

## README how to best use data released from the TRY database

- a) Generalities, data structure and recommendations for data extraction and import
- b) Column Headers
- c) Recommendations in the case of publication

Please note: categorical traits for many species (plant growth form, woodiness, leaf type, etc.) and information about the measurement locations (soil, climate, biome-type) is publicly available at the TRY website: <http://www.try-db.org/TryWeb/Data.php>

### a) Generalities and data structure

The data are provided as zipped text files (.txt) with ASCII Latin-1 encoding (Linux 64-bit operating system). Columns are tab-delimited. The first row contains the header. The decimal point is given as a point (e.g. 0.123) not as a comma.

- 1) You receive only the trait values for which you have the permissions from data contributors.
- 2) A trait may consist of different sub-groups, which have been called "Data", given as DataName and DataID.
- 3) You receive the trait entries and all covariates related to the respective trait entries. The covariates you receive may not always make sense in the context of your trait data.
- 4) You receive the original values of traits and covariates (Column: OrigValueStr). The units of some trait data and the geographical references have been standardized. In these cases you additionally receive the standardized values (Column: StdValue and Unit\_1\_UnitName (unit of standardized values)). The standardized values of traits and geographical references (latitude and longitude) are checked for general consistency.
- 5) Each row in the data table contains an entry in the database (identified by ObsDataID). This entry is categorized either as trait – in this case it has a TraitID - or as covariate – in this case it has no TraitID.
- 6) The "ObservationID" relates database entries (measurements of traits and covariates) on the same entity to observations (e.g. two traits measured on the same leaf). **Different rows in the data table with the same ObservationID are directly related to each other.**
- 7) The data you receive may contain data from plants grown in natural environment and from plants grown under experimental conditions. If plants have grown under experimental conditions this is reported as a covariate entry (related to the trait entry via the ObservationID; same for mature and juvenile plants).
- 8) The data may contain duplicates, if the same data have been contributed to TRY from different contributors. If we have identified an entry as duplicate you will find the ID of the original entry in OrigObsDataID.
- 9) A general consistency check for the standardized values of traits and geographic coordinates (Column: StdValue) has been applied. For standardized trait values we provide the largest distance of the trait record from either species-mean, genus-mean, family mean or the mean of all data in terms of respective Standard deviation (ErrorRisk). For Example, an error risk value of 3 indicates that this trait record has a maximum distance of 3 Standard deviations from any of the above mentioned mean values. Error risk > 4 may indicate likely

problems in the data (similar to Kattge et al. 2011 Global Change Biology, supplementary material). Geographic coordinates are standardized to longitude and latitude values in decimal format and checked against ESA CCI Land Cover Map of Global Water Bodies (<https://www.esa-landcover-cci.org/?q=node/162>).

10) Data extraction and import: You should be able to extract the zip file in any operating system. You can read the data file (.txt) with a text editor. It should also seamlessly open in Excel. However, in some cases the file may contain too many rows to be accessible via Excel.

To import the data into R, we recommend the 'fread' function, which also works well for large files (>100000 traits). In case of large files first install a 64-bit version of R. Then download the 'data.table' package from an R repository and make it available via 'require' and finally import the data:

```
require(data.table)
TRYdata <- fread("file.txt", header = T, sep = "\t", dec = ".", quote = "", data.table = T)
```

'TRYdata': example name of the R data frame into which the data will be imported; 'file.txt': path and name of the file to be imported; 'header = T': the file has a header; 'sep = "\t": columns are tab delimited, 'dec = "."': decimals are given by a point; 'quote = ""': text within a column is not characterized by an additional sign.

After import into R you can check the data:

dim(TRYdata): returns the number of rows and columns (number of columns should be 28)

ls(TRYdata): should return the names of headers in alphabetical order

head(TRYdata): should return the header and first rows

tail(TRYdata): should return the header and last rows

Please do not hesitate to contact us in case you have questions with respect to the data, if you have problems importing the data, or if you find inconsistencies:

[jkattge@bgc-jena.mpg.de](mailto:jkattge@bgc-jena.mpg.de); [boenisch@bgc-jena.mpg.de](mailto:boenisch@bgc-jena.mpg.de)

## b) Column headers

Column	Comment
1. LastName	Surname of data contributor
2. FirstName	First name of data contributor
3. DatasetID	Unique identifier of contributed dataset
4. Dataset	Name of contributed dataset
5. SpeciesName	Original name of species
6. AccSpeciesID	Unique identifier of consolidated species name
7. AccSpeciesName	Consolidated species name
8. ObservationID	Unique identifier for each observation
9. ObsDataID	Unique identifier for each record
10. TraitID	Unique identifier for traits (only if the record is a trait)
11. TraitName	Name of trait (only if the record is a trait)
12. DataID	Unique identifier for each sub-trait or context information
13. DataName	Name of sub-trait or context information
14. OriginalName	Original Name of sub-trait or context information
15. OrigValueStr	Original value as text string
16. OrigUnitStr	Original unit as text string
17. ValueKindName	Value kind (single measurement, mean, median, etc.)

18. OrigUncertaintyStr	Original uncertainty as text string
19. UncertaintyName	Kind of uncertainty (standard deviation, standard error,...)
20. Replicates	Count of replicates
21. StdValue	Standardized value: available for standardized traits
22. StdUnit	Standard unit: available for standardized traits
23. RelUncertaintyPercent	Relative uncertainty in %
24. OrigObsDataID	Unique identifier for duplicate entries
25. ErrorRisk	Indication for outliers: distance to mean in standard deviations
26. Reference	Reference to be cited if trait record is used in analysis
27. Comment	Explanation for the OriginalName in the contributed dataset

### **c) Recommendations in case of publication**

The intellectual property guidelines of TRY provide the following suggestions:

- 1) Contact data contributors and database management for collaboration and offer authorship, if this has been agreed at data request. You can find the information in the Request PI Center on the Data Portal of the TRY website ([try-db.org/TryWeb/dp.php](http://try-db.org/TryWeb/dp.php)).
- 2) The TRY standard reference and the references of individual datasets, as indicated by the custodians, should be cited in the methods section. You can find the information in the Request PI Center on the Data Portal of the TRY website ([try-db.org/TryWeb/dp.php](http://try-db.org/TryWeb/dp.php)).
- 3) Acknowledge the TRY initiative and its funding organisations:

The study has been supported by the TRY initiative on plant traits (<http://www.try-db.org>). The TRY initiative and database is hosted, developed and maintained by J. Kattge and G. Boenisch (Max Planck Institute for Biogeochemistry, Jena, Germany). TRY is currently supported by Future Earth/bioDISCOVERY and the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig.

Based on first experiences we would like to explain the suggestions in more detail:

---

#### 1) Invitation for collaboration and authorship

In the context of the updated TRY intellectual property guidelines data contributors and the TRY database management indicate at data request if they would like to be involved in the given analyses. Please respect these conditions and contact respective data contributors and database management, when the first version of your manuscript is suitable for comments and edits.

Database management:

Gerhard Boenisch: [boenisch@bgc-jena.mpg.de](mailto:boenisch@bgc-jena.mpg.de)  
 Jens Kattge: [jkattge@bgc-jena.mpg.de](mailto:jkattge@bgc-jena.mpg.de)

---

#### 2) Citation of references

The TRY intellectual property guidelines suggest that the TRY standard reference (Kattge et al. 2011, Global Change Biology) and the references of individual datasets, as indicated by

the custodians, should be cited in the methods section of any product that involves the dataset.

According to recent developments with respect to citation guidelines, the original publications, under which the data have been published should be cited as well (see: Joint Declaration of Data Citation Principles; <https://www.force11.org/datacitation> and San Francisco Declaration on Research Assessment; [http:// am.ascb.org/dora/](http://am.ascb.org/dora/)). You can find these original References in the data release you received from TRY under DataID 113 (Original Reference).

We request that the references of datasets should be cited in the main manuscript - not in supplementary material - because citations in supplementary material are not evaluated by the relevant literature databases, e.g. Web of Science, and thus do not provide reward for data contribution. You received the references to be cited in the context of the data release. Else you can as well find them in the Request PI Center on the Data Portal of the TRY website ([try-db.org/TryWeb/dp.php](http://try-db.org/TryWeb/dp.php)).

---

### 3) Acknowledgements

The TRY intellectual property guidelines suggest the following text:

The study has been supported by the TRY initiative on plant traits (<http://www.try-db.org>). The TRY initiative and database is hosted, developed and maintained by J. Kattge and G. Boenisch (Max Planck Institute for Biogeochemistry, Jena, Germany). TRY is currently supported by Future Earth/bioDISCOVERY and the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig.

---

If data via TRY have made a decidedly minor contribution to the manuscript, offering authorship and citing original references in the main part of the manuscript may not be appropriate. However, these exceptions need to be discussed with the TRY database management in each case.

Apart from these general requirements of authorship, citation and acknowledgement, **please respect all individual agreements with data contributors made at the time of data requests!**