The Biofuel Ecopysiological Traits and Yields Database: Database Description and User's Guide Version 1.0

David LeBauer, Dan Wang, Patrick Mulrooney, Mike Dietze

Please cite this document regarding the implementation and structure of BETY:

David LeBauer, Dan Wang, Patrick Mulrooney, Mike Dietze. 2011. The Biofuel Ecopysiological Traits and Yields Database: Database Description and User's Guide, Version 1.0

Please cite the use of data contained in BETY:

David LeBauer, Dan Wang, and Michael Dietze, 2010. Biofuel Ecophysiological Traits and Yields Database Version 1.0. Energy Biosciences Institute, Urbana, IL `http://ebi-forecast.igb.uiuc.edu/bety`

Contents

List of Tables

List of Figures

1 Quick Start

Open web interface: ebi-forecast.igb.uiuc.edu/bety/
Download data: subsection 4.3.
Enter data: see the Data Entry Workflow.
Read about table contents: see Table 2 and section 5.
View summary of core tables and relationships : Figure 1
View comprehensive schema, Figure 3 .

2 Background

A major motivation of the biofuel industry is to reduce greenhouse gas emissions by providing ecologically and economically sustainable sources of fuel and dependence on fossil fuel. The goal of this database is to provide a clearinghouse of existing research on potential biofuel crops, to provide a source of data on plant ecophysiological traits and yields, and to present ecosystem scale re-analysis and forecasts that can support the agronomic, ecological, policy, and economic aspects of the biofuel industry. This database will facilitate the scientific advances and assesments that this transition will require.

3 Introduction

This document describes the purpose, design, and use of the Biofuel Ecophysiological Traits and Yields database (BETYdb). BETYdb is a database of plant trait and yield data that supports research, forecasting, and decision making associated with the development and production of cellulosic biofuel crops. While the content of BETYdb is agronomic, the structure of the database itself is general and can therefore be used more generally for ecosystem studies.

BETY-db can be accessed online at ebi-forecast.igb.uiuc.edu/bety/. For developers interested in the database description in SQL syntax, e.g. to explore, create, and modify its structure, the betydb_schema is available. or further information about the proceedures that are used to enter data into the database, see the accmpanying Data Entry Workflow.

3.1 Objectives

The objectives of this database are to allow other users access data that has been collected from previously published and ongoing research in a consistent format, and to provide a streamlined interface that allows users to enter their own data. These objectives will support specific research and collaboration, advance agricultural practices, and inform policy decisions. Specifically, BETYdb supports the following uses:

1. Carry out statistical analyses to explore the relationships between traits
2. Identify differences among species and functional groups
3. Access BETY-db from simulation models to look up values for traits and parameter
4. Identify gaps in knowledge about biofuel crop traits and model parameters to aid rational planning of research activities

BETYdb provides a central clearinghouse of biofuel crop physiological traits and yields in a consitently organized framework that simplifies the use of these data for further analysis and interpretation. Scientific applications include the development, assessment, and prediction of crop yields and ecosystem services in biofuel agroecosystems. The database directly supports parameterization and validation of ecological, agronomic, engineering, and economic models. The initial target end-users of BETY-db version 1.0 are users within EBI who aim to support sustainable biofuel production through statistical analysis and ecological modeling. By streamlining the process of data summary, we hope to inspire new scientific perspectives on biofuel crop ecology that are based on a comprehensive evaluation of available knowledge.

Published data and analyses will be provided to other scientists and the public in an easy to understand, interactive web front end to the database.

4 Scope

The database contains trait, yield, and ecosystem service data. Because all plants have the potential to be used as biofuel feedstock, BETYdb supports data from all plant species. In practice, the species included in the database reflect available data and the past and present research interests of contributors. Trait and yield data are provided at the level of species, with cultivar and clone information provided where available.

The yield data not only includes end of season harvestable yield, but also includes measurements made over the course of the growing season. These yield data are useful in the assessment of historically observed crop yields, and they can also be used in the validation of plant models. Yield data includes peak biomass, harvestable biomass, and the biomass of the crop throughout the growing season.

The trait data represent phenotypic traits; these are measurable characteristics of an organism. The primary objective of the trait data is to allow researchers to model second generation biofuel crops such as Miscanthus and Switchgrass. In addition, these data enable evaluation of new plant species as potenial biofuel crops. Ecosystem service data reflect ecosystem-level observations, and these data are included in the traits table.

4.1 Data Content

BETYdb includes data obtained through extensive literature review of target species in addition to data collected from the Energy Farm at the University of Illinois, and by our collaborators. The BETYdb database contains trait and yield data for a wide range of plant species so that it is possible to estimate the distribution of plant traits for broad phylogenetic groups and plant functional types.

BETYdb contains data from intensive efforts to find data for specific species of interest as well as from previous plant trait and yield syntheses, and other databases.Most of the data currently in the database is from plant groups that are the focus of our current research (Table 1). These species include perennial grasses, such as Miscanthus (*Miscanthus sinensis*) Switchgrass (*Panicum virgatum*), and sugarcane (*Saccharyn* spp.). BETY also includes short-rotation woody species, including poplar (*Populus* spp.) and willow (*Salix* spp.) and a group of species that are being evaluated at the energy farm as novel woody crops. In addition to these herbaceous species, we are collecting data from a species in an experimental low-input, high diversity prairie.

4.2 Design

BETYdb is a relational database that comprehensively documents available trait and yield data from diverse plant species (Figure 1). The underlying structure of BETY-db is designed to support meta-analysis and ecological modeling. A key feature is the PFT (plant functional type) table which allows a user to group species for analysis. On top of the database, we have created a web-portal that targets a larger range of end users, including scientists, agronimists, foresters, and those in the biofuel industry.

4.3 Data Access

Data is made available for analysis after it is submitted and reviewed by a database admistrator. These data are suitable for basic scientific research and modeling. All reviewed data are made publicly available after publication to users of BETY-db who are conducting primary research. Access to these raw data is provided to users based on affiliation and contribution of data.

Data can be downloaded as a `.csv=` file, and data from previously published syntheses can be downloaded without login. For example, to download all of the Switchgrass (*Panicum virgatum* L.) yield data,

1. Open the BETY homepage ebi-forecast.igb.uiuc.edu

Genus	Traits	Yields
Miscanthus	2741	506
Populus	1740	755
Panicum	606	1904
Salix	146	136
Andropogon	92	
Agave	88	
Betula	70	

PFT		
forb	287	
tree / shrub	194	3
sedge	50	32
C4 grass	43	
C3 grass	36	
nitrogen fixer	8	

Table 1: Data from targeted species-specific data collection for BETYdb. Data are summarized by genus for the top seven genera, and the rest of the data are summarized by plant functional type.

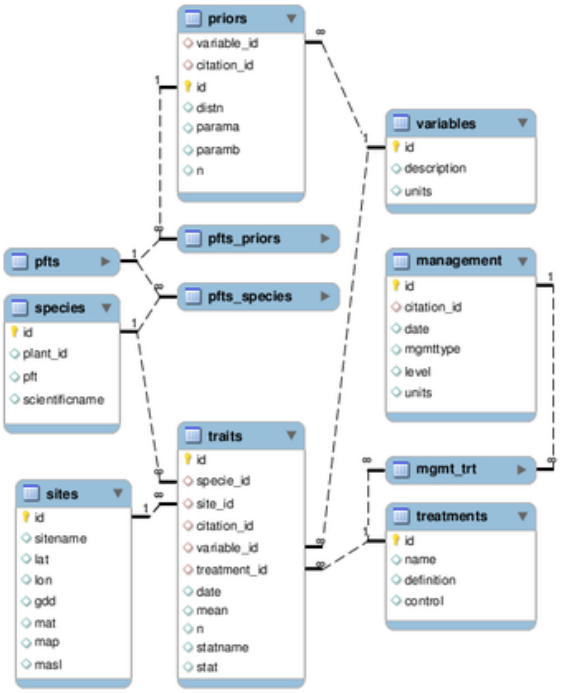
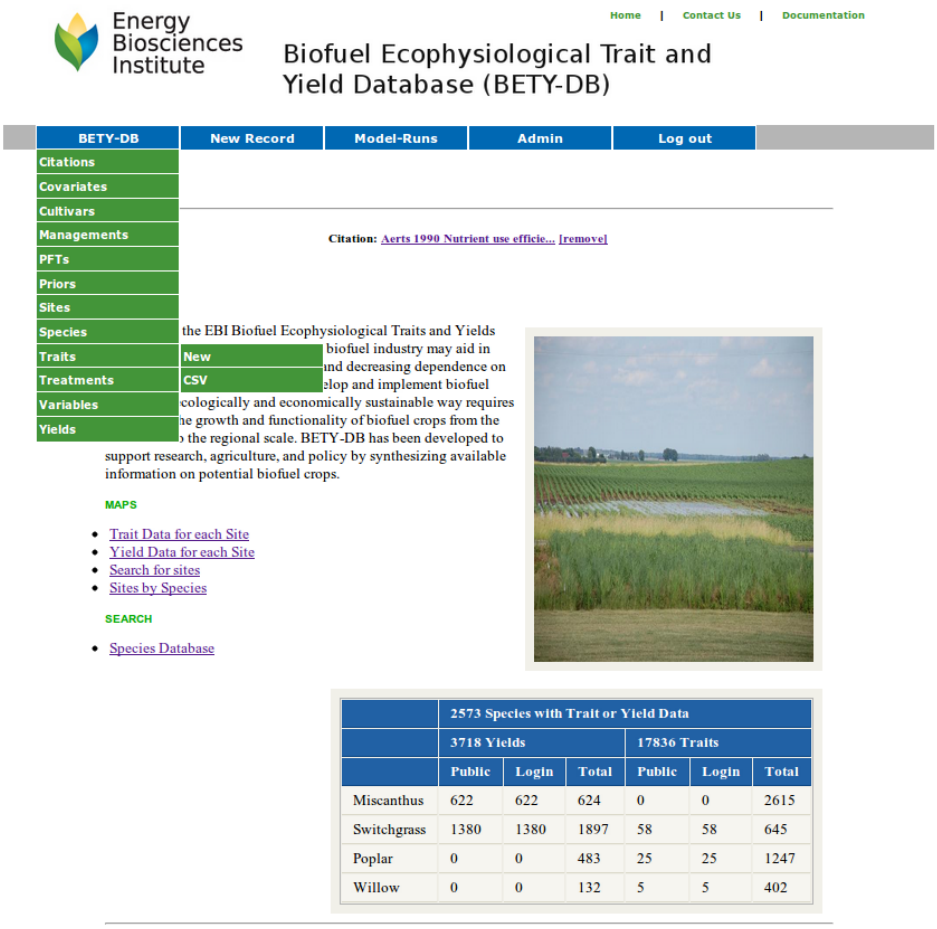


Figure 1: Abbreviated schema for BETYdb (Zoom).

2. Select `Species database` under **Search**
3. Select `Click Here` under Yields
4. to download all records as a comma-delimited (`.csv`) file, scroll down and select the link `http://ebi-forecast.igb.uiuc.edu/bety/maps/yields?format=csv&species=938CSV Format`

4.4 Web Interface

The web interface to BETYdb provides an interactive portal in which available data can be visualized, accessed, and entered (Figure 2).



	2573 Species with Trait or Yield Data					
	3718 Yields			17836 Traits		
	Public	Login	Total	Public	Login	Total
Miscanthus	622	622	624	0	0	2615
Switchgrass	1380	1380	1897	58	58	645
Poplar	0	0	483	25	25	1247
Willow	0	0	132	5	5	402

Figure 2: The BETYdb web interface home page.

4.5 Data Entry

The Data Entry Workflow provides a complete description of the data entry process. BETY's web interface has been developed to facilitate accurate and efficient data entry. This interface provides logical workflow to guide the user through comprehensively documenting data along with species, site information, and experimental methods. This workflow is outlined in the BETYdb Data Entry. Data entry requires a login with `Create` permissions, this can be obtained by contacting David LeBauer or Mike Dietze.

5 Tables

The database is designed as a relationship database management system (RDBMS), following the normalization Figure 1. Each table has a primary key field, `id`, which is a unique identifier for each record in the table. In addition, each record has `created_at` and `updated_at` fields. The traits and yields tables each has a `user_id` field to record the user who originally entered the data.

A complete list of tables is provided in Table 2, and a comprehensive description of the contents of each table is provided below.

Table 2: Comprehensive list, overview, and brief description of tables in BETY

Table	Name	Description
3	citations	Citation information, links
4	citations_sites	associates sites with citations
5	citations_treatments	associates citations with treatments
6	covariates	covariates are required for some traits
7	cultivars	cultivars associated with species
8	error_logs	
9	managements	quantifies managements, including treatment levels; provides dates associated with treatments
10	managements_treatments	associates managements with specific treatments
11	pfts	defines plant functional types (PFTs), users may choose existing pfts can be used, or user can enter pfts
12	pfts_priors	associates prior parameterizations with pfts used in modeling
13	pfts_species	associates species with pfts used in modeling
14	priors	PFT level summaries of available information for use in Bayesian meta-analysis
15	sites	Site level information
16	species	Based on USDA Plants database
17	traits	Trait data table
18	treatments	identifies experimental treatment name
19	variables	Description, including units, associated with variables used to define traits, trait covariates, and priors
20	yields	Yield data table

5.1 Table and field naming conventions

Each table is given a name that describes the information that it contains. For example, the table containing trait data is called `traits`, the table containing yield data is `yields`, and so on. Each table also has a *primary key*; the primary key is always `id`, and the primary key of a specific table might be identified as `yields.id` . One table can reference another table using a *foreign key*; the foreign key is given a name using the singular form of the foreign table, and underscore, and id, e.g. `traits_id` or `yields_id`.

In some cases, two tables can have multiple references to one another, known as a 'many to many' or 'm:n' relationship. For example, one citation may contain data from many sites; at the same time, data from a single site may be included in multiple citations. Such relationships use lookup tables. Lookup tables (e.g. Tables 4, 5, 4, 10, 12, 13) combine the names of the two tables being related, in the case of this example, the table used to link `citations` and `sites` is named `citations_sites`. These lookup tables have two foreign keys, e.g. `citation_id` and `site_id` but do not have a primary key The foreign keys are identified by `FK: table.column` in the comment fields of the database tables where `table` is either a) for 1:many relationships the name of the master table in which `column` is the primary key or b) for many to many (m:n) relationships, to the auxillary table with `column` adjacent to another column with which the m:n relationship is simplified into 1:m and 1:n relationships.

5.2 Data Tables

The two data tables, **traits** and **yields**, contain the primary data of interest; all of the other tables provide information associated with these data points. These two tables are structurally very similar as can be seen

in Tables 17 and 20.

traits

The **traits** table contains trait data (Table 17). Traits are measurable phenotypes that are influenced by a plants genotype and environment. Most trait records presently in BETY describe tissue chemistry, photosynthetic parameters, and carbon allocation by plants.

yields

The **yields** table includes aboveground biomass in units of Mg ha^{-1} (Table 20). Biomass harvested in the fall and winter generally represents what a farmer would harvest, whereas spring and summer harvests are generally from small samples used to monitor the progress of a crop over the course of the growing season. Managements associated with Yields can be used to determine the age of a crop, the fertilization history, harvest history, and other useful information.

5.3 Auxillary Tables

sites

Each site is described in the **sites** table (Table 15). A site can have multiple studies and multiple treatments. Sites are identified and should be used as the unit of spatial replication; treatments are used identify independent units within a site, and these can be compared to other studies at the same site with shared management. "Studies" are not identified explicitly but independent studies can be identified via shared management entries at the same site.

treatments

The **treatments** table provides a categorical identifier of a study's experimental treatments, if any (Table 18).

Any specific information such as rate of fertilizer application should be recorded in the managements table (section. A treatment name is used as a categorical (rather than continuous) variable, and the name relates directly to the nomenclature used in the original citation. The treatment name does not have to indicate the level of treatment used in a particular treatment - if required for analysis, this information is recorded as a management.

Each study includes a control treatment, when there is no experimental manipulation, the treatment is considered 'observational' and listed as control. In studies that compare plant traits or yields across different genotypes, site locations, or other factors that are built in to the database, each record is associated with a separate cultivar or site so these are not considered treatments.

For ambiguous cases, the control treatment is assigned to the treatment that best approximates the background condition of the system in its non-experimental state, for this reason, a treatment that approximates conventional agronomic practice may be labeled 'control'.

managements

The **managements** table provides information on management types, including planting time and methods, stand age, fertilization, irrigation, herbicides, pesticides, as well as harvest method, time and frequency.

The **managmenets** and **treatments** tables are linked through the `managements_treatments` lookup table (10).

Managements are distinct from treatments in that a management is used to describe the agronomic or experimental intervention that occurs at a specific time and may have a quantity whereas Treatment is a categorical identifier of an experimental group. Managements include actions that are done to a plant or ecosystem, for example the planting density or rate of fertilizer application.

In other words, managements are the way a treatment becomes quantified. Each treatment can be associated with multiple managements. The combination of managements associated with a particular treatment will distinguish it from other treatments. Each management may be associated with one or more treatments. For example, in a fertilization experiment, planting, irrigation, and herbicide managements would be applied to all plots but the fertilization will be specific to a treatment. For a multi-year experiment, there may be multiple entries for the same type of management, reflecting, for example, repeated applications of herbicide or fertilizer.

covariates

The **covariates** table is used to record one or more covariates associated with each trait record (Table 6). Covariates generally indicate the environmental or experimental conditions under which a measurement was made. The definition of specific covariates can be found in the **variables** table (Table 19). Covariates are required for many of the traits because without covariate information, the trait data will have limited value.

The most frequently used covariates are the temperature at which some respiration rate or photosynthetic parameter was measured. For example, photosynthesis measurements are often recorded along with irradiance, temperature, and relative humidity.

Other covariates include the size or age of the plant or plant part being measured. For example, root respiration is usually measured on fine roots, and if the authors define fine root as ¡2mm, the covariate root_minimum_diameter has a value of 2.

pfts

The plant functional type (PFT) table, **pfts** is used to group plants for statistical modeling and analysis. Each record in **pfts** contains a PFT that is linked to a subset of species in the **species** table. This relationship requires the lookup table **pfts_species** (Table 13). Furtheromre, each PFT can be associated with a set of trait prior probability distributions in the **priors** table (Table 14). This relationship requires the lookup table **pfts_priors** (Table 12).

In many cases, it is appropriate to use a pre-defined default PFT (e.g. `tempdecid` is temperate deciduous trees) In other cases, a user can define a new pft to query a specific set of priors or subset of species. For example, there is a PFT for each of the functional types found at the EBI Farm prairie. Such project-specific PFTs can be defined as `‘projectname‘.‘pft‘` (i.e. `ebifarm.c4grass` instead of `c4grass`).

variables

The **variables** table includes definitions of different variables used in the traits, covariates, and priors tables (Table 19). Each variable has a `name` field, and is associated with a standardized value for `units`. The `description` field provides additional information or context about the variable.

5.4 Lookup Tables

Lookup tables are required when each record on one table can be related to many records on another table, and vice-versa; this is called a 'many to many' relationship.

citations_sites

Because a single study can use multiple sites and multiple studies can use the same site, these relationships are tracked in the **citation_sites** table (Table 4).

citations_treatments

Because a single study can include multiple treatments and each treatment can be associated with multiple citations, these relationships are measured in **citations_treatments** table (Table 5).

managements_treatments

It is clear that one treatment can have many managements, e.g. tillage, planting, fertilization. It is also important to note that any managements applied to a control plot should, by definition, be associated with all of the treatments in an experiment; this is why the many-to-many lookup table **managements_treatments** is required.

pfts_priors

The **pfts_priors** table allows a many to many relationship between the **pfts** and **priors** tables (Table 12). This allows each pft to be associated with multiple priors and each prior to be associated with multiple pfts.

pfts_species

The **pfts_species** table allows a many to many relationship between the **pfts** and **species** tables (Table 13).

6 Acknowlegments

BETY-db is a product of the Energy Biosciences Institute at the University of Illinois at Urbana-Champaign. Funding for this research was provided by British Petroleum through a grant to the Energy Biosciences institute. We gratefully acknowledge the great effort of other researchers who generously made their own data available for further study.

7 Appendix

7.1 Full Schema: Enhanced Entity-Relationship Model

Figure 3 provides a visualization of the complete schema, including interrelationships among tables, of the biofuel database.

7.2 Software

The BETY-db has beeen developed in MySQL using Ruby on Rails and is hosted on a RedHat Linux Server (ebi-forecast.igb.uiuc.edu). BETY-db is a relational database designed in a generic way to facilitate easy implementation of additional traits and parameters.

Table 3: citations table

Field	Type	Null	Default	Comments
id	int(11)	No		
Field	Type	Null	Default	Comments
id	int(11)	No		
author	varchar(255)	Yes	NULL	last name of first author
year	int(11)	Yes	NULL	year of publication
title	varchar(255)	Yes	NULL	article title
journal	varchar(255)	Yes	NULL	Journal name
vol	int(11)	Yes	NULL	
pg	varchar(255)	Yes	NULL	page range of article
url	varchar(512)	Yes	NULL	link to article url
pdf	varchar(255)	Yes	NULL	link to pdf version of article
created_at	datetime	Yes	NULL	
updated_at	datetime	Yes	NULL	
doi	varchar(255)	Yes	NULL	Digital Object Identifier

Table 4: citations_sites table

Field	Type	Null	Default	Comments
citation_id	int(11)	Yes	NULL	
site_id	int(11)	Yes	NULL	
created_at	datetime	Yes	NULL	
updated_at	datetime	Yes	NULL	

Table 5: citations_treatments table

Field	Type	Null	Default	Comments
citation_id	int(11)	Yes	NULL	
treatment_id	int(11)	Yes	NULL	
created_at	datetime	Yes	NULL	
updated_at	datetime	Yes	NULL	

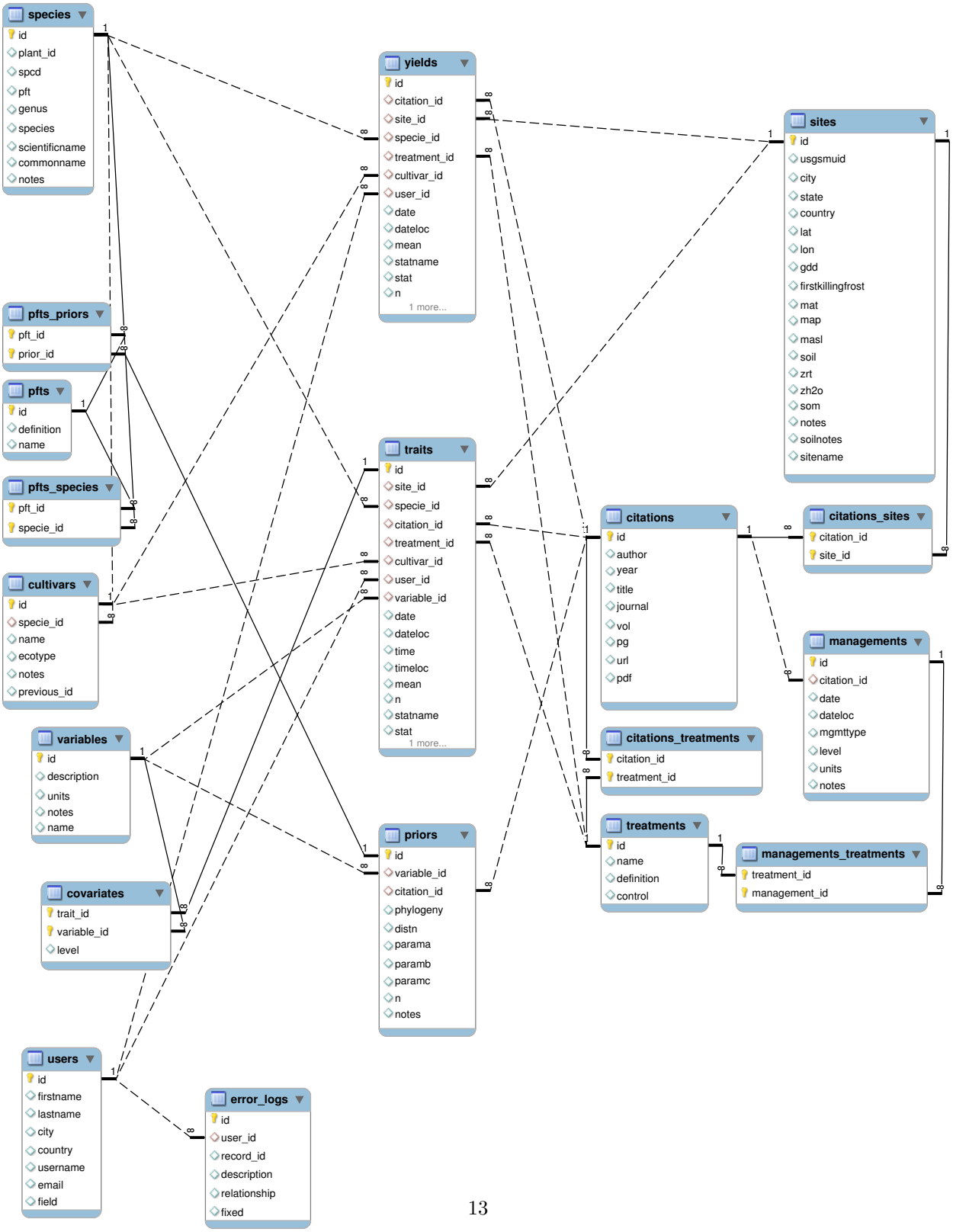


Figure 3: Full Schema of BETYdb, showing all tables and relations in the database

Table 6: covariates table

Field	Type	Null	Default	Comments
id	int(11)	No		
trait_id	int(11)	Yes	NULL	
variable_id	int(11)	Yes	NULL	
level	decimal(16,4)	Yes	NULL	Value of covariate, units are determined in variables table by the variable_id foreign key.
created_at	datetime	Yes	NULL	
updated_at	datetime	Yes	NULL	

Table 7: cultivars table

Field	Type	Null	Default	Comments
id	int(11)	No		
specie_id	int(11)	Yes	NULL	
name	varchar(255)	Yes	NULL	Cultivar name given by breeder or reported in citation.
ecotype	varchar(255)	Yes	NULL	Does not apply for all species, used in the case of switchgrass to differentiate lowland and upland genotypes.
notes	text	Yes	NULL	
created_at	datetime	Yes	NULL	
updated_at	datetime	Yes	NULL	
previous_id	varchar(255)	Yes	NULL	

Table 8: error_logs table

Field	Type	Null	Default	Comments
id	int(11)	No		
record_id	int(11)	Yes	NULL	
description	varchar(255)	Yes	NULL	Description of error that needs to be addressed.
relationship	varchar(255)	Yes	NULL	
user_id	int(11)	Yes	NULL	Identifies user responsible for handling error.
fixed	int(11)	Yes	0	Set to 0 when error is reported, 1 after error has been checked and fixed.
created_at	datetime	Yes	NULL	
updated_at	datetime	Yes	NULL	

Table 9: managements table

Field	Type	Null	Default	Comments
id	int(11)	No		
citation_id	int(11)	Yes	NULL	
date	date	Yes	NULL	Date on which management was conducted.
dateloc	decimal(4,2)	Yes	NULL	Level of confidence in value given as date. See documentation for details.
mgmttype	varchar(255)	Yes	NULL	Type of management
level	decimal(16,4)	Yes	NULL	Amount applied, not always required.
units	varchar(255)	Yes	NULL	units, standardized for each management type.
notes	text	Yes	NULL	
created_at	datetime	Yes	NULL	
updated_at	datetime	Yes	NULL	

Table 10: managements_treatments table

Field	Type	Null	Default	Comments
treatment_id	int(11)	Yes	NULL	
management_id	int(11)	Yes	NULL	
created_at	datetime	Yes	NULL	
updated_at	datetime	Yes	NULL	

Table 11: pfts table

Field	Type	Null	Default	Comments
id	int(11)	No		
definition	text	Yes	NULL	Defines the creator and context under which the pft will be used.
created_at	datetime	Yes	NULL	
updated_at	datetime	Yes	NULL	
name	varchar(255)	Yes	NULL	unique identifier used by PEcAn.

Table 12: pfts_priors table

Field	Type	Null	Default	Comments
pft_id	int(11)	Yes	NULL	
prior_id	int(11)	Yes	NULL	
created_at	datetime	Yes	NULL	
updated_at	datetime	Yes	NULL	

Table 13: pfts_species table

Field	Type	Null	Default	Comments
pft_id	int(11)	Yes	NULL	
specie_id	int(11)	Yes	NULL	
created_at	datetime	Yes	NULL	
updated_at	datetime	Yes	NULL	

Table 14: priors table

Field	Type	Null	Default	Comments
id	int(11)	No		
citation_id	int(11)	Yes	NULL	
variable_id	varchar(255)	Yes	NULL	Links to variable for which prior is used.
phylogeny	varchar(255)	Yes	NULL	Used to note the group of plants for which the prior was specified, often the group of plants represented by the data used to specify the prior.
distn	varchar(255)	Yes	NULL	Name of the probability distribution, using R naming convention (e.g. 'beta','f', 'gamma', 'lnorm', 'norm', 'pois', 't', 'unif', 'weibull'.
parama	decimal(16,4)	Yes	NULL	First parameter for distribution, as specified by R.
paramb	decimal(16,4)	Yes	NULL	Second parameter for distribution, as specified by R.
paramc	decimal(16,4)	Yes	NULL	A third parameter, if required.
n	int(11)	Yes	NULL	number of observations used to specify prior.
notes	text	Yes	NULL	
created_at	datetime	Yes	NULL	
updated_at	datetime	Yes	NULL	

Table 15: sites table

Field	Type	Null	Default	Comments
id	int(11)	No		
usgsmuid	varchar(255)	Yes	NULL	
city	varchar(255)	Yes	NULL	Nearest city to site.
state	varchar(255)	Yes	NULL	If in the United States, state in which study is conducted.
country	varchar(255)	Yes	NULL	
lat	decimal(9,6)	Yes	NULL	Latitude, in decimal degrees
lon	decimal(9,6)	Yes	NULL	Longitude, in decimal degrees.
gdd	int(11)	Yes	NULL	Depreciated
firstkillingfrost	date	Yes	NULL	Depreciated

Table 15: sites table (continued)

Field	Type	Null	Default	Comments
mat	int(11)	Yes	NULL	Mean Annual Temperature (C)
map	int(11)	Yes	NULL	Mean Annual Precipitation (mm)
masl	int(11)	Yes	NULL	Elevation (m above sea level)
soil	varchar(255)	Yes	NULL	Soil type, 'sand', 'loamy sand', 'sandy loam', 'silt loam', 'loam', 'sandy clay loam', 'silty clay loam', 'clay loam', 'sandy clay', 'silty clay', 'clay', 'peat'.
zrt	decimal(4,2)	Yes	NULL	Depreciated
zh2o	decimal(4,1)	Yes	NULL	Depreciated
som	decimal(4,2)	Yes	NULL	Depreciated
notes	text	Yes	NULL	
soilnotes	text	Yes	NULL	
created_at	datetime	Yes	NULL	
updated_at	datetime	Yes	NULL	
sitename	varchar(255)	Yes	NULL	
greenhouse	tinyint(1)	Yes	NULL	Boolean: indicates if study was conducted in a field (0) or greenhouse, pot, or growth chamber (1)

Table 16: species table

Field	Type	Null	Default	Comments
id	int(11)	No		
plant_id	int(11)	Yes	NULL	
spcd	int(11)	Yes	NULL	
pft	int(11)	Yes	NULL	Depreciated: moved to pfts_species table
genus	varchar(255)	Yes	NULL	
species	varchar(255)	Yes	NULL	
scientificname	varchar(255)	Yes	NULL	
commonname	varchar(255)	Yes	NULL	
notes	varchar(255)	Yes	NULL	
created_at	datetime	Yes	NULL	
updated_at	datetime	Yes	NULL	
...				other columns imported from USDA Plants database

Table 17: traits table

Field	Type	Null	Default	Comments
id	int(11)	No		
site_id	int(11)	Yes	NULL	Site at which measurement was taken.

Table 17: traits table (continued)

Field	Type	Null	Default	Comments
specie_id	int(11)	Yes	NULL	Species on which measurement was taken.
citation_id	int(11)	Yes	NULL	Citation in which data was originally reported.
cultivar_id	int(11)	Yes	NULL	Cultivar information, if any.
treatment_id	int(11)	Yes	NULL	Experimental treatment identification. Required, can indicate observational study.
date	datetime	Yes	NULL	Date on which measurement was made.
dateloc	decimal(4,2)	Yes	NULL	Level of confidence in date. See documentation.
time	time	Yes	NULL	Time at which measurement was taken. Sometimes necessary, e.g. for photosynthesis measurements.
timeloc	decimal(4,2)	Yes	NULL	Level of confidence in time.
mean	decimal(16,4)	Yes	NULL	Mean value of trait.
n	int(11)	Yes	NULL	Number of experimental replicates used to estimate mean and statistical summary.
statname	varchar(255)	Yes	NULL	Name of reported statistic.
stat	decimal(16,4)	Yes	NULL	Value of reported statistic.
notes	text	Yes	NULL	
created_at	datetime	Yes	NULL	
updated_at	datetime	Yes	NULL	
variable_id	int(11)	Yes	NULL	Links to information in variables table that describes trait being measured.
user_id	int(11)	Yes	NULL	ID of user who entered data.
checked	tinyint(1)	Yes	0	Boolean, indicates if data have been checked after original entry.
access_level	int(11)	Yes	NULL	Level of access required to view data.

Table 18: treatments table

Field	Type	Null	Default	Comments
id	int(11)	No		
name	varchar(255)	Yes	NULL	Name of treatment, should be easy to associate with treatment name in original study.
definition	varchar(255)	Yes	NULL	Description of treatment, e.g. levels of fertilizer applied, etc. This information may be redundant with 'levels' information recorded in Managements table.

Table 18: treatments table (continued)

Field	Type	Null	Default	Comments
created_at	datetime	Yes	NULL	
updated_at	datetime	Yes	NULL	
control	tinyint(1)	Yes	NULL	Boolean, indicates if treatment is a control or observational (1) or experimental treatment (0).

Table 19: variables table

Field	Type	Null	Default	Comments
id	int(11)	No		
description	varchar(255)	Yes	NULL	Description or definition of variable.
units	varchar(255)	Yes	NULL	units in which data must be entered.
notes	text	Yes	NULL	
created_at	datetime	Yes	NULL	
updated_at	datetime	Yes	NULL	
name	varchar(255)	Yes	NULL	variable name, this is the name used by PEcAn and in other modeling contexts.

Table 20: yields table

Field	Type	Null	Default	Comments
id	int(11)	No		
citation_id	int(11)	Yes	NULL	Citation in which data originally reported.
site_id	int(11)	Yes	NULL	Site at which crop was harvested.
specie_id	int(11)	Yes	NULL	Species for which yield was measured.
treatment_id	int(11)	Yes	NULL	Experimental treatment identification. Required, can indicate observational study.
cultivar_id	int(11)	Yes	NULL	Cultivar information, if any.
date	date	Yes	NULL	Date on which crop was harvested.
dateloc	decimal(4,2)	Yes	NULL	Level of confidence in harvest date. See documentation.
statname	varchar(255)	Yes	NULL	Name of reported statistic.
stat	decimal(16,4)	Yes	NULL	Value of reported statistic.
mean	decimal(16,4)	Yes	NULL	Mean yield reported.
n	int(11)	Yes	NULL	Number of replicates used to estimate mean and statistical summary.
notes	text	Yes	NULL	

Table 20: yields table (continued)

Field	Type	Null	Default	Comments
created_at	datetime	Yes	NULL	
updated_at	datetime	Yes	NULL	
user_id	int(11)	Yes	NULL	ID of user who entered data.
checked	tinyint(1)	Yes	0	Boolean, indicates if data have been checked after original entry.
access_level	int(11)	Yes	NULL	Level of access required to view data.