

TraitDB: Web application database of phenotypic trait data.

Dan Leehr¹ and Mercedes Gosby¹

¹ National Evolutionary Synthesis Center

DOI: [10.21105/joss.01201](https://doi.org/10.21105/joss.01201)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Submitted: 20 July 2018

Published: 27 March 2019

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

Summary

TraitDB is a web application that synthesizes phenotypic trait datasets from varying structures and formats. It organizes data into a consistent format and taxonomy, facilitating querying, subsetting, and sharing over a range of user criteria. While many other software packages exist to store structured and unstructured relational data, TraitDB is designed to meet the needs of research teams by fitting into their data gathering practices. Commonly-used packages like spreadsheets and SQL databases may excel at direct ingest, storage, and management of homogeneous datasets, but fall short dealing with real-world heterogeneous data structures and formats. They require dedicated effort to build and adapt complex processes in order to maintain data integrity.

Through work with multiple [NESC](#) working groups, we observed and supported real-world efforts to build a database of traits by collecting and organizing thousands of trait data observations. We evaluated software including [OpenRefine](#), [SQLite Manager](#), and [mx](#) that aim to address many of these challenges.

While these software are capable of enforcing complex data models, we found that individual researchers were most productive tabulating data from specimens or literature into local spreadsheets rather than an online application. At this level, they needed frictionless data entry of the records they had available, rather than tedious interactions. After collecting their data, they needed a robust process to validate, reconcile, and ingest records into the larger body.

Through collaboration, and iteration, we designed and built TraitDB to serve the use cases of these researchers. Primarily, TraitDB aids ingest of datasets from different researchers working to synthesize a single, uniform dataset. Data ingest is driven by a flexible YAML template system, managed by the group of researchers. The process provides validation and observation-level feedback on input CSV datasets. It indicates which fields are required, warns of duplicate records, and performs validation on categorical and continuous variables. Prior to ingesting the data, TraitDB provides a structured report of warnings and requires that any errors are corrected.

Once ingested, TraitDB provides several mechanisms for querying, browsing, subsetting, and fetching data via its web interface. It implements simple summary calculations of trait data over taxonomic levels, and provides CSV download of synthesized datasets for further analysis.

TraitDB has been used to gather and produce datasets hosted by the [Tree of Sex](#).

Acknowledgements

TraitDB is a project of [National Evolutionary Synthesis Center](#)