

DASA

Digital Animal Sound Archive

Contract - B2/233/P2/DASA

Contexte

Une grande variété d'animaux produit des signaux ou appels acoustiques, souvent spécifiques à chaque espèce. La bioacoustique est ainsi devenue un outil essentiel pour étudier la répartition, l'écologie, le comportement et la phénologie d'organismes souvent difficiles à observer visuellement (par ex. les mammifères marins, les chauves-souris). Cependant, les ensembles de données acoustiques sont souvent fragmentés entre institutions et formats, ce qui freine leur réutilisation et leur intégration pour des analyses écologiques à grande échelle. Le projet Digital Animal Sound Archive (DASA) a été lancé en février 2023 pour relever ces défis en créant une infrastructure unifiée pour le stockage, la gestion et la valorisation des enregistrements bioacoustiques collectés en Belgique. Le projet, coordonné par l'Institut des Sciences naturelles (IRSNB) en collaboration avec deux ONG environnementales (Natuurpunt et Natagora), s'est d'abord concentré sur les enregistrements de chauves-souris (Chiroptera) comme preuve de concept.

Objectifs

L'objectif principal de DASA était de centraliser les ensembles de données bioacoustiques dispersés au sein d'une infrastructure conforme aux principes FAIR (Findable, Accessible, Interoperable et Reusable), garantissant la préservation des données pour la recherche future et offrant une collection de référence fiable de sons validés au niveau de l'espèce.

Les autres objectifs sont :

- Concevoir et mettre en œuvre un modèle de base de données robuste pour les données et métadonnées acoustiques, en s'inspirant de cadres existants tels que ecoSound-web, Tethys et waarnemingen.be.
- Développer une plateforme web conviviale pour la recherche et la validation humaine des données.
- Mettre en place des solutions de stockage à long terme, capables de gérer des ensembles de données de plusieurs dizaines de téraoctets.
- Encourager la collaboration entre professionnels et scientifiques citoyens afin de renforcer la validation et la réutilisation des données.
- Promouvoir l'interopérabilité entre les réseaux de science citoyenne et les plateformes internationales de biodiversité telles que le Global Biodiversity Information Facility (GBIF).

Méthodologie

Un modèle de base de données complet a été développé, intégrant des entités pour le projet, le déploiement, la détection, l'occurrence et la détermination, permettant des processus d'identification humains et automatisés. Une procédure d'importation de données a été élaborée,

combinant des modèles de métadonnées standardisés et des scripts de validation automatisée garantissant la qualité et la cohérence taxonomique des données. L'architecture du système comprend une base de données relationnelle (PostgreSQL/PostGIS), un index Elasticsearch pour les recherches avancées, et une application web (<https://dasa.naturalsciences.be>) développée avec Spring Boot et React, offrant des API, des recherches à facettes, la visualisation de sonogrammes et des interfaces multilingues. L'infrastructure de stockage à l'IRSNB comprend 20 To de stockage primaire en réseau (NAS) avec sauvegardes hors site et archivage sur bande à long terme.

Résultats

Le projet DASA a mis en place avec succès une infrastructure numérique fonctionnelle intégrant près de deux millions d'enregistrements de chauves-souris, dont environ un million sont directement liés à des fichiers audio. Pour les autres enregistrements, les métadonnées sont préservées.

La plateforme prend en charge des déterminations humaines et automatiques, une gestion transparente des métadonnées et une validation collaborative. Elle est pleinement conforme aux principes FAIR et à la Directive européenne sur les données ouvertes, garantissant ouverture, interopérabilité et accessibilité à long terme.

Bien que le projet se soit initialement concentré sur les chauves-souris, la conception du système permet une extension aisée à d'autres taxons et domaines écoacoustiques.

Recommandations

- Afin d'exploiter pleinement le potentiel des développements de DASA, les recommandations suivantes sont proposées :
- Les futurs développements devraient donner la priorité à la simplification des flux de soumission des données, permettant aux contributeurs de télécharger directement enregistrements et métadonnées via l'interface web.
- Une implication continue des scientifiques citoyens et des réseaux professionnels sera essentielle pour maintenir les activités de validation et élargir la répartition des données.
- Toutes les données acoustiques, y compris celles des espèces communes et du bruit de fond, devraient être conservées, car elles sont précieuses pour les études écologiques et méthodologiques à long terme.
- Un plan d'acquisition de stockage devra être mis en place à l'IRSNB pour faciliter le stockage de nouveaux ensembles de données.
- Le renforcement de l'interopérabilité avec des pipelines de classification externes et des plateformes de données sur la biodiversité augmentera encore l'impact scientifique et la durabilité de l'infrastructure DASA.
- Un meilleur jeu de données de référence est nécessaire : des confirmations visuelles des sons enregistrés afin de remplir pleinement l'objectif de collection de référence.

Mots-clés

Bioacoustique ; Infrastructure numérique ; Gestion des données ; Sciences citoyennes ; Chauves-souris ; Suivi écologique ; Stockage à long terme.