Nomenclature over 5 years in TaxonWorks: Approach, implementation, limitations and outcomes

Matthew Yoder[‡], Dmitry A Dmitriev[‡]

[‡] University of Illinois at Urbana-Champaign, Champaign, United States of America

Corresponding author: Matthew Yoder (diapriid@gmail.com)

Abstract

We are now over four decades into digitally managing the names of Earth's species. As the number of federating (i.e., software that brings together previously disparate projects under a common infrastructure, for example TaxonWorks) and aggregating (e.g., International Plant Name Index, Catalog of Life (CoL)) efforts increase, there remains an unmet need for both the migration forward of old data, and for the production of new, precise and comprehensive nomenclatural catalogs. Given this context, we provide an overview of how TaxonWorks seeks to contribute to this effort, and where it might evolve in the future.

In TaxonWorks, when we talk about governed names and relationships, we mean it in the sense of existing international codes of nomenclature (e.g., the International Code of Zoological Nomenclature (ICZN)). More technically, nomenclature is defined as a set of objective assertions that describe the relationships between the names given to biological taxa and the rules that determine how those names are governed. It is critical to note that this is not the same thing as the relationship between a name and a biological entity, but rather nomenclature in TaxonWorks represents the details of the (governed) relationships between names. Rather than thinking of nomenclature as changing (a verb commonly used to express frustration with biological nomenclature), it is useful to think of nomenclature as a set of data points, which grows over time. For example, when synonymy happens, we do not erase the past, but rather record a new context for the name(s) in question. The biological concept changes, but the nomenclature (names) simply keeps adding up.

Behind the scenes, nomenclature in TaxonWorks is represented by a set of nodes and edges, i.e., a mathematical graph, or network (e.g., Fig. 1). Most names (i.e., nodes in the network) are what TaxonWorks calls "protonyms," monomial epithets that are used to construct, for example, bionomial names (not to be confused with "protonym" sensu the ICZN). Protonyms are linked to other protonyms via relationships defined in NOMEN, an ontology that encodes governed rules of nomenclature. Within the system, all data, nodes

©

.

and edges, can be cited, i.e., linked to a source and therefore anchored in time and tied to authorship, and annotated with a variety of annotation types (e.g., notes, confidence levels, tags). The actual building of the graphs is greatly simplified by multiple user-interfaces that allow scientists to review (e.g. Fig. 2), create, filter, and add to (again, not "change") the nomenclatural history.

As in any complex knowledge-representation model, there are outlying scenarios, or edge cases that emerge, making certain human tasks more complex than others. TaxonWorks is no exception, it has limitations in terms of what and how some things can be represented. While many complex representations are hidden by simplified user-interfaces, some, for example, the handling of the ICZN's Family-group name, batch-loading of invalid relationships, and comparative syncing against external resources need more work to simplify the processes presently required to meet catalogers' needs.

The depth at which TaxonWorks can capture nomenclature is only really valuable if it can be used by others. This is facilitated by the application programming interface (API) serving its data (https://api.taxonworks.org), serving text files, and by exports to standards like the emerging Catalog of Life Data Package. With reference to real-world problems, we illustrate different ways in which the API can be used, for example, as integrated into spreadsheets, through the use of command line scripts, and serve in the generation of public-facing websites.

Behind all this effort are an increasing number of people recording help videos, developing documentation, and troubleshooting software and technical issues. Major contributions have come from developers at many skill levels, from high school to senior software engineers, illustrating that TaxonWorks leads in enabling both technical and domain-based contributions. The health and growth of this community is a key factor in TaxonWork's potential long-term impact in the effort to unify the names of Earth's species.

Presenting author

Matthew Yoder

Presented at

TDWG 2021

Acknowledgements

Many people have contributed to refining TaxonWorks and the cloud of resources surrounding it, see https://github.com/SpeciesFileGroup/taxonworks/blob/development/CONTRIBUTORS.md for more.

Conflicts of interest

References

- Dmitriev DA (2003) 3i World Auchenorrhyncha Database. http://dmitriev.speciesfile.org/. Accessed on: 2021-9-13.

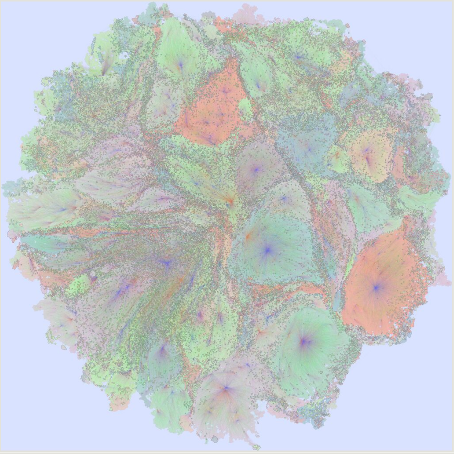
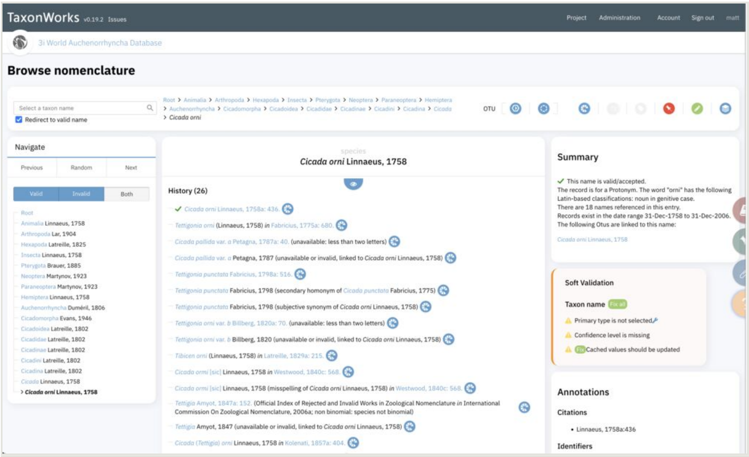


Figure 1.

Visualizing nomenclature in TaxonWorks as a network (graph). Over 700K data points corresponding to over 120K names, their relationships, status, and citations in the 3i World Auchenorrhyncha Database (Dmitriev 2003). Encoded in the DOT graph description language and formatted in Graphviz.



Browse nomenclature

☑ Redirect to valid name

Root > Animalia > Arthropoda > Hexapoda > Insecta > Pterygota > Neoptera > Paraneoptera > Hemiptera > Auchenorrhyncha > Cicadomorpha > Cicadoidea > Cicadidae > Cicadinae > Cicadini > Cicadina > *Cicada* > *Cicada orni*

Navigate

- Root
- Animalia Linnaeus, 1758
- Arthropoda Lar, 1904
- Hexapoda Latreille, 1825
- Insecta Linnaeus, 1758
- Pterygota Brauer, 1885
- Neoptera Martynov, 1923
- Paraneoptera Martynov, 1923
- Hemiptera Linnaeus, 1758
- Auchenorrhyncha Duméril, 1806
- Cicadomorpha Evans, 1946
- Cicadoidea Latreille, 1802
- Cicadidae Latreille, 1802
- Cicadinae Latreille, 1802
- Cicadini Latreille, 1802
- Cicadina Latreille, 1802
- *Cicada* Linnaeus, 1758
- > ***Cicada orni* Linnaeus, 1758**

species

Cicada orni Linnaeus, 1758

History (26)

- ✓ *Cicada orni* Linnaeus, 1758a: 436.
- *Tettigonia orni* (Linnaeus, 1758) *in* Fabricius, 1775a: 680.
- *Cicada pallida* var. *a* Petagna, 1787a: 40. (unavailable: less than two letters)
- *Cicada pallida* var. *a* Petagna, 1787 (unavailable or invalid, linked to *Cicada orni* Linnaeus, 1758)
- *Tettigonia punctata* Fabricius, 1798a: 516.
- *Tettigonia punctata* Fabricius, 1798 (secondary homonym of *Cicada punctata* Fabricius, 1775)
- *Tettigonia punctata* Fabricius, 1798 (subjective synonym of *Cicada orni* Linnaeus, 1758)
- *Tettigonia orni* var. *b* Billberg, 1820a: 70. (unavailable: less than two letters)
- *Tettigonia orni* var. *b* Billberg, 1820 (unavailable or invalid, linked to *Cicada orni* Linnaeus, 1758)
- *Tibicen orni* (Linnaeus, 1758) *in* Latreille, 1829a: 215.
- *Cicada ormi* [sic] Linnaeus, 1758 *in* Westwood, 1840c: 568.
- *Cicada ormi* [sic] Linnaeus, 1758 (misspelling of *Cicada orni* Linnaeus, 1758) *in* Westwood, 1840c: 568.
- *Tettigia* Amyot, 1847a: 152. (Official Index of Rejected and Invalid Works in Zoological Nomenclature *in* International Commission On Zoological Nomenclature, 2006a; non binomial: species not binomial)
- *Tettigia* Amyot, 1847 (unavailable or invalid, linked to *Cicada orni* Linnaeus, 1758)
- *Cicada (Tettigia) orni* Linnaeus, 1758 *in* Kolenati, 1857a: 404.

Summary

✓ This name is valid/accepted.
The record is for a Protonym. The word "orni" has the following Latin-based classifications: noun in genitive case.
There are 18 names referenced in this entry.
Records exist in the date range 31-Dec-1758 to 31-Dec-2006.
The following Otus are linked to this name:

Cicada orni Linnaeus, 1758

Soft Validation

Taxon name

- ⚠ Primary type is not selected
- ⚠ Confidence level is missing
- ⚠ Cached values should be updated

Annotations

Citations

- Linnaeus, 1758a:436

Identifiers

Figure 2.

Screenshot of the "Browse nomenclature" interface in TaxonWorks, showing the nomenclatural history of a name used for cicadas.