

## Conference Abstract

# Current progress in the development of taxonomic and anatomical ontologies within the scope of BIOfid

Markus Koch<sup>‡,§</sup>, Christine Driller<sup>‡</sup>, Marco Schmidt<sup>‡,|</sup>, Thomas Hörschemeyer<sup>¶</sup>, Claus Weiland<sup>‡</sup>, Istvan Miko<sup>#</sup>, Matthew Yoder<sup>□</sup>, Thomas Hickler<sup>‡</sup>

‡ Senckenberg Biodiversity and Climate Research Centre, Frankfurt am Main, Germany

§ Institute of Evolutionary Biology and Ecology, University of Bonn, Bonn, Germany

| Palmengarten der Stadt Frankfurt, Frankfurt am Main, Germany

¶ Senckenberg Research Institute and Natural History Museum Frankfurt, Frankfurt am Main, Germany

# Pennsylvania State University, University Park, United States of America

□ University of Illinois, Champaign, United States of America

Corresponding author: Markus Koch ([markus.koch@senckenberg.de](mailto:markus.koch@senckenberg.de))

Received: 06 Apr 2018 | Published: 17 May 2018

Citation: Koch M, Driller C, Schmidt M, Hörschemeyer T, Weiland C, Miko I, Yoder M, Hickler T (2018) Current progress in the development of taxonomic and anatomical ontologies within the scope of BIOfid. Biodiversity Information Science and Standards 2: e25585. <https://doi.org/10.3897/biss.2.25585>

## Abstract

The *Specialized Information Service Biodiversity Research* (BIOfid; <http://biofid.de/>) has recently been launched to mobilize valuable biodiversity data hidden in German print sources of the past 250 years. The partners involved in this project started digitisation of the literature corpus envisaged for the pilot stage and provided novel applications for natural language processing and visualization. In order to foster development of new text mining tools, the Senckenberg Biodiversity Informatics team focuses on the design of ontologies for taxa and their anatomy. We present our progress for the taxa prioritized by the target group for the pilot stage, i.e. for vascular plants, moths and butterflies, as well as birds. With regard to our text corpus a key aspect of our taxonomic ontologies is the inclusion of German vernacular names. For this purpose we assembled a taxonomy ontology for vascular plants by synchronizing taxon lists from the *Global Biodiversity Information Facility* (GBIF) and the *Integrated Taxonomic Information System* (ITIS) with K.P. Buttler's *Florenliste von Deutschland* (<http://www.kp-buttler.de/florenliste/>). Hierarchical

classification of the taxonomic names and class relationships focus on rank and status (validity vs. synonymy). All classes are additionally annotated with details on scientific name, taxonomic authorship, and source. Taxonomic names for birds are mainly compiled from ITIS and the *International Ornithological Congress (IOC) World Bird List*, for moths and butterflies mainly from GBIF, both lists being classified and annotated accordingly. We intend to cross-link our taxonomy ontologies with the *Environment Ontology (ENVO)* and anatomy ontologies such as the *Flora Phenotype Ontology (FLOPO)*. For moths and butterflies we started to design the *Lepidoptera Anatomy Ontology (LepAO)* on the basis of the already available *Hymenoptera Anatomy Ontology (HAO)*. LepAO is planned to be interoperable with other ontologies in the framework of the OBO foundry. A main modification of HAO is the inclusion of German anatomical terms from published glossaries that we add as scientific and vernacular synonyms to make use of already available identifiers (URIs) for corresponding English terms. International collaboration with the founders of HAO and teams focusing on other insect orders such as beetles (CoIAO) aims at development of a unified *Insect Anatomy Ontology*. With a restriction on terms applicable on all insects the unified Insect Anatomy Ontology is intended to establish a basis for accelerating the design of more specific anatomy ontologies for any particular insect order. The advancement of such ontologies aligns with current needs to make knowledge accumulated in descriptive studies on the systematics of organisms accessible to other domains. In the context of BIOfid our ontologies provide exemplars on how semantic queries of yet untapped data relevant for biodiversity studies can be achieved for literature in non-English languages. Furthermore, BIOfid will serve as an open access platform for professional international journals facilitating non-commercial publishing of biodiversity and biodiversity-related data.

## Keywords

Specialized Information Service, Biodiversity Research, Taxonomy, Anatomy Ontology, Tracheophyta, Lepidoptera, Aves, Text Mining

## Presenting author

Markus Koch

## Presented at

SPNHC+TDWG Conference 2018

## Funding program

Acquisition and Provision (Scientific Library Services and Information Systems DFG-LIS)

## **Grant title**

Specialized Information Service for Biodiversity Research (DFG-Project number 326061700)