

Connecting IndExs Editors and exsiccata IDs with Wikidata for Disambiguation of People Names and Work in Botanical and Mycological Collections

Dagmar Triebel[‡], Camila Uribe-Holguin[§], Stefan Seifert[‡], Markus Weiss[‡], Peter Scholz[§]

[‡] Staatliche Naturwissenschaftliche Sammlungen Bayerns, SNSB IT Center, Munich, Germany

[§] Staatliche Naturwissenschaftliche Sammlungen Bayerns, Botanische Staatssammlung München, Munich, Germany

Corresponding author: Dagmar Triebel (triebels@snsb.de), Camila Uribe-Holguin (uribe@snsb.de)

Abstract

"IndExs—Index of Exsiccatae" is an online database with bibliographic information on exsiccatae and exsiccata-like series launched in 2001 (Triebel and Scholz 2022). This type of series is a specific system in botany and mycology to create, publish and distribute well identified and documented reference material. In most cases the distributed specimens are numbered and each number consists of uniform material (herbarium duplicates) from a single collection event. Exsiccatal series are regularly published with small booklets containing the printed labels/ schedae of each numbered entity. The title of the series often shows the geographic and taxonomic focus of the series, e.g., "[Delogne & Gravet, Hépat. Ardenne](#)" and "[Hertel, Lecideaceae Exs.](#)". The persons editing the series are specialists, often recognized botanists and taxonomists. They are mostly not identical with the persons collecting and identifying the specimens distributed. Examples are E. M. Fries with "[Fries, Herb. Norm. Pl. Suec.](#)", G. L. Rabenhorst who published 24 series with more than 6,000 numbered entities and K. H. Rechinger with "[Rechinger & Polunin, Exs. Herb. Baghdad](#)". In the minority of cases the editors are anonymous persons and organisations devoted to plant exchange like "Société Dauphinoise pour l'échange des plantes". The more than 2.200 known series are widely distributed in public herbaria, either kept separately or integrated in the main collection. We estimate that more than 10 million specimens belong to such a series with printed labels. Approximately 70 series are running.

The eldest exsiccata might be that of Johann Balthasar Ehrhart (from 1732 see [here](#)). It is followed by the better known exsiccatae edited by Jakob Friedrich Ehrhart, e.g., the series "[Ehrhart, Pl. Crypt. Linn.](#)" starting with 1785. The two newest series started in 2020 and are bryophyte series from Taiwan and Vietnam.

The online database IndExs categorizes the series according to the group of organisms distributed and delivers editors, full title, standard abbreviation, editing institution, place of publication, range of (suggested) publication dates, range of numbered entities,

exemplary images of printed label as well as information sources and literature (Triebel et al. 2004). A stable and persistent exsiccata identifier, so-called "IndExs Exsiccata ID" is given. This set of standardized information is available via the [IndExs search interface](#), and ready to be downloaded via several formats (csv, xls, xml). A machine readable REST web service is under development. IndExs information and services with stable IndExs Exsiccata ID are used by data portals like the [Macroalgal Herbarium Consortium Portal](#) powered by [Symbiota](#), the [JACQ herbarium management system](#) and integrated in collection management systems like [DiversityCollection](#), a module of the [Diversity Workbench \(DWB\)](#) tool suite. It is envisaged to be included in future terminology services like the [GFBio Terminology Service](#). IndExs is appropriate to build the curated reference list for exsiccatae in the frame of herbarium digitization approaches (Borsch et al. 2020).

IndExs is storing information on the work of 1,300 editors of exsiccatae who are persons from nearly 300 years. According to the data models of DiversityAgents (Weiss et al. 2016) and DiversityExsiccatae (Hagedorn et al. 2008) the information is managed in freely accessible interlinked instances of SQL RDBMS [DiversityAgents](#) and [DiversityExsiccatae](#). The applications are installed as part of the data network at the [SNSB IT Center](#).

In 2012, the Wikidata project started and acts as central storage for the structured data of its Wikimedia sister projects (Anonymous 2022, Vrandečić and Krötzsch 2014). The content of Wikidata is available under a free license, exported using standard formats, and can be interlinked to other open data sets on the linked data web including life sciences (Anonymous 2022, Mitra et al. 2015).

The study will use existing IndExs services for the 1,300 IndExs editors and 2,200 disambiguated exsiccata series and expand them for [linked data](#) / semantic web approaches. It will explore the usability of Wikidata:

- for disambiguation of person names (=editors) in IndExs by adapting [Wikidata Identifiers](#),
- for adding information to existing Wikidata person Q-entities (items) via [statement s](#) on persons' work,
- for integrating IndExs information in Wikidata with URI for "IndExs Exsiccata ID" via a newly proposed P-entity ([property](#)) for this special kind of creative work in natural science,
- for adding new Q-entities in Wikidata for IndExs editors.

These editors of published booklet series with distributed physical material fulfill the Wikidata criteria of [notability](#). They are often more or less well-known botanists and mycologists. By their published work they might even more fulfill the criteria than certain persons categorized as botanical collectors with assignment of a Wikidata ID through activities of [CETAF](#), [DiSSCo](#) and [COST MOBILISE](#).

Keywords

linked data, external identifiers

Presenting author

Camila Uribe-Holguin

Presented at

TDWG 2022

Funding program

The study was inspired by discussions in the COST MOBILISE working groups WG3, WG4 (EU COST Action CA17106).

Conflicts of interest

References

- Anonymous (2022) Wikidata: Introduction. <https://www.wikidata.org/wiki/Wikidata:Introduction>. Accessed on: 2022-6-22.
- Borsch T, Stevens A, Häffner E, Güntsch A, Berendsohn W, Appelhans M, Barilaro C, Beszteri B, Blattner F, Bossdorf O, Dalitz H, Dressler S, Duque-Thüs R, Esser H, Franzke A, Goetze D, Grein M, Grünert U, Hellwig F, Hentschel J, Hörandl E, Janßen T, Jürgens N, Kadereit G, Karisch T, Koch M, Müller F, Müller J, Ober D, Porembski S, Poschlod P, Printzen C, Röser M, Sack P, Schlüter P, Schmidt M, Schnittler M, Scholler M, Schultz M, Seeber E, Simmel J, Stiller M, Thiv M, Thüs H, Tkach N, Triebel D, Warnke U, Weibulat T, Wesche K, Yurkov A, Zizka G (2020) A complete digitization of German herbaria is possible, sensible and should be started now. Research Ideas and Outcomes 6 <https://doi.org/10.3897/rio.6.e50675>
- Hagedorn G, Weiss M, Triebel D (2008) DiversityExsiccatae information model (version 2.0). https://www.diversityworkbench.net/Portal/DiversityExsiccataeModel_v2.0. Accessed on: 2022-6-29.
- Mitraka E, Waagmeester A, Burgstaller-Muehlbacher S, Schriml LM, Su AI, Good BM (2015) Wikidata: A platform for data integration and dissemination for the life sciences and beyond. bioRxiv <https://doi.org/10.1101/031971>
- Triebel D, Scholz P, Hagedorn G, Weiss M (2004) History of exsiccatal series in cryptogamic botany and mycology as reflected by the web-accessible database of exsiccatae "IndExs – Index of Exsiccatae" - In: Döbbeler P, Rambold G (Eds) Contributions to Lichenology. Festschrift in Honour of Hannes Hertel. Biblioth. Lichenol. 88: 671-690.
- Triebel D, Scholz P (2022) IndExs - Index of Exsiccatae. <http://indexs.botanischestaatssammlung.de>. Accessed on: 2022-6-29.

- Vrandečić D, Krötzsch M (2014) Wikidata. Communications of the ACM 57 (10): 78-85. <https://doi.org/10.1145/2629489>
- Weiss M, Hagedorn G, Triebel D (2016) DiversityAgents information model (version 2.1.9). https://www.diversityworkbench.net/Portal/DiversityAgentsModel_2.1.9. Accessed on: 2022-6-29.