquatique
- Écologie des populations
- Écologie des communautés
- Écologie du paysage
- Biologie du changement climatique
- Biologie de l'évolution
- Phylogénétique

- Génétique des populations
- Génomique
- Autres : _____

(12) La langue utilisée dans l'ensemble de données est (cochez toutes les réponses qui s'appliquent) :

- Anglais
- Français

(13) Veuillez indiquer toutes autres compétences ou éléments nécessaires à la sauvegarde de votre ensemble de données qui n'ont pas été abordés.

Section D : Mise en œuvre du projet

(14) Qui, au sein de votre organisation, sera principalement responsable de l'encadrement du projet ? Cette personne devra se charger de fournir les données, de répondre aux questions relatives aux données et à leur collecte, et de rencontrer les formateurs et formatrices du programme PLD ainsi que les étudiant-es de l'équipe de sauvetage de données. Nous prévoyons cinq à dix rencontres au cours de l'automne.

(15) Outre la mise en pratique des compétences en gestion et archivage des données, quels autres avantages les étudiants et étudiantes pourront-ils retirer de leur participation à ce projet ? Nous souhaitons notamment qu'ils découvrent des carrières hors du milieu

universitaire et la possibilité de s'impliquer dans l'analyse de données, la publication et la diffusion des connaissances auprès du grand public.

Section E : Résultats

(16) Énumérez les résultats attendus (produits) du projet. Précisez si chaque résultat est obligatoire ou facultatif, selon le temps disponible. Exemples : élaboration d'un plan de gestion des données en vue de leur archivage futur, fusion de plusieurs sources de données, contrôle qualité des données, rédaction d'un article de synthèse des données dans le cadre du processus d'archivage. Remarque : tous les projets doivent obligatoirement déposer leurs données dans un dépôt public et rédiger une brève description du ou de la chercheuse et du jeu de données pour le site web de l'ICEE.

(17) Avez-vous un référentiel de données préféré (ou mandaté par l'employeur) pour les données à archiver ? Si tel est le cas, veuillez l'indiquer ici (sinon laissez vide, et nous pourrions décider ensemble) :

Nous acceptons tout référentiel de données qui utilise les meilleures pratiques en matière de stockage permanent et d'accès ouvert aux données. Un registre global de ces référentiels peut être trouvé sur r3data.org (veuillez nous contacter si votre référentiel préféré ne figure pas sur cette liste). Les référentiels de données populaires dans les

domaines de l'écologie, de l'évolution et des sciences de l'environnement incluent : KNB, GBIF, Environmental Data initiative, Genbank, Pangea, Sequence Read Archive. Les référentiels de données populaires pour les données scientifiques ou académiques en général incluent : Zenodo, Dryad, Figshare et Borealis. Les organismes gouvernementaux fédéraux sont généralement obligés de s'assurer que les données archivées dans un référentiel peuvent être trouvées par l'entremise du portail du gouvernement ouvert.

(18) Avez-vous des inquiétudes concernant la mise à disposition publique de ces données ? Veuillez nous le faire savoir, afin que nous puissions travailler avec vous sur des solutions possibles. Par exemple, les préoccupations concernant une utilisation abusive potentielle des données peuvent parfois être traitées en codant dans les données toute information qui devrait être transmise aux futurs utilisateurs et utilisatrices (par exemple, des changements dans la méthode de collecte ou la fiabilité des données). Les préoccupations concernant les emplacements d'espèces menacées ou ayant une valeur commerciale peuvent être résolues en diminuant la résolution spatiale. Les préoccupations concernant le fait de laisser du temps pour terminer une publication avant que les données ne soient rendues ouvertes pourraient être résolues par la fonction d'embargo du référentiel de Pangaea, ou par la politique d'embargo sans questions d'un an de Dryad et Figshare. Les préoccupations concernant l'utilisation commerciale des données peuvent être résolues par la licence [Creative Commons](#) appropriée (la licence CC NC empêche l'utilisation commerciale).

19. Je reconnais qu'il est de ma responsabilité de déterminer les variables à archiver, ainsi que les ensembles de données exacts à archiver, et que des données supplémentaires ne peuvent être ajoutées au projet une fois celui-ci lancé.

Oui Non