

Conference Abstract

# NaturalHeritage: Bridging Belgian natural history collections

Franck Theeten<sup>‡,§</sup>, Marielle Adam<sup>§</sup>, Thomas Vandenberghe<sup>§</sup>, Mathias Dillen<sup>l</sup>, Patrick Semal<sup>§</sup>, Serge Scory<sup>§</sup>, Jean-Marc Herpers<sup>‡</sup>, Didier Van den Spiegel<sup>‡</sup>, Patricia Mergen<sup>‡,l</sup>, Larissa Smirnova<sup>‡,l</sup>, Henry Engledow<sup>l</sup>, Ana Casino<sup>¶</sup>, Karsten Gödderz<sup>¶</sup>

<sup>‡</sup> Royal Museum for Central Africa (RMCA), Tervuren, Belgium

<sup>§</sup> Royal Belgian Institute of Natural Sciences (RBINS), Brussels, Belgium

<sup>l</sup> Meise Botanic Garden (MBG), Meise, Belgium

<sup>¶</sup> Consortium of European Taxonomic Facilities (CETAF, AISBL), Brussels, Belgium

Corresponding author: Franck Theeten ([franck.theeten@africamuseum.be](mailto:franck.theeten@africamuseum.be))

Received: 02 Jul 2019 | Published: 04 Jul 2019

Citation: Theeten F, Adam M, Vandenberghe T, Dillen M, Semal P, Scory S, Herpers J, Van den Spiegel D, Mergen P, Smirnova L, Engledow H, Casino A, Gödderz K (2019) NaturalHeritage: Bridging Belgian natural history collections. Biodiversity Information Science and Standards 3: e37854. <https://doi.org/10.3897/biss.3.37854>

## Abstract

The Royal Belgian Institute of Natural Sciences (RBINS), the Royal Museum for Central Africa (RMCA) and Meise Botanic Garden house more than 50 million specimens covering all fields of natural history.

While many different research topics have their own specificities, throughout the years it became apparent that with regards to collection data management, data publication and exchange via community standards, collection holding institutions face similar challenges (James et al. 2018, Rocha et al. 2014). In the past, these have been tackled in different ways by Belgian natural history institutions. In addition to local and national collaborations, there is a great need for a joint structure to share data between scientific institutions in Europe and beyond. It is the aim of large networks and infrastructures such as the Global Biodiversity Information Facility (GBIF), the Biodiversity Information Standards (TDWG), the Distributed System of Scientific collections (DiSSCo) and the Consortium of European Taxonomic Facilities (CETAF) to further implement and improve these efforts, thereby gaining ever increasing efficiencies.

In this context, the three institutions mentioned above, submitted the NaturalHeritage project ([http://www.belspo.be/belspo/brain-be/themes\\_3\\_HebrHistoScien\\_en.stm](http://www.belspo.be/belspo/brain-be/themes_3_HebrHistoScien_en.stm)) granted in 2017 by the Belgian Science Policy Service, which runs from 2017 to 2020.

The project provides links among databases and services. The unique qualities of each database are maintained, while the information can be concentrated and exposed in a structured way via one access point. This approach aims also to link data that are unconnected at present (e.g. relationship between soil/substrate, vegetation and associated fauna) and to improve the cross-validation of data.

(1) The NaturalHeritage prototype (<http://www.naturalheritage.be>) is a shared research portal with an open access infrastructure, which is still in the development phase. Its backbone is an ElasticSearch catalogue, with Kibana, and a Python aggregator gathering several types of (re)sources: relational databases, REpresentational State Transfer (REST) services of objects databases and bibliographical data, collections metadata and the GBIF Internet Publishing Toolkit (IPT) for observational and taxonomical data. Semi-structured data in English are semantically analysed and linked to a rich autocomplete mechanism. Keywords and identifiers are indexed and grouped in four categories ("what", "who", "where", "when"). The portal can act also as an Open Archives Initiatives Protocol for Metadata Harvesting (OAI-PMH) service and ease indexing of the original webpage on the internet with microdata enrichment.

(2) The collection data management system of DaRWIn (Data Research Warehouse Information Network) of RBINS and RMCA has been improved as well.

- External (meta)data requirements, i.e. foremost publication into or according to the practices and standards of GBIF and OBIS (Ocean Biogeographic Information System: <https://obis.org>) for biodiversity data, and INSPIRE (<https://inspire.ec.europa.eu>) for geological data, have been identified and evaluated. New and extended data structures have been created to be compliant with these standards, as well as the necessary procedures developed to expose the data.
- Quality control tools for taxonomic and geographic names have been developed. Geographic names can be hard to confirm as their lack of context often requires human validation. To address this a similarity measure is used to help map the result. Species, locations, sampling devices and other properties have been mapped to the World Register of Marine Species and DarwinCore (<http://www.marinespecies.org>), Marine Regions and GeoNames, the AGRO Agronomy and Vertebrate trait ontologies and the British Oceanographic Data Centre (BODC) vocabularies (<http://www.obofoundry.org/ontology/agro.html>). Extensive mapping is necessary to make use of the ExtendedMeasurementOrFact Extension of DarwinCore (<https://tools.gbif.org/dwca-validator/extensions.do>).

## Keywords

natural history collections, standardisation, webservices, search portal, interoperable databases, data analysis, data quality and cleaning

## Presenting author

Franck Theeten

## Presented at

Biodiversity\_Next 2019

## Funding program

BRAIN-be Belgian Research Action through Interdisciplinary Networks

## Grant title

NaturalHeritage: BR/175/A3/NATURALHERITAGE

Système de base de données modulaire interopérable et portail pour les collections belges d'Histoire Naturelle

Modulair interoperabel databasesysteem en portal voor de Belgische Natuurhistorische collecties

## Hosting institution

Royal Belgian Institute of Natural Sciences (RBINS)

29 rue Vautier, B-1000 Bruxelles

## References

- James SA, Soltis PS, Belbin L, Chapman AD, Nelson G, Paul DL, Collins M (2018) Herbarium data: Global biodiversity and societal botanical needs for novel research. Applications in Plant Sciences 6 (2): e1024. <https://doi.org/10.1002/aps3.1024>

- Rocha LA, Aleixo A, Allen G, Almeda F, Baldwin CC, Barclay MV, Bates JM, Bauer AM, Benzon F, Berns CM, et al. (2014) Specimen collection: An essential tool. Science 344 (6186): 814-815. <https://doi.org/10.1126/science.344.6186.814>