# SEV metadata template

Comprehensive metadata are necessary to support data quality control, efficient dataset archiving and retrieval, and functional re-use of the data both by owners and secondary users. The Sevilleta LTER requires complete documentation for long-term maintenance and distribution of study datasets. Metadata elements requested in this form are consistent with elements required by the Ecological Metadata Standard (EML), which was adopted by the LTER Coordinating Committee in 2004.

Sevilleta metadata consists of information about:

1. Why the study was done
2. Who is involved with study design, data collection, analysis and data management
3. Detailed study methods, so that a secondary user will be able to understand what was done without contacting the study principal investigator.
4. Where the study was done, with GPS points so the site or plots can be relocated.
5. Detailed variable information

Tips for completing a data package (data + metadata):

* Data file and variable names should be brief, yet descriptive.
* Data file names should not contain spaces or symbols (except '\_', '-', and '.'). File names should be descriptive of the data contained in the file.
* Data files should generally be in a flat table format, similar to a SQL database table. Rows are observations, and columns are variables.
* Data files should be in a csv or other delimited text file format. Excel (e.g., .xlsx, .xls, etc.) files are not acceptable.
* Variable names should be descriptive, short, and machine readable. They should not begin with a number or symbol. They should only contain letters, numbers, and underscores ('\_'). They should not contain spaces. (e.g., 'air\_temp' or 'airTemp')
* Variable names used in the data file(s) should match the names used in this metadata document.
* Use ISO 8601 date formats. (e.g., YYYY-MM-DD)

Data should be in csv text file. If starting with an Excel spreadsheet, make sure it does not contain any formulas and comments on cells. If you need comments, put them in their own columns, such as 'field\_notes' and 'qaqc\_notes'. If data were used in a database and major table linking is necessary to analyze the data, de-normalize data into a flat files - do not just export database tables.

Please fill in this metadata form to the best of your ability and submit it to the Sevilleta Information Manager, Kristofer Hall, [khall001@unm.edu](mailto:khall001@unm.edu). If you have questions at any stage, contact the SEV IM.Allow several weeks for review and revision of your data package.

## Dataset Title

Demographic data from long-term symbiont removal experiments with grasses and *Epichloë* fungal endophytes

## Short name or nickname you use to refer to this dataset:

Indiana endo-demog

## Abstract

This project was designed to understand the demographic effects of vertically transmitted fungal endophytes (*Epichloë* spp.) on their grass hosts. The experiment includes seven host-symbiont taxonomic pairs: *Agrostis perennans - E. amarillans, Elymus villosus - E. elymi, Elymus virginicus - E. elymi or EviTG-1,* *Festuca subverticillata - E. starrii, Poa alsodes - E. alsodes, Poa sylvestris - E. PsyTG-1,* *Schedonorus arundinaceus - E. coenophiala*. Experimental plots were established at the Indiana University Lilly-Dickey Woods Research and Teaching Preserve in south-central Indiana, USA in 2007. For each species, 5-10 plots were planted with naturally symbiotic (S+) hosts, and 5-10 plots were plated with hosts that were disinfected of fungal endophytes by heat treatment (S-). Over 14 years (2007-2021) we collected demographic data on the survival, growth, reproduction, and recruitment of all plants in all plots. Beginning in 2018 we also collected data on the locations of all plants in every plot.

## Investigators

List in order as you would for a paper. Include an e-mail address, organization and preferably an ORCID ID. If you don't have an ORCID ID, get one, it's easy and free! Go to <http://orcid.org/> .

Required Roles include the 'creator' who is the author of the dataset, 'PI' who is the principal investigator, and 'contact' who is the dataset contact. If individuals have multiple roles, list them on a separate line for each role. Other roles are acceptable, but a creator, PI and contact are required.

Add table rows as needed.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| First Name | Middle Initial | Last Name | Organization | Role | e-mail address | ORCID ID (optional) |
| Joshua | C | Fowler | University of Miami | creator, contact | joshuacfowler1@gmail.com | 0000-0001-8956-0896 |
| Shaun | M | Ziegler | University of New Mexico | creator | shaun.ziegler@gmail.com |  |
| Tom | EX | Miller | Rice Universiy | creator, PI | tom.miller@rice.edu | 0000-0003-3208-6067 |
| Jennifer | A | Rudgers | University of New Mexico | creator, PI, contact | jrudgers@unm.edu | 0000-0001-7094-4857 |
| Kenneth | D | Whitney | University of New Mexico | creator, PI | whitneyk@unm.edu | 0000-0002-2523-5469 |

## Other personnel names and roles

Field crew, data entry, etc. Include an e-mail address, organization and preferably an ORCID ID.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| First Name | Middle Initial | Last Name | Organization | Role | e-mail address | ORCID ID (optional) |
| Mark |  | Sheehan | Rice University | creator | markcsheehan@msn.com |  |
|  |  |  |  |  |  |  |

## License

Select a license for release of your data. We have 2 recommendations: [CC0 – most accommodating of data reuse](https://creativecommons.org/publicdomain/zero/1.0/), and [CCBY – requires attribution](https://creativecommons.org/licenses/by/4.0/)

**CCO**

## Keywords

List keywords and separate with commas. Using keywords from a controlled vocabulary (CV) will improve the future discovery and reuse of your data. The LTER CV is effective at describing ecological and environmental data. [Access the LTER CV here](http://vocab.lternet.edu/vocab/vocab/index.php). [Try this text mining service to extract LTER CV keywords from your abstract or methods](http://vocab.lternet.edu/keywordDistiller/). Additionally, determine one or two keywords that best describe your lab, station, and/or project (e.g., Sevilleta Field Station, SEV LTER). This will help others discover your data by site/project.

Demography; symbiosis; fungal endophyte; Lilly-Dickey Woods; population dynamics

## Funding of this work:

Add rows to table if several grants were involved, list only the main PI, start with main grant first:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| PI First Name | PI Middle Initial | PI Last Name | PI ORCID ID (optional) | Title of Grant | Funding Agency | Funding Identification Number |
| Tom | EX | Miller | 0000-0003-3208-6067 | LTREB: Host-symbiont interactions through the lens of stochastic demography | National Science Foundation | 1754468 |

## Timeframe

* Begin date: May 2007
* End date: September 2022
* Data collection ongoing

## Geographic location

Tips: SEV core sites are: Deep Well, Five Points Grass, Five Points Creosote, Blue Grama, Sepultura Canyon, Cerro Montoso, Goat Draw, Black Butte, Sevilleta Field Station, Bronco Well, Rio Salado Grass, Rio Salado Creosote, Red Tank. If you do not know the coordinates of your site, then enter the coordinates for the bounding box of the Sevilleta: North: 34.42, South = 34.19, East = -106.513, West = -107.08. If you are working at multiple sites or plots within the SEV, provide GPS coordinates for each site and/or plot. High accuracy GPS recordings can be made by contacting Lauren Baur ([lbaur@unm.edu](mailto:lbaur@unm.edu)). COORDINATES MUST BE ENTERED in DECIMAL DEGREES AND NAD83. There is a degrees/minutes/seconds to decimal converter at <http://www.fcc.gov/mb/audio/bickel/DDDMMSS-decimal.html>.

Include:

* Verbal description (elevation, landform, geology, soils, hydrology, vegetation, climate, site history, etc.): Lilly-Dickey Woods is part of the Eastern broadleaf forests of southern Indiana. Plots are situated among rolling hills in the forest understory. Elevation spans 289 to 343 m.
* North bounding coordinates (decimals): 39.245723, -86.214618
* South bounding coordinates (decimals): 39.241273, -86.214610
* East bounding coordinates (decimals): 39.243131, -86.210058
* West bounding coordinates (decimals): 39.245655, -86.220144

## Taxonomic species or groups

Tips: For plants, include the USDA plant codes and most up to date plant taxonomy from USDA PLANTS website <https://plants.sc.egov.usda.gov/java/> so that data are concordant with SEV LTER plant codes. For animals, check the SEV LTER website to match species names and species codes with any overlapping species monitored by the LTER.

*Agrostis perennans –* AGPE

*Elymus villosus –* ELVI

*Elymus virginicus -* ELVI3

*Festuca subverticillata -* FESU3

*Poa alsodes* – POAL3

*Poa sylvestris -* POSY

*Schedonorus arundinaceus -* SCAR7

## Methods

Be specific. Include instrument descriptions (name, manufacturer, model number), or point to a protocol online. If this is a data compilation, specify datasets used, preferably their DOI or URL plus general citation information. If the data package is associated with a manuscript, you may paste the methods from the manuscript.

Seeds from naturally symbiotic populations of the seven focal host species were collected during summer-fall 2006 from Lilly-Dickey Woods and the nearby Bayles Road Teaching and Research Preserve (39.220167, -86.542683). To generate symbiotic (S+) and symbiont-free (S-) plants from the same genetic lineages, seeds from each species were disinfected with a heat treatment described in Table S1 or left untreated. The heat treatment created symbiont-free plants by warming seeds to temperatures at which the fungus becomes inviable but the host seeds can still germinate.Both heat-treated and untreated seeds were surface sterilized with bleach to remove epiphyllous microbes, cold stratified for up to 4 weeks, then germinated in a growth chamber before transfer to the greenhouse at Indiana University where they grew for ~8 weeks. We confirmed endophyte status by staining thin sections of inner leaf sheath with aniline blue and examining tissue for fungal hyphae at 200X magnification. We established experimental populations with vegetatively propagated clones of similar sizes. By starting the experiment with plants of similar sizes and the same number of unique genotypes, we aimed to limit any potential effects of heat treatments on initial plant growth.

During the fall of 2007 and spring of 2008, we established 10 3x3 m plots for *A. perennans*, *E. villosus*, *E. virginicus*, *F. subverticillata*, and *S*. *arundinaceus* and 18 plots for *P. alsodes* and *P. sylvestris*. Half of the plots were randomly assigned to be planted with either symbiotic or with symbiont-free plants, and initiated with 20 evenly spaced individuals labeled with aluminum tags. In spring 2008, we placed plastic deer net fencing around each plot to limit deer herbivory and disturbance; damaged fences were regularly replaced.

Each summer starting in 2008 through 2021, we censused all individuals in each plot for survival, growth and reproduction, adding new recruits to the census. We censused each species during its peak fruiting stage (May: *P. alsodes*, *P. sylvestris*; June: *F. subverticillata*; July: *E. villosus*, *E. virginicus*, *S*. *arundinaceus*; September: *A. perennans*), such that the censuses were pre-breeding and new recruits came from the previous years' seed production. Leaf litter was cleared out of each plot prior to the census, to aid in locating plants. For each plant remaining from the previous year, we determined survival, measured its size as a count of tillers, and collected reproductive data as counts of reproductive tillers and seed-bearing spikelets on all reproductive tillers to a maximum of three. We also tagged all unmarked individuals that were recruits from the previous years' seed production and collected the same demographic data. New recruits typically had one tiller and were non-reproductive. In 2008 and 2009, we took additional counts of seeds per inflorescence for all reproducing individuals in the plots to ground-truth our sub-sample estimates. In 2018, we stopped collecting data for the *S*. *arundinaceus* plots, which had very high survival and low recruitment, and consequently very low variation across years. Beginning in 2018, we used a laser measuring device to measure the distances of every plant to two plot corner posts, labeled “A” and “B”. With these pairs of distances, x/y coordinates can be derived using Pythagorean algebra.

In 2021 and 2022, because plot numbers had strongly declined, we planted new greenhouse-reared from the same source population and endophyte status as the 2007 founders. We planted 6-10 plants in all plots of *P. alsodes*, *P. sylvestris*, *F. subverticillata*, *E. villosus,* and *A. perennans*.

## Data Table

* Column name: exactly as it appears in the dataset. Avoid special characters, dashes and spaces. See tips at the beginning of this template.
* Description of the variable: be specific, it can be lengthy
* Unit: avoid special characters and describe units in this pattern: e.g., microSiemenPerCentimeter, microgramsPerLiter, absoptionPerMolePerCentimeter
* Code explanation: if you use codes in your column, explain them in this way: e.g., LR=Little Rock Lake, A=Sample suspect, J=Nonstandard routine followed
* Data format: specify exactly how the date and time is formatted. Use ISO 8601 date formats. If reporting datetimes, specify the time zone and whether or not daylight savings was observed if that is important to data interpretation. (e.g., YYYY-MM-DD (2019-02-22), YYYY-MM-DDThh:mm:ssZ (2019-02-22 14:33:23)). See the [Wikipedia 'ISO 8601'](https://en.wikipedia.org/wiki/ISO_8601) page for more info. If you are trying to convert Excel date formats to ISO 8601, see the most upvoted answer by Dirk Bester on [this Stack Overflow page](https://stackoverflow.com/questions/27388761/how-to-convert-a-date-in-excel-to-iso-8601-format).
* If a code for no data or missing data is used, specify what the code is. (e.g., -99999 or NA)

**Table file name:** LDW\_LTREB\_20072022.csv

**Table description:** Long-term demographic data from symbiont-removal experiment. Each row represents one transition year of one individual (from the census in year t to the census in year t+1). Individuals that survive one or more years occur in multiple rows.

Add rows as needed. Use a separate table for each data file in your data package.

|  |  |  |  |
| --- | --- | --- | --- |
| Column name | Description | Unit or  code explanation or date format | Empty value code |
| species | Host grass species | AGPE = *A. perennans*  ELRI = *E. villosus*  ELVI = *E. virginicus*  FESU = *F.* subverticillata  LOAR = *S*. *arundinaceus* (formerly *Lolium arundinaceum*)  POAL = *P. alsodes*  POSY = *P. sylvestris* | NA |
| plot | Plot identified | Arbitrary integer | NA |
| endo\_01 | Endophyte status assigned to plot | 0 = endophyte-free (S-), 1 = endophyte-symbiotic (S+) | NA |
| id | Unique ID of focal plant. For original founders, a unique four-digit tag number. For recruit plants, a concatenation of plot number and recruit number (e.g., the first recruit observed in plot 111 has id = 111\_1) | Arbitrary integer | NA |
| origin\_01 | Whether or not the individual was greenhouse-reared and planted by the investigators or recruited into the population | 0 = greenhouse-reared “original” plants (2007 founders plus 2021-22 additions), 1 = natural recruit | NA |
| dist\_a | Distance to post A | meters | NA |
| dist\_b | Distance to post B | meters | NA |
| birth | Year in which the plant recruited into the population; NA when birth year is unknown | Year (YYYY) | NA |
| endo\_status\_from\_check | Individual-level endophyte status observed during spot-checks | 0 = endophyte-free (S-), 1 = endophyte-symbiotic (S+) | NA |
| year\_t | Calendar year at the start of the census interval | Year (YYYY) | NA |
| size\_t | Number of tillers at the start of the census interval | Count | NA |
| flw\_count\_t | Number of flowering tillers at the start of the census interval (always less than or equal to number of tillers) | Count | NA |
| spike\_a\_t | Count of spikelets on one of three arbitrarily chosen flowering tillers at the start of the census interval | Count | NA |
| spike\_b\_t | Count of spikelets on one of three arbitrarily chosen flowering tillers at the start of the census interval | Count | NA |
| spike\_c\_t | Count of spikelets on one of three arbitrarily chosen flowering tillers at the start of the census interval | Count | NA |
| spike\_d\_t | Count of spikelets on a fourth flowering tiller at the start of the census interval (recorded infrequently) | Count | NA |
| year\_t1 | Calendar year at the end of the census interval | Year (YYYY) | NA |
| surv\_t1 | Survival status at the end of the census interval | 0 = dead, 1 = alive | NA |
| size\_t1 | Number of tillers at the end of the census interval | Count | NA |
| flw\_count\_t1 | Number of flowering tillers at the end of the census interval (always less than or equal to number of tillers) | Count | NA |
| spike\_a\_t1 | Count of spikelets on one of three arbitrarily chosen flowering tillers at the end of the census interval | Count | NA |
| spike\_b\_t1 | Count of spikelets on one of three arbitrarily chosen flowering tillers at the end of the census interval | Count | NA |
| spike\_c\_t1 | Count of spikelets on one of three arbitrarily chosen flowering tillers at the end of the census interval | Count | NA |
| spike\_d\_t1 | Count of spikelets on a fourth flowering tiller at the end of the census interval (recorded infrequently) | Count |  |

## Articles

List articles that cite this dataset (optional)

|  |  |  |
| --- | --- | --- |
| Article DOI or URL (DOI is preferred) | Article title | Journal title |
| https://github.com/joshuacfowler/Grass-Endophyte-Stochastic-Demography/blob/master/Manuscript/submission\_to\_PNAS/Endo\_stoch\_demo.pdf | Microbial symbionts buffer host-plants from the demographic costs of environmental stochasticity | In review |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

## Scripts/code (software)

List any software scripts/code you would like to archive along with your data. These may include processing scripts you wrote to create, clean, or analyze the data. Submit copies of the scripts as part of the data package.

|  |  |  |
| --- | --- | --- |
| File name | Description | Scripting language |
| endo\_demog\_data\_processing.R | This script cleans and harmonizes several raw data sources to produce the derived data product archived here. | R |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

## Data provenance

Were these data derived from other data? If so, document this information so data users will know where the data came from.

|  |  |  |  |
| --- | --- | --- | --- |
| Dataset title | Dataset DOI or URL | Creator (name & email) | Contact (name & email) |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

## QA/QC Procedures

Data distributions and variable classes were inspected in R. Any data issues were corrected in our data processing script, providing a record of how derived data relate to raw data inputs.

## Notes and Comments

If you are a graduate student, provide an estimated/projected date of public release for your data.