

Résumé du projet

Les collections européennes de sciences naturelles contiennent la partie la plus importante et la plus significative des connaissances scientifiques mondiales sur la structure, l'environnement et la biosphère de la Terre. Les collections belges couvrent tous les domaines de l'histoire naturelle, y compris la zoologie, la botanique, la géologie, la paléontologie et l'anthropologie. Les institutions scientifiques belges abritent plus de 55 millions de spécimens, ce qui en fait l'une des plus importantes collections de sciences naturelles au monde.

Les partenaires du projet sont également impliqués dans la mise en place de DiSSCO (dissco.eu), le nouvel ESFRI lié aux collections d'histoire naturelle qui vise à assurer un accès ouvert aux collections européennes d'histoire naturelle, élargissant ainsi la communauté des utilisateurs et trouvant des solutions innovantes grâce à l'utilisation de données liées aux sciences naturelles.

L'évolution rapide de la programmation et des échanges de données entre différents systèmes et appareils (ordinateurs de bureau, ordinateurs portables, tablettes, smartphones, etc.), ainsi que le développement exponentiel des applications web donnant accès à une grande variété d'informations géocodées ou géoréférencées sur les sciences naturelles, obligent à adopter une approche dont le cœur est l'interconnexion de modules et de données séparés.

Le projet NaturalHeritage a étudié et évalué comment une architecture modulaire peut répondre à la diversité des systèmes de gestion des données et des collections en offrant un outil de recherche commun. Cette architecture est basée sur deux niveaux d'interopérabilité entre petits modules et avec les fichiers d'autorité internationaux. La nouvelle architecture se concentre sur l'interopérabilité au sein du portail NaturalHeritage.be (entre les différents sous-modules) et à l'extérieur avec les données disponibles auprès de ressources électroniques certifiées et les données publiées via des agrégateurs internationaux. L'organisation modulaire proposée pour NaturalHeritage permettra également une meilleure utilisation du système par les scientifiques internes et externes, les gestionnaires de collections et d'autres utilisateurs potentiels tels que les experts des différentes disciplines (zoologie, botanique, géologie, paléontologie, etc.).

Les objectifs du projet sont les suivants :

- Réaliser une meilleure gestion intégrée des collections numériques au niveau belge basée sur l'interopérabilité des différents ensembles de données et modules avec un portail de recherche commun ; Ceci concerne les données individuelles déjà numérisées dans les systèmes de gestion des collections mais aussi la description des collections qui ne sont pas encore numérisées (environ 90% des collections).
- Fournir des outils de validation adéquats, en utilisant des services web conformes aux normes internationalement reconnues pour la conservation numérique du contenu en tant que source (c'est-à-dire la vérification croisée des noms scientifiques, des noms de localités, l'utilisation de normes pour l'importation/exportation de données, l'utilisation de normes pour les outils de visualisation, etc.)

- Étendre le(s) modèle(s) de données à de nouveaux types de collections hébergées par les partenaires qui ne sont pas centrées sur les données taxonomiques, afin de parvenir à une interopérabilité plus large et à une plus grande application ;
- Se conformer aux exigences en matière de données conformément aux directives européennes (INSPIRE et OpenData) et aux protocoles (Nagoya) et règlements internationaux (APA) ;
- Fournir un accès libre et fiable à des données de haute qualité provenant des différentes parties prenantes et des différents utilisateurs par le biais d'un portail de recherche commun permettant un suivi historique et un aperçu global de la biodiversité et des données scientifiques via un système de recherche trié géographiquement et/ou thématiquement ;
- Promouvoir la nouvelle architecture parmi les collègues et les pairs utilisant le CETAF en élargissant la capacité de la plateforme à la fois géographiquement (couvrant l'Europe et l'international) et conceptuellement (à d'autres disciplines connexes), à la lumière des défis sociétaux existants (espèces invasives, sécurité sanitaire, etc.).

La méthodologie adoptée par le projet a permis de consolider et d'élargir la collaboration, l'intégration et l'interopérabilité entre les institutions belges et de développer le noeud technique pour les partenaires belges de DiSSCo. La priorité a été donnée aux technologies Open Source qui sont déjà utilisées par les partenaires pour fédérer les efforts et produire une interopérabilité complète entre les partenaires belges, mais aussi avec des systèmes internationaux plus grands et plus larges comme le portail GBIF et qui seront compatibles avec le système Elvis de DiSSCo.

Sur la base de cette approche, le projet a développé un nouvel écosystème de base de données et un portail de recherche afin qu'il y ait une interopérabilité entre les 3 partenaires collections (RBINS, RMCA et APM).

Le projet a étendu le système de gestion de collection DaRWIn existant à de nouvelles catégories d'objets et de données/métadonnées associées. Le modèle de données couvre des données et des métadonnées compatibles avec d'autres efforts internationaux dans le domaine de l'environnement ou des archives (GBIF, BIOCASE, GEOGASE, OAI-PMH) et remplit les obligations des États membres de l'Union européenne (dérivées essentiellement de INSPIRE, Open Data et du Protocole de Nagoya).

Le projet a également développé de nouveaux modules dédiés à des types d'objets spécifiques comme les images, le multimédia, les archives du protocole de Nagoya et de la CITES, la bibliographie, etc.

L'interopérabilité de toutes ces données/métadonnées aidera les scientifiques et les décideurs à aborder les collections d'histoire naturelle belges en utilisant des approches de données croisées et de big data. Le projet a également prototypé le noeud technique belge pour le nouveau DiSSCO ESFRI.

Le projet a également développé un nouveau portail de recherche commun permettant aux utilisateurs d'utiliser 3 niveaux de questions :

- une simple recherche sur Google,
- QUI ?, QUAND ?, OU ?, COMMENT ? comme simple question "humaine".
- ou une recherche détaillée à facettes

L'utilisation des identifiants uniques permet de toujours établir un lien avec les données originales. Ceci est important pour la propriété des données, la traçabilité et la citation de la source.

Certaines collections spécifiques comme les herbiers permettent des programmes de numérisation massive mais la plupart des collections doivent être numérisées et le ratio global de la collection numérisée au niveau des spécimens ne dépasse pas 10%. La numérisation complète des spécimens est donc un objectif à très long terme. Il est très important de fournir aussi des informations au niveau des collections et sous-collections même si elles ne sont pas encore numérisées. C'est pourquoi le consortium NaturalHeritage a développé un outil utilisant l'approche et les résultats du projet NH et permettant de fournir des données/métadonnées au niveau des collections en donnant des informations sur la taille, la diversité et la couverture taxonomique et/ou géographique de ces collections.

Le projet a permis aux partenaires d'accomplir des développements importants dans le cadre de la préparation des collections belges à l'infrastructure DiSSCo et la plupart des développements sont basés sur des technologies Open source, et disponibles sur les dossiers Github des partenaires du projet pour être réutilisés par d'autres institutions/collections en Belgique ou à l'étranger.

Néanmoins, la division des compétences et des ressources financières rend (très) difficile le transfert d'expertise/de technologies à des partenaires intéressés qui ne font pas partie du même pilier institutionnel (régional ou fédéral). Le programme BRAIN encourage la collaboration entre les FSI et les institutions régionales dans le domaine de la recherche, mais il n'existe aucune possibilité de financement pour le transfert des technologies/expertise à long terme. Par exemple, le programme d'infrastructure du FWO ne permet pas le financement d'institutions fédérales même en tant que sous-traitant et le programme fédéral soutenant les infrastructures ne finance que les FSI alors que les réductions budgétaires obligent à concentrer les activités sur les compétences fédérales exclusives. Une voie de collaboration claire/facile entre le niveau fédéral et le niveau régional doit être établie pour renforcer la ou les collaboration(s) entre les FSI et les institutions scientifiques d'autres niveaux.

Mots-clés:

Portail de recherche, Open Source, Système de gestion des collections, Webservices, DiSSCo ESFRI