DaRWIN: An open source natural history collections data management system

Marielle Adam†, Franck Theeten§, Jean-Marc Herpers§, Thomas Vandenberghe†, Patrick Semal‡, Didier Van den Spiegel¶, Paul-André Duchesne

† Royal Belgian Institute of Natural Sciences (RBINS), Brussels, Belgium
‡ Royal Museum for Central Africa (RMCA), Tervuren, Belgium
¶ Natagora, Namur, Belgium

Corresponding author: Marielle Adam (madam@naturalsciences.be)

Received: 13 Aug 2019 | Published: 20 Aug 2019


Abstract

DaRWIN (Data Research Warehouse Information Network) is an in-house solution developed by the Royal Belgian Institute of Natural Sciences (RBINS), as a Natural History collections management system for biological and geological samples in collections. In 2014, the Royal Museum for Central Africa (RMCA) adopted this system for its collections and started to take part in new developments.

The DaRWIN database currently manages information on more than 600,000 records (about 4 million specimens) housed at the RBINS and more than 650,000 records (more than 1 million specimens) at the RMCA.

DaRWIN is an open source system, consisting of a PostgreSQL database and a customizable web-interface based on the Symfony framework (https://symfony.com).

DaRWIN is divided into 2 parts:

- one public section that gives a “read-only” access to digitised specimens,
- one section for registered users, with different levels of access rights (user, encoder, conservator and administrator), customizable for each collection and...
allowing update of specimens and collections, daily management of collections, 
and the potential for dealing with sensitive information.

DaRWIN stores sample data and related information such as place and date of collection, 
missions and collectors, identifiers, technicians involved, taxonomy, identification 
information (type, stage, state, etc.), bibliography, related files, storage, etc. Other features 
that deal with day-to-day curation operations are available: loans, printing of labels for 
storage, statistics and reporting. DaRWIN features its own JSON (JavaScript Object 
Notation) webservice for specimens and scientific names and can export data in tab-delimited, Excel, PDF and GeoJSON formats.

More recently, a procedure for importing batches of data has been developed, based on 
tab-delimited files, making integration of data from (old/historical) databases faster and 
more controlled.

Additional improvements of the user interface and database model have been made. For 
example, parallel taxonomical hierarchies can be created, allowing users to work with 
temporary taxonomies, old scientific names (basionyms and synonyms) and document the 
history of type specimens.

Finally, quality control and data cleaning on several tables have been implemented, e.g. 
mapping of locality names with vocabularies like Geonames, adding ISO 3166 two-letter 
country codes (https://www.iso.org/iso-3166-country-codes.html), cleaning duplicates from 
people/institutions and taxonomy catalogues. A tool for checking taxonomical names on 
GBIF (Global Biodiversity Information Facility), WoRMS (World Register of Marine Species) and DaRWIN itself, based on webservices and tab-delimited files, has been developed.

Last year, RBINS, RMCA and Meise Botanic Garden (MBG) defined a new framework of 
collaboration in the NaturalHeritage project (http://www.naturalheritage.be), in order to 
foster interoperability among their collection data sources. This new framework presents 
itself as one common research portal for data on natural history collections (from DaRWIN 
and other existing collection databases) of the three partnered institutions and makes data 
compliant to a standard agreed by the partners. See Poster "NaturalHeritage: Bridging 
Belgian Natural History Collections" for more information.

DaRWIN is accessible online (http://darwin.naturalsciences.be). A Github repository is also 

Keywords

data management, natural history collections, standardization, webservices, data quality 
and cleaning
Presenting author
Marielle Adam

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 database systeem en portal voor de Belgische Natuurhistorische collecties

Hosting institution
Royal Belgian Institute of Natural Sciences, 29 rue Vautier, B-1000 Brussels, Belgium