Summary of the project

European natural science collections contain the largest and most significant part of the world's scientific knowledge of the earth's structure, environment and biosphere. The Belgian collections cover all the fields of Natural History including Zoology, Botany, Geology, Palaeontology and Anthropology. Belgian scientific institutions house more than 55 million specimens, some of the most important collections relating to Natural Science worldwide.

The partners of the project are also involved in DiSSCO (dissco.eu), the new ESFRI related to the Natural History collections aiming to ensure open access to European natural history collections. This will broaden the user community and promote innovative solutions through the use of natural science data.

The fast evolution in programming and data exchange between different systems and devices (desktops, laptops, tablets, smartphones, etc.), together with the exponential development of web applications providing access to a wide variety of geo-coded or geo-referenced natural sciences information, forces one to adopt an approach with the interconnection of separated modules and data at its core.

The NaturalHeritage project researched and evaluated how a modular architecture can address the diversity of data and collection management systems offering a common research tool. This architecture is based on two levels of interoperability between small modules and with international authority files. The new architecture focuses on interoperability within the NaturalHeritage.be portal (amongst the various sub-modules) and outside with the data available from certified e-resources and data published via international aggregators. The proposed modular organization of NaturalHeritage will also enable better usability of the system by internal and external scientists, collection managers and other potential users including experts from different disciplines (Zoology, Botany, Geology, Palaeontology, etc.).

The project's objectives are:

• To achieve a better integrated digital collections management at a Belgian level based on the interoperability of the different datasets and modules with a common search portal; This concerns the individual data already digitized in the Collection Management Systems but also the description of the collections that are not yet digitized (about 90% of the collections).

• To provide adequate validation tools, by using web-services compliant with internationally recognized standards for digital curation at source (i.e. cross-checking scientific names, locality names, using standards for import/export of data, using standards for visualisation tools, etc.)

• To extend the data model(s) to new types of collections housed by partners that are not centred on taxonomic data, in order to reach a broader interoperability and greater application;

• To be compliant with data requirements according to European directives (INSPIRE & OpenData) and international protocols (Nagoya) and regulations (ABS);

• To provide reliable open access to high quality data from the different stakeholders and users through a common research portal allowing an historical track and a global overview on biodiversity and scientific data via a search system sorted geographically and/or thematically;

• To promote the new architecture among colleagues and peers using CETAF by enlarging the capacity of the platform both geographically (covering Europe and internationally) and conceptually (to other related disciplines), in the light of existing societal challenges (invasive species, health security, etc.).

The methodology adopted by the project consolidated and broadened the collaboration, integration and interoperability amongst Belgian institutions leading to the development of the technical hub for the Belgian DiSSCo partners. Priority was given to Open Source technologies already used by ICT partners to combine efforts to produce full interoperability between the Belgian partners, as well as the broader international initiatives like GBIF and the DiSSCo Elvis system.

The project developed a new database ecosystem and search portal with interoperability between the 3 collection partners (RBINS, RMCA and APM).

The project extended the existing DaRWIN collection management system to new categories of objects and associated data/metadata. The data model produces data and metadata compatible with other international efforts in the environmental domain or for the archives (GBIF, BIOCASE, GEOGASE, OAI-PMH) and fulfil the obligations of the European Union Member States (derived basically from INSPIRE, Open Data and Nagoya Protocol).

The project also developed new modules dedicated to specific types of objects like images, multimedia, Nagoya protocol and CITES archives, bibliography, etc.

The interoperability of all these data/metadata will help scientists and decision makers to access Belgian Natural History Collections, using cross-linked and big-data approaches. This will aid the design of the Belgium data hub for the new DiSSCO ESFRI.

The project developed also a new common research portal allowing users to use 3 levels of questions:

- simple google search,
- WHO?, WHEN?, WHERE?, HOW? as simple “human” questions
- or detailed faceted search

The use of UUID allows it to always link to the original data. This is important for data ownership, traceability and citation of source.

Some specific collections, like the herbarium, have mass digitization programs, but most of the other collections still need to be digitized where the global ratio of the collection digitized at the specimen level does not exceed 10%. The complete digitisation of the specimens is thus a long term objective. It is very important to also provide information at the collection and sub-collection level even if they are not yet digitized. This is why the NaturalHeritage consortium developed a tool to provide data/metadata at the collection level concerning the size, diversity, taxonomy and geographic scope of these collections.

The project allowed partners to accomplish important developments within the framework of Belgian collections in the DiSSCo infrastructure. Where most of the developments are based on Open source technologies, and are available on the Github repository folders for reuse by other institutions in Belgium or abroad.

Nevertheless, the division of competences and funding resources make it difficult to transfer expertise/technologies to interested partners that are not in the same institutional category (regional or federal). The BRAIN program promotes collaboration between FSI’s and regional institutions within research, however there is no funding opportunity for the transfer of the technologies/expertise in the long term. As an example, the infrastructure program of the FWO does not allow the funding of federal institutions even as subcontractors, while the federal program supporting infrastructure only funds FSI’s where the budgetary cuts oblige one to focus activities on exclusively federal institutions. A clear and easy collaboration pathway between federal and regional levels should be established to reinforce collaboration(s) between FSIs and other scientific institutions.

Keywords:

Search portal, Open Source, Collection management system, Webservices, DiSSCo ESFRI