On the coexistence of taxonomic botanical databases – a user study

Authors: Schrumpf, Annika, Killinger, Marcel, Schiessle, Pascal, and Scherp, Ansgar

On the coexistence of taxonomic botanical databases – a user study

Annika Schrumpf[1], Marcel Killinger[2], Pascal Schiessle[2] & Ansgar Scherp[2]

Abstract: There is a variety of taxonomic and nomenclatural databases, curated at different intervals and using different criteria to decide which species names are listed as accepted and which are considered synonyms. Botanical collections, such as herbaria or botanical gardens, maintain data that link names to plant material. The choice of the used database affects the naming of the plant specimens. If data from different institutions are to be matched, inconsistencies may arise. A solution that allows the use of different databases side by side would be beneficial in our opinion. The linking of botanical taxonomic and nomenclatural datasets by utilizing standardized Semantic Web technologies allows the coexistence of several lists. In this study, we conducted a mapping of The Plant List, World Flora Online, and the Integrated Taxonomic Information System provided in Darwin Core Archive files to the Simple Knowledge Organization System and created an interlinked version of these datasets. The developed graphical user interface visualized the contents of the included taxonomic databases. The usefulness and usability of the GUI were positively evaluated in interviews with five scientific employees and curators of four different botanical gardens.

Introduction

All over the world, botanical gardens cultivate plant specimens. Herbaria store over 390 million specimens (Thiers 2023+) for the purposes of research, education, and conservation. An important information about the specimens is their scientific name and their respective family. Taxonomic experts working in the respective institution, scientific employees, and curators of the botanical collections decide which taxonomic and/or nomenclatural list is used to apply the names to the specimens. Depending on the source, the cornflower has either the accepted name *Centaurea cyanus* L. (e.g. POWO 2023) or *Cyanus segetum* Hill (e.g. Euro+Med PlantBase, Greuter 2006+). In common German identification literature, the accepted genus for the cornflower changed from *Centaurea* L. to *Cyanus* Mill. (Seybold 2009; Parolly & Rohwer 2019). Even while conducting this study, the accepted genus changed in the World Flora Online contrarily to the mentioned identification literature due to the update of the *Asteraceae* family through the respective taxonomic expert network (WFO 2023). The publication date and the source are therefore important characteristics of the published names.

Several approaches unite existing publications and taxonomic experts' evaluations (Banki & al. 2019; Borsch & al. 2020; Govaerts & al. 2021) and are building large, easily accessible checklists (Godfray 2002). Additionally, a workflow that allows cross-talking between important checklists shall be established in the future (Schellenberger Costa & al. 2023). Other publications present concepts and prototypes for merging taxonomic and/or nomenclatural databases (Ytow & al. 2001; Laurenne & al. 2014; Michel & al. 2017) and for the preservation of changes throughout time (Chawuthai & al. 2016; Kohlbecker & al. 2021). Nevertheless, in our experience, it is not sufficient to assess a single taxonomic database in the day-to-day work in a botanical collection.

Checking the scientific name in various lists becomes relevant for instance, when plant material is exchanged between botanical collections working with different taxonomical sources. Some specimens might have a name in the local database that was accepted several years ago but has changed now. This can be caused by an irregular update of both the applied names in the local databases and the underlying taxonomic and/or nomenclatural data set. Existing name resolution services (Wagner 2016; Grenié & al. 2023) can check against a single taxonomic dataset, but do not allow the coexistence of different synonyms caused by the use of different databases in botanical col-

1 Institute of Botany, Ulm University, Ulm, Germany.

2 Data Science & Big Data Analytics Group, Ulm University, Ulm, Germany.

Authors for correspondence: Annika Schrumpf, annika.schrumpf@uni-ulm.de; Ansgar Scherp, ansgar.scherp@uni-ulm.de

lections. Another example is the use of plant material in teaching. A botanical garden might choose to abide to the latest version of the World Flora Online and to name the cornflower *Centaurea cyanus*. In case of German identification courses, the garden would have to keep the name *Cyanus segetum*, nevertheless, to agree with the respective identification literature. Furthermore, a botanical collection might be required to follow an older taxonomic and nomenclatural database for the curation of herbarium specimens.

We suggest accepting the coexistence of several taxonomies and supporting the concurrent search within them. We propose a mapping model for the purpose of searching within accepted names in different taxonomic and nomenclatural datasets. Our model puts the different databases next to each other without judging the databases or the resulting meta-taxonomy, or defining one of them as the ultimate truth.

For our study, we selected three representative taxonomic databases, namely The Plant List (TPL, The Plant List 2013), the World Flora Online (WFO, Borsch & al. 2020) and the Integrated Taxonomic Information System (ITIS, Integrated Taxonomic Information System 2021). We defined a lightweight mapping between the concepts in the taxonomies and created a prototypical search GUI for our system. We conducted a user study with scientific personnel from German botanical gardens to evaluate whether there is a need for the described system and whether the prototype would be a helpful tool in everyday work life.

Material and methods

We defined a mapping of existing taxonomic datasets to the Simple Knowledge Organization System (SKOS) to publish the datasets on the web but maintain them separately. The mapping was conducted on the levels of the scientific species and family names for this first version. Additional information such as the orders or subspecies could be included in an updated version. The mapping was used to interlink the datasets with a matching algorithm and visualize the contents in a graphical user interface.

Datasets

Three datasets were selected for the use in our prototype. They were chosen following discussions at the authors' institution. Although it has been static since 2013, TPL was, to our knowledge, still used as source for accepted names by botanical gardens in 2021. The WFO (WFO 2021) is used as a source for taxonomic information in botanical gardens and, as successor of TPL, there had to be an overlap between the datasets of TPL and WFO. This secured the functioning of the mapping. The similarity of TPL and WFO allowed to check whether

Table 1. Number of concepts and relations within three considered databases and added, cross-taxonomy relations within prototype.

	TPL	WFO	ITIS	Cross-taxonomy
Concepts	1,334,558	1,329,050	146,605	
Relations	1,965,401	1,984,117	216,341	1,706,646

the matching algorithm operated as it should. ITIS includes documented taxonomic information of multiple kingdoms. ITIS was chosen for this study because it was one of the first projects organizing taxonomic information online and because it is part of the GBIF Taxonomy Backbone (GBIF Secretariat 2022) and the Catalogue of Life (Banki & al. 2019).

We focused on the presented three datasets for our proof of concept and the user study. ITIS was assessed on July 12th, 2021, TPL was assessed on July 16th, 2021, for TPL, and WFO was assessed on November 22nd, 2021, respectively. About a fourth of the taxon entries were accepted names in TPL and WFO (number of taxon entries in Table 1). Other databases such as the World Checklist of Vascular Plants (WCVP, Govaerts & al. 2021) or the Leipzig Catalogue of Vascular Plants (LCVP, Freiberg & al. 2020) were not considered in this work but could be easily integrated in future versions.

Mapping the taxonomic databases to SKOS

The mapping of the taxonomies for the use in our prototype was based on the Simple Knowledge Organization System (SKOS, W3C 2009), a W3C standard to link and share taxonomic information online. SKOS was chosen as representation format due to its versatility to model taxonomies. The standard is used in various disciplines such as agriculture, social sciences or economics. Basing a wrapper on SKOS therefore has the additional benefit that other, existing SKOS-based thesauri, as databases from medical institutions or the Wikipedia, could be connected to each other and to the presented model.

The mapping modelled taxonomic entries as concepts and their relations. We performed an exact matching of the scientific names, to create cross-taxonomy relations. We chose not to apply fuzzy matching because it is not possible to automatically decide when names only contain typing errors and when the names are synonyms. Although we noticed some typing errors in taxon names, the scientific names provided by the databases were reliable and comparable for the main part of the data (see also Schellenberger Costa & al. 2023). The number of additional relations created through this matching were 1,706,646 cross-taxonomy relations (Table 1). We used Skosify, a framework to validate and enhance SKOS vocabularies (Suominen & Hyvönen 2012), to assess the quality of the mapping of the taxonomies. The mapping was conducted with the goal to have an integrated taxonomy that could be used for the search in the prototype
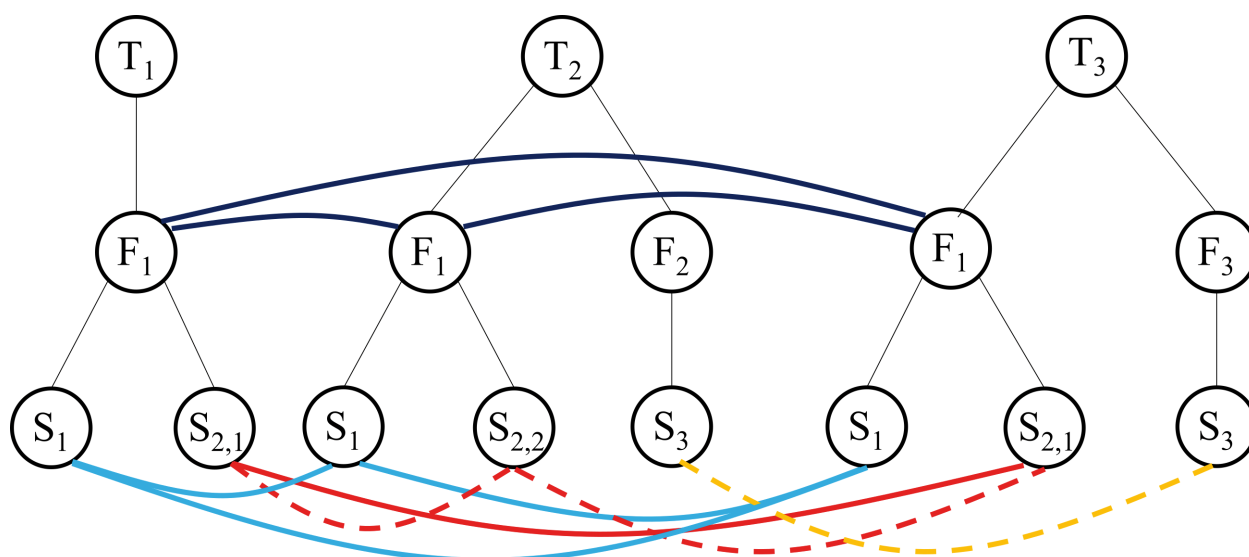
Fig. 1. Scheme of concepts, i.e. family (F_n) and species (S_n) names, in three databases (T_n) with resulting edges and synonym relations. Solid lines: edges between concepts without botanical synonyms. Dashed lines: edges between synonyms. Synonyms on species level are represented with two indices: first index represents the species; enumeration within second index represents the synonym. A concrete corresponding subgraph can be found in Supplementary Fig. S1.

and evaluate it with experts. It does not assess whether the mapping between taxa in the three taxonomic databases was correct.

The three taxonomic databases were provided as Darwin Core Archive files (Wieczorek & al. 2012). These archives consist of one or more Tab-Separated-Value (TSV) files and an XML metafile. While the TSV files store the actual taxonomic and nomenclatural information, the metafile describes which Darwin Core Terms are represented. A Darwin Core Term represents the column description of the dataset (Darwin Core Maintenance Group 2021). While all three datasets were provided in the same format, they modelled their hierarchical structure information differently. Therefore, it was required to implement one sub-parser for each of the three datasets.

The ITIS taxonomy used the Darwin Core Term dwc:*parentNameUsageID* for each entry to model a link to the entry of the next higher hierarchy level, unless the entry represented was a synonym. This allowed us to parse the ITIS dataset in a recursive way, starting at the *Kingdom* level. All synonyms of an entry newly added to our graph could be found and parsed by searching for the current accepted taxon ID in the column dwc:*acceptedNameUsageID*.

The term dwc:*parentNameUsageID* was not used in the TPL dataset, which necessitated the implementation of a different linking. The taxonomic information, i.e. kingdom, family and genus, was stored in columns inside each data row, but not as separate data rows. To parse this structure, we implemented an iterative algorithm, which added the families and genera to the graph structure and generated an ID for them. After this, we were able to parse the entries of the species, subspecies, varieties, and forms in an iterative way. Parsing the synonyms worked

the same as for the dataset of ITIS. The WFO dataset was parsed analogously to the TPL dataset.

Besides the differences in the linking structure, each of the three sub-parsers currently only took taxon entries of the kingdom *Plantae* into account. To allow future extensions of the parsers without much effort, they have been implemented in such a way that other kingdoms can be added. The three sub-parsers converted the information on the species name and taxonomic status to the tree structure (dwc:*scientificName* with dwc:*scientificNameAuthorship*, dwc:*taxonomicStatus*). These three terms were sufficient for the described task and were available in all datasets. The parser implementation allows to include more information, e.g. orders or other Darwin Core Terms, when necessary.

The merging stage combined the three taxonomies to one graph as described above. This stage was crucial to provide a cross-taxonomic graph without creating a new taxonomy and use it in the prototype to search within the taxonomies. This resulted in bidirectional edges between the taxonomies and denoted synonym relations (Fig. 1). The SKOS class skos:conceptScheme was used to model families, genera and species. This is the recommended approach to group multiple concepts and when dealing with concepts coming from multiple sources in the SKOS standard. The taxonomic structures in TPL, ITIS and WFO were modelled as skos:Concept class. The relationships between classes within a taxonomy, i.e. families and genera, were modelled by skos:narrower relations, and inversely, skos:broader, as indicated by the edges in Fig. 1 (see also Supplementary Fig. S1 for an exemplary subgraph). Related concepts across taxonomies were modelled using the skos:related property. This was applied, as shown in Fig. 1, on the level of families, gen-

era, and species. The matching was restricted to link only concepts that were within the same taxonomic rank. The developed user interface worked directly on the SKOS data.

User study

Five participants from four botanical gardens in Germany could be recruited for the user study in summer 2022. Prior to starting the interviews, the participants were provided with an informed consent form, and the goal of the study was clarified. The participants could withdraw from the study at any point in time. The goal was to find out whether taxon conflicts are an issue in the every-day work at botanical gardens, whether there is the need for a solution and whether the prototype would be a helpful tool (the questionnaire of the user study can be found in the Supplementary Table S1). The expert interviews were semi-structured to obtain comparable, qualitative feedback as well as suggestions for further improvements (Lazar & al. 2017).

The participants were given access to the web interface of the prototype. They were asked to perform representative tasks in resolving synonym relationships of a name across multiple taxonomies. The participants were also encouraged to try out own queries with other names. Goal of this usability test was to determine whether the prototype met the requirements for task appropriateness (Lazar & al. 2017), i.e. whether the tool helped to solve the challenge of finding synonymous concepts in several botanical taxonomies. The participants were asked to evaluate the suitability of the prototype for solving the tasks with questions based on ISO 9241 part 110, *Interaction principles* (International Organization for Standardization 2020) provided by Gediga & Hamborg (1999). We concluded with an open discussion, where further suggestions for improvement were solicited.

Results

Prototype for searching in multiple botanical taxonomies

Our prototype allows the users to search in the taxonomies via string matching. The internal representation and mapping of taxonomies (Fig. 1) allows to search along the different levels, i.e. family, genus and species. For each item in the result list, the taxonomic hierarchy can be requested by selecting the item descriptor. The web-based prototype provides a mid-heavy, three column layout (Fig. 2) with focus on a minimal, easy to understand design with high functionality.

The left column incorporates a search field with the corresponding result list (Fig. 2). When typing a query in the search field, the users receive a cross-taxonomy response list of possible synonyms. After selecting an item in the result list, the middle column provides a list of found taxa and possible synonym relations. An entry is represented by the scientific name and the author, the taxonomic or nomenclatural source dataset, and the taxonomic rank. In addition, the taxonomic status, provided by the taxonomic source, can be accessed.

The result list is ordered by the taxonomic status, with the accepted names up top, followed by synonyms, and not accepted names. When a user selects an item in the mid column, additional taxonomic information, i.e. the family, is provided in the right column.

Expert responses

All participants confirmed that they regularly work with taxonomies and plant material registrations, and that conflicts occur due to different scientific names in different taxonomies (the responses within the user study can be found in the Supplementary Table S1). Four of the five study participants stated that they had already searched for solutions for taxonomy conflicts within their professional lives. Three of the participants dealt with the resolution of taxonomy conflicts on a weekly basis. They would invest more time into this issue if there would be more time left in the day-to-day work. The priority of solving the issue of taxonomy conflicts for their respective gardens was rather high for four of five participants.

The WFO und TPL datasets were the most familiar for the participants. The TPL data set was only used in one of the botanical gardens; the other participants did not use it, as it is outdated. Three of the five participants stated that they were not familiar with the dataset of ITIS. Another one stated, that he/she was aware of the dataset, but rarely used it him-/herself.

The participants gave positive feedback regarding the usability of the prototype in the sense of ISO 9241 part 110. The prototype allowed to insert the search term as required and supported the participants in conducting the given tasks. Three of five participants stated that the graphic interface itself was designed and structured in a clear manner. In addition, one participant stated, that he/she found the user interface to be intuitive. All experts confirmed that the information required for the tasks were always shown at the right place on the screen. Four of the experts agreed that the presentation of the information on the screen supported them in conducting their tasks.

The main point of criticism toward the developed prototype was the missing display of the publication date of the respective scientific name. Furthermore, three of the five participants would have liked to see more structure and a better overview in the synonym list. It was also suggested to add a filter and sorting function to improve the overview of the search results. Other suggestions made by individual participants included the integration of more datasets, a function to save the results of the synonym search as a CSV-file, links to the three databases, and displaying the type of synonym relationship,
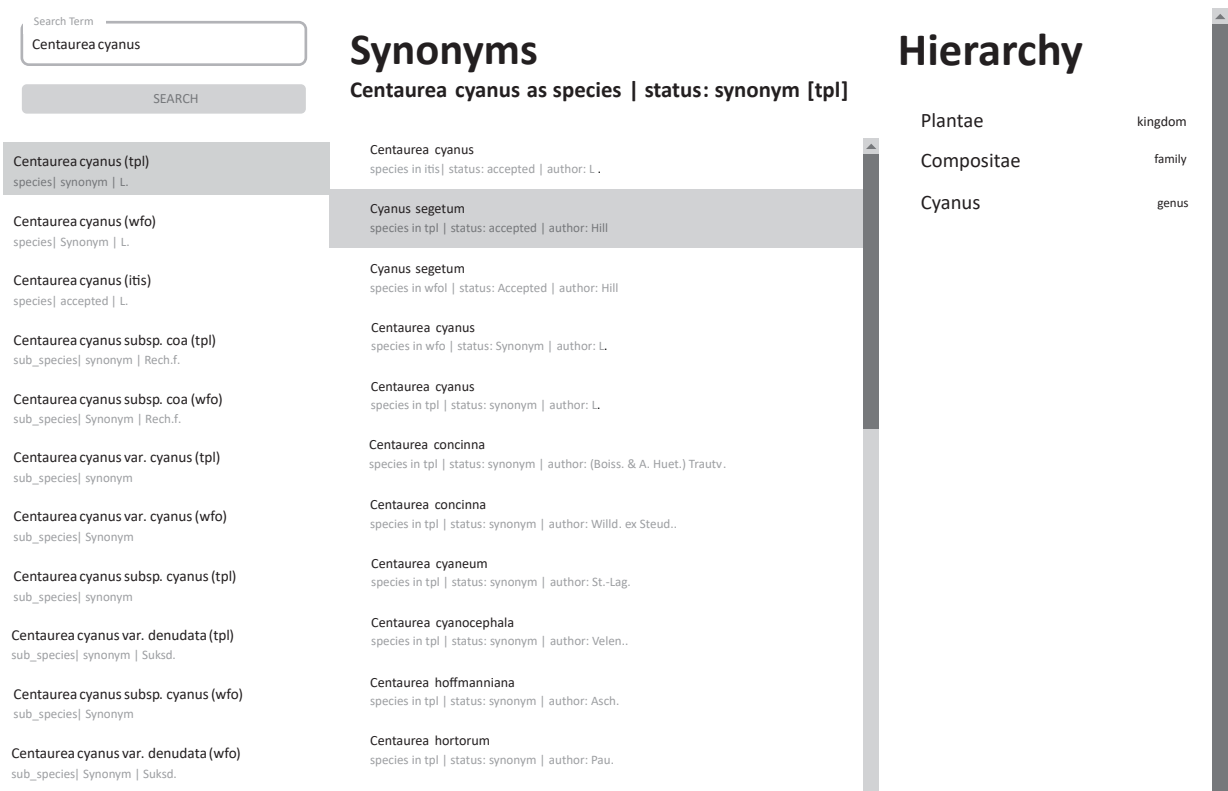
Fig. 2. Screenshot of web layout of prototype. Left column consists of search field and result list. Middle column provides list of found synonym relations for selected item. Additional information on hierarchy is displayed on right side. A screencast of the user interface has been uploaded to a GitHub repository and to YouTube (see Supplemental content online).

i.e. heterotypic or homotypic. Furthermore, two participants mentioned that the prototype could be interesting for other fields, e.g. in zoology.

Discussion

Mapping of Darwin Core to SKOS

The current parser relied on datasets that were provided in the Darwin Core Standard (Wieczorek & al. 2012). Other studies presented solutions to integrate datasets which are based on different mappings (Michel & al. 2017). This idea could be implemented into a next version of our model. This would allow integrating a plethora of additional nomenclatural and taxonomic databases. A next version of the model could also be able to trace changes in the respective datasets throughout time. It is important to consider the change in time, when assessing the accepted scientific names (Kohlbecker &al. 2021). Other studies have shown in models and prototypes how this change can be followed (Chawuthai & al. 2016; Michel & al. 2017; Kohlbecker & al. 2021).

The publication date of an (accepted) name within the respective taxonomic or nomenclatural database is a relevant information. The publication date was not displayed in the current version of the GUI. It can be, and

will be, implemented in a next version of our model, as the information is available in the Darwin Core formatted datasets. In a next version, the authors names should be included into the name matching to avoid larger amounts of mismatching. Even without the display of the publication date and the matching of the authors name, the feedback on our prototype GUI was positive; the missing publication date did, therefore, not have an impact on our proof of concept.

From a technical perspective, we have encountered some inconsistencies when working with taxonomic sources in the Darwin Core format. While Darwin Core Terms are standardized, it is not specified which Darwin Core Terms must be present in a Darwin Core Archive. This led to a situation where the structure of different taxonomies was represented by different Darwin Core Terms (Löffler & al. 2021), which complicated the development of a generalized parser. Better harmonized models consistently using the same Darwin Core Terms are needed in future.

Depending on the type of mapping, loss of information may be the result when mapping information from Darwin Core Archive files to SKOS (Michel & al. 2018). In this study, only the scientific name including the author and the taxonomic status were considered. The three mentioned types of information were sufficient to enable the taxon matching. Therefore, the other information

saved in the DWC files of the three databases was not included in the prototype. The possible information loss therefore is not a problem in this study.

Reliability of the user study

The user study was conducted with five scientific employees or curators of German botanical gardens. This guaranteed a high quality and reliability of feedback. The number of experts needed in such qualitative studies to gain sufficient and representative feedback has long been investigated (Virzi 1992; Nielsen & Landauer 1993; Gubrium & al. 2012). For human computer interaction, Virzi stated in 1992 that "80 % of the usability problems are detected with four or five subjects". Also, Nielsen & Landauer reflected on this question in 1993 and concluded that for usability testing, such as a qualitative study done here with experts, the best results with regard to cost and benefit are obtained with three to seven users. Baker & Edwards (2018) surveyed 14 social scientists and five junior scientists on this issue and concluded in their study that the recurring answer to the question "how many" is "it depends". We consider the number of five study participants as sufficient and can confirm from our experience in this user study that we won the most important new information with the first three participants.

The participants were not familiar with all of the considered databases. This did not lead to problems in understanding the tasks and the main idea of the project. Therefore, the choice of the three databases, TPL, WFO and ITIS, did not impact the study and the reception of the prototype. Ever since the user study was conducted in summer 2022, WFO has been updated and TPL is becoming less popular because it is superseded. Furthermore, other databases are used by botanical collections. Other additional taxonomic and nomenclatural databases should, therefore, be implemented in their updated versions in a future version of the prototype.

Coexistence of databases

The overall feedback from the participants of the user study on the presented prototype was positive. It was received as a useful, time-saving tool for the day-to-day work life. The participants showed through their feedback that there is a certain necessity to compare the different taxonomic and nomenclatural datasets instead of relying on a single one. This makes our prototype a valuable addition to the existing name resolution services (Wagner 2016; Grenié & al. 2023) and authoritative checklists (Schellenberger Costa & al. 2023). The Unified Taxonomic Information Service (UTIS, Unified Taxonomic Information Service 2023) intends to follow the same principal as presented in this study by enabling the search of a scientific name in several European databases. The GBIF Checklist Bank (GBIF Checklist Bank 2023) is an API that allows to look up a scientific name

in all checklists in the GBIF network. The information accessible via this API could serve as fruitful source for a more extensive, updated version of our mapping model.

Our model allows the coexistence of several taxonomic and nomenclatural databases without judging the accepted names of the individual databases or the resulting meta-taxonomy. Mapping the taxonomies to SKOS has several advantages, when compared to the existing systems. The standard is used in various disciplines; therefore, existing SKOS-based thesauri could be connected to each other and to the presented mapping of the botanical taxonomies. Furthermore, data sets from different botanical collections could be brought together, without having to run the names through a name resolution service and adjusting synonyms according to one list. The names in each data set could remain as entered, but the different entries of synonyms could still be matched by our model.

This opens the discussion on whether there must be (or actually can be) a single, correct taxonomic database, which serves as the ultimate truth. In our opinion, there are certain cases in which it is more practicable to see several datasets in an equal coexistence. Our presented study and prototype are a first idea of how this coexistence could look like.

Author contributions

ASchrumpf and AScherp designed this study, wrote the manuscript and supervised the project. MK and PS developed the prototype and conducted the user study.

Acknowledgements

References

Baker S. & Edwards R. 2018: How many qualitative interviews is enough? Expert voices and early career reflections on sampling and cases in qualitative research. – Southampton: National Centre for Reseach Methods. https://eprints.ncrm.ac.uk/id/eprint/2273

Banki O., Hobern D., Döring M. & Remsen D. 2019: Catalogue of Life Plus: a collaborative project to complete the checklist of the world's species. – Biodivers. Inform. Sci. & Standards **3**(e37652). https://doi.org/10.3897/biss.3.37652

Borsch T., Berendsohn W., Dalcin E., Delmas M., Demissew S., Elliott A., Fritsch P. & al. 2020: World Flora Online: placing taxonomists at the heart of a definitive and comprehensive global resource on the world's plants. – Taxon **69:** 1311–1341. https://doi.org/10.1002/tax.12373

Chawuthai R., Takeda H., Wuwongse V. & Jinbo U. 2016: Presenting and preserving the change in taxonomic knowledge for linked data. – Semantic Web **7:** 589–616. https://doi.org/10.3233/SW-150192

Darwin Core Maintenance Group 2021: List of Darwin Core terms. Biodiversity information standards (TDWG). – Published at https://dwc.tdwg.org/list/2021-07-15

Freiberg M., Winter M., Gentile A., Zizka A., Muellner-Riehl A. N., Weigelt A. & Wirth C. 2020: LCVP, the Leipzig catalogue of vascular plants, a new taxonomic reference list for all known vascular plants. – Scientific Data **7**(416). https://doi.org/10.1038/s41597-020-00702-z

GBIF Checklist Bank 2023: Published at https://api.gbif.org/v1/species [accessed 19 Sep 2023].

GBIF Secretariat 2022: GBIF backbone taxonomy. Checklist dataset. – Published at https://doi.org/10.15468/39omei via GBIF.org [accessed 17 May 2023].

Gediga G., Hamborg K.-C. & Düntsch, I. 1999: The IsoMetrics usability inventory: an operationalization of ISO 9241-10 supporting summative and formative evaluation of software systems. – Behav. Inform. Technol. **18:** 151–164. https://doi.org/10.1080/014492999119057

Godfray H. C. J. 2002: Challenges for taxonomy. – Nature **417:** 17–19. https://doi.org/10.1038/417017a

Govaerts R., Nic Lughadha E., Black N., Turner R. & Paton A. 2021: The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. – Sci. Data **8**(215). https://doi.org/10.1038/s41597-021-00997-6

Grenié M., Berti E., Carvajal-Quintero J., Dädlow G. M. L., Sagouis A. & Winter M. 2023: Harmonizing taxon names in biodiversity data: a review of tools, databases and best practices. – Meth. Ecol. Evol. **14:** 12–25. https://doi.org/10.1111/2041-210X.13802

Greuter W. 2006+: *Compositae* (pro parte majore). – In: Greuter W. & Raab-Straube E. von (ed.), *Compositae*. Euro+Med PlantBase – the information resource for Euro-Mediterranean plant diversity. – Published at https://www.europlusmed.org/cdm_dataportal/taxon/db14424b-b2db-4b35-a76f-177b5d5be984 [accessed 23 Mar 2023].

Gubrium J. F., Holstein J. A., Marvasti A. B. & McKinney K. D. 2012: The SAGE handbook of interview research: the complexity of the craft. – Thousand Oaks: SAGE Publications Inc. https://doi.org/10.4135/9781452218403

Integrated Taxonomic Information System (ITIS) 2021: Published at https://doi.org/10.5066/F7KH0KBK [accessed 12 Jul 2021].

International Organization for Standardization 2020: Ergonomics of human-system interaction. Part 110: interaction principles. Ed. 2. – Published at https://www.iso.org/standard/75258.html [accessed 25 May 2023].

Kohlbecker A., Karam N., Paschke A. & Güntsch A. 2021: Preserving taxonomic change and subsequent taxon relationships over time. – CEUR Workshop Proceedings **2969.**

Laurenne N., Tuominen J., Saarenmaa H. & Hyvönen E. 2014: Making species checklists understandable to machines – a shift from relational databases to ontologies. – J. Biomed. Semantics **5**(40). https://doi.org/10.1186/2041-1480-5-40

Lazar J., Feng J. H. & Hochheiser H. 2017: Research methods in human-computer interaction, ed. 2. – Cambridge: Morgan Kaufmann. https://www.sciencedirect.com/book/9780128053904/research-methods-in-human-computer-interaction

Löffler F., Wesp V., König-Ries B. & Klan F. 2021: Dataset search in biodiversity research: do metadata in data repositories reflect scholarly information needs? – PLoS ONE **16**(e0246099). https://doi.org/10.1371/journal.pone.0246099

Michel F., Faron-Zucker C., Tercerie S. & Olivier G. 2018: Modelling biodiversity linked data: pragmatism may narrow future opportunities. – Biodivers. Inform. Sci. Standards **2**(e26235). https://doi.org/10.3897/biss.2.26235

Michel F., Gargominy O., Tercerie S. & Faron-Zucker C. 2017: A model to represent nomenclatural and taxonomic information as linked data. Application to the French taxonomic register, TAXREF. – CEUR Workshop Proceedings **1933.**

Nielsen J. & Landauer T. K. 1993: A mathematical model of the finding of usability problems. – Pp. 206–213 in: CHI '93. Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems. – New York: Association for Computing Machinery. https://doi.org/10.1145/169059.169166

Parolly G. & Rohwer J. G. (ed.) 2019: Schmeil-Fitschen – Die Flora Deutschlands und angrenzender Länder, ed. 97. – Wiebelsheim: Quelle & Meyer Verlag.

POWO 2023: *Centaurea cyanus* L. – In: Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew – Published at https://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:190310-1 [accessed 17 May 2023].

Schellenberger Costa D., Boehnisch G., Freiberg M., Govaerts R., Grenié M., Hassler M., Kattge J., Muellner-Riehl A. N., Rojas Andrés B. M., Winter M., Watson M., Zizka A. & Wirth C. 2023: The big four of plant taxonomy – a comparison of global check-

lists of vascular plant names. – New Phytol. **240:** 1687–1702. https://doi.org/10.1111/nph.18961

Seybold S. (ed.) 2009: Schmeil-Fitschen – Flora von Deutschland und angrenzender Länder, ed. 94. – Wiebelsheim: Quelle & Meyer Verlag.

Suominen O. & Hyvönen E. 2012: Improving the quality of SKOS vocabularies with Skosify. – Pp. 383–397 in: Teije A., Völker J., Handschuh S., Stuckenschmidt H., d'Acquin M., Nikolov A., Aussenac-Gilles N. & Hernandez N. (ed.), Knowledge engineering and knowledge management. – Berlin & Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-642-33876-2_34

The Plant List 2013: Version 1.1 – Published at http://www.theplantlist.org/ [accessed 16 Jul 2021].

Thiers B. M. 2023+ [continuously updated]: Index herbariorum. – Published at https://sweetgum.nybg.org/science/ih/ [accessed 23 Mar 2023].

Unified Taxonomy Information Service (UTIS) 2023: Published at https://cybertaxonomy.org/eu-bon/utis/1.3/doc.html [accessed 19 Sep 2023].

Virzi R. A. 1992: Refining the test phase of usability evaluation: how many subjects is enough? – Human Factors **34:** 457–468. https://doi.org/10.1177/001872089203400407

W3C 2009: SKOS Simple Knowledge Organization System reference. – Published at https://www.w3.org/TR/skos-reference/ [accessed 19 Oct 2022].

Wagner V. 2016: A review of software tools for spellchecking taxon names in vegetation databases. – J. Veg. Sci. **27:** 1323–1327. https://doi.org/10.1111/jvs.12432

WFO 2021: World Flora Online. Version 2019.10. – Published at https://www.worldfloraonline.org/ [accessed 22 Nov 2021].

WFO 2023: *Centaurea cyanus* L. – Published at https://www.worldfloraonline.org/taxon/wfo-0000008997 [accessed 23 Mar 2023].

Wieczorek J., Bloom D., Guralnick R., Blum S., Döring M., Giovanni R., Robertson T. & Vieglais D. 2012: Darwin Core: an evolving community-developed biodiversity data standard. – PLoS ONE **7**(e29715). https://doi.org/10.1371/journal.pone.0029715

Ytow N., Morse D. R. & Roberts D. M. 2001: Nomencurator: a nomenclatural history model to handle multiple taxonomic views. – Biol. J. Linn. Soc. **73:** 81–98. https://doi.org/10.1111/j.1095-8312.2001.tb01348.x

Supplemental content online

See https://doi.org/10.3372/wi.53.53308

Supplementary Fig. S1. Excerpt of the graph modelling *Cyanus segetum* within the TPL database and its relation to a synonym in TPL and the related entry in WFO in Turtle syntax. The entries mapped by skos:broader are the respective genera.

Supplementary Table S1. Questionnaire and answers from the user study. The originally asked questions in German are attached at the end. Questions that could lead to the identification of the study participants were deleted from this publication.

The source code and the SKOS graph are made publicly available under a MIT license at GitHub: https://github.com/LeBraveLittleToaster/taxonomic-coexistence

A screencast of the user interface has been uploaded to the GitHub repository and to YouTube: https://www.youtube.com/watch?v=m9eYlbOl8B4