# Digitization Protocol

*Sam Levin*

*2018-03-09*

## Contents

# PADRINO
Plant IPM Database

# MADRINA
Animal IPM Database

## Padrino digitization protocol

Welcome to the `Padrino` and `Madrina` digitization team! The following is an in-depth guide for how to translate a published IPM into the *Excel* sheets and what to do with that sheet once you have completed the process.

There is a good chance I will have forgotten to include some important aspects of what this work entails. If that happens, please create an *Issue* in the GitHub repository as opposed to emailing me. This helps me (and, hopefully, you) keep all discussion related to the problem in a centralized location. Most importantly, this centralized location is accesible and searchable by your fellow digitizers, so they can also participate in the discussion or return to it later to reference it as needed.

This document assumes you have a general understanding of an IPM and why we want to have all of them in one place.

## Data base structure

`Padrino` and `Madrina` are remotely hosted relational data bases that consider an individual IPM as the "atomic unit", as opposed to each vital rate or parameter. This helps keep relations tidy and avoid unnecessary data duplication, but can introduce some confusion at first. Hopefully, after reading this, the reasons for this decision will become more clear.

Relational data bases can be a bit tricky to work with if you aren't already familiar with them. My (admittedly limited) experience has been that this is not the most accesible format for most ecologists who are usually more accustomed to working in *R* or *Excel*. Therefore, we have created a "flat" version of it that is a set of 5 *Excel* sheets. The columns of each one are described below. The format of each description is: **data type**, description of variable, *constraint*.

### Metadata

This is the table you will find when opening the spreadsheet. It contains important information about the study site, species, and authors that aren't necessarily relevant to construction of the IPM, but provide important context nonetheless.

### ipm_id

**Character**

This column contains a unique identifier for each IPM in the data base. There should only be **1** row per IPM in this table.

*This column cannot be blank.*

### species_author

**Character**

The genus and species epithet of the organism used by the author in the publication.

*This column cannot be blank.*

**species_accepted**

**Character**

The currently accepted genus and species epithet of the organism. For `Padrino` entries, this will come from the The Plant List. For `Madrina` entries, this will come from the Catalogue of Life.

*This column cannot be blank.*

**tax_genus**

**Character**

The accepted genus name of the organism. For `Padrino` entries, this will come from the The Plant List. For `Madrina` entries, this will come from the Catalogue of Life.

*This column cannot be blank.*

**tax_family**

**Character**

The accepted family of the organism. For `Padrino` entries, this will come from the The Plant List. For `Madrina` entries, this will come from the Catalogue of Life.

*This column may be blank.*

**tax_order**

**Character**

The accepted order of the organism. For `Padrino` entries, this will come from the The Plant List. For `Madrina` entries, this will come from the Catalogue of Life.

*This column may be blank.*

**tax_class**

**Character**

The accepted class of the organism. For `Padrino` entries, this will come from the The Plant List. For `Madrina` entries, this will come from the Catalogue of Life.

*This column may be blank.*

**tax_phylum**

**Character**

The accepted phylum of the organism. For `Padrino` entries, this will come from the The Plant List. For `Madrina` entries, this will come from the Catalogue of Life.

*This column may be blank.*

**kingdom**

**Character**

The accepted kingdom of the organism. For `Padrino` entries, this will come from the The Plant List. For `Madrina` entries, this will come from the Catalogue of Life.

*This column may be blank.*

**organism_type**

**Character**

`Padrino entries`

This is the general plant/algae type. This will usually come from the publication itself, but sometimes you may need to use other sources (e.g. other publications or taxonomic data bases) to find this information. Possible values are as follows

- Algae: brown, green or red. Green algae are in the *Plantae* kingdom, but are still considered algae for the purposes of this variable.

- Fungi: This includes fungus species, yeasts, molds, and multicellular fungi.

- Annual: This includes annuals and biennials. Annuals complete their entire lifecycle (birth, growth, reproduction, death) within a year wherease biennials can stretch that window to two years. For the sake of simplicity (and because both will die following reproduction), they are both classified as "Annual" in `Padrino`.

- Bryophyte: All bryophytes.

- Epiphyte: All epiphytes.

- Fern: All ferns species.

- Herbaceous perennial: All plants that herbaceous and have the potential to live for more than two years.

- Liana: All lianas.

- Palm: All palm species.

- Shrub: Woody upright plants that are not trees or palms.

- Succulent: All succulent species.

- Tree: All tree species.

`Madrina entries`

This is generally the same as `Class` for animals (except humans, which are recorded using their genus and species epithet). Non-animal species that are also not plants are typically recorded as *Bacteria* or *Virus*.

*This column may be blank.*

**dicot_monocot**

**Character**

Indicates whether a species is a dicot or monocot. Not applicable for `Madrina` entries.

*This column may be blank.*

**angio_gymno**

**Character**

Indicates whether a species is a angiosperm or gymnosperm. Not applicable for `Madrina` entries.

*This column may be blank.*

**authors**

**Character**

The last name of all authors. Multiple entries should be separated with semicolon (";").

*This column cannot be blank.*

**journal**

**Character**

The document that the information comes from. Possible values are listed below.

- Abbreviated name of the journal: We use the BIOSIS system for abbreviating journal names. More information on how to use it is in the link.
- Book: Models are sourced from a book.
- PhD Thesis: Models are sourced from a PhD thesis.
- MSc Thesis: Models are sourced from an MSc thesis.
- Report: Models are sourced from a report.
- Conference talk: Models are sourced from a conference talk.
- Conference poster: Models are sourced from a conference poster.

**pub_year**

**Integer**

The year that the model was published.

*This column cannot be blank.*

**doi**

**Character**

The *DOI* or *ISBN* for the publication.

*This column may be blank.*

**corresponding_author**

**Character**

The author to whom correspondance should be directed.

*This column may be blank.*

**email_year**

**Character**

The email address of the corresponding author with the year it is from in parentheses. If the email address is no longer in use, add the word "Dead" after a comma in the parentheses.

Example: levisc8@gmail.com (2018); levisc8@wfu.edu (2010, Dead)

*This column may be blank.*

**remark**

**Character**

Any observations you have about the model that are not captured by the other columns in `Metadata`.

*This column may be blank.*

**apa_citation**

**Character**

The full APA citation for the source.

*This column may be blank.*

**demog_appendix_link**

**Character**

If the model parameters are contained in an appendix, then include the link to said appendix here.

*This column may be blank.*

**duration**

**Integer**

Model duration is defined as the end year −
This table contains information on the state variables used by the authors to generate their IPM.

**ipm_id**

**state_variable**

**discrete**

**discrete_type**

**Domains**

This table contains information on the domains associated with each state variable in `States`. Keep in mind that one `state_variable` can have multiple domains (which themselves may be defined by a different `state_variable`!).

**ipm_id**

**state_variable**

**domain**

**lower**

**upper**

**n_meshpoints**

### Model Expressions (`ModelExpr`)

This table contains textual expressions of the models used to create IPM. A given IPM will have many rows in this table.

**ipm_id**

**demographic_parameter**

**formula**
See Writing Model Formulae for additional details

**model_type**

**model_family**

### Model Values (`ModelValues`)

This contains the actual values for all of the parameters described in `ModelExpr`. A single IPM will have many rows in this table.

Every column in this table **must** be filled in to be able to enter the data base.

**ipm_id**

**demographic_parameter**

**state_variable**

**parameter_type**

**parameter_name**

**parameter_value**

## Writing Model Formulae

Details on conventions and lots of examples!