# 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

Be descriptive, more than 5 words:

Creosotebush demography along transects spanning grass-shrub ecotones at Sevilleta National Wildlife Refuge

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

Creosote demography

## Abstract

Provide an abstract describing the purpose of the research project *-* include what, why, where, when, and how

Encroachment of shrubs into adjacent grasslands has become an increasingly reported phenomenon across the world. In many cases, such encroachment is either pulled forward by high population growth at the low-density encroachment front or pushed forward by higher-density areas behind the front. However, at sites such as Sevilleta National Wildlife Refuge in central New Mexico, little is known about whether encroachment is pushed or pulled, and the dynamics of encroachment are not well-understood. Here, long-term encroachment of creosotebush (*Larrea tridentata*), a native perennial shrub, stands in stark contrast with the stagnation in encroachment observed in recent decades.

In order to better understand creosotebush encroachment dynamics at the Sevilleta National Wildlife Refuge, we quantify this encroachment using a spatially structured population model where a wave of individuals travels at a speed governed by both dispersal and density-dependence. Density dependence was studied along a gradient of creosotebush densities, using observational and experimental approaches such as transplant experiments and annual surveys; dispersal was separately modelled using seed terminal velocity estimates in a wind-based dispersal framework. Demographic and dispersal data were then incorporated into the spatial population model, which estimates the speed of encroachment and determines whether it is pushed or pulled.

## 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) |
| Trevor | H. | Drees | Pennsylvania State University | Creator | thd5066@psu.edu |  |
| Brad | M. | Ochocki | Rice University | Creator | brad.ochocki@gmail.com |  |
| Tom | E. | Miller | Rice University | Creator, contact,  PI | tom.miller@rice.edu |  |

## 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) | Role in project |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

## License

Select a license for release of your data. We have 2 recommendations: [CCO – 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, ecology, shrub, ecotone

## 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 |
| Miller | EX | Tom |  | N/A | Rice University start-up funds | N/A |

## Timeframe

* Begin date
* End date
* Data collection ongoing/completed

## 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.):
* North bounding coordinates (decimals)
* South bounding coordinates (decimals)
* East bounding coordinates (decimals)
* West bounding coordinates (decimals)

This study was conducted at four sites at the Sevilleta National Wildlife Refuge. We named these sites Five Points (FPS), Palo Duro Canyon (PDC), Middle of Desert (MOD), and Southern Los Pinos (SLP). At each site there were three transects (e.g., FPS-1, FPS-2, FPS-3) spanning shrub-grass ecotones that were variable in length (200-600m). These are the coordinates of the starting positions (meter zero) of each transect:

PDC-1: -106.725300 34.330244

PDC-2: -106.725487 34.329019

PDC-3: -106.725588 34.330793

FPS-1: -106.701584 34.320695

FPS-2: -106.700229 34.318763

FPS-3: -106.702532 34.322238

MOD-1: -106.690047 34.290288

MOD-2: -106.689847 34.289510

MOD-3: -106.691007 34.294074

SLP-1: -106.670569 34.273684

SLP-2: -106.669399 34.273660

SLP-3: -106.671538 34.273571

## 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.

LATR2 – *Larrea tridentata*

## 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.

Collection of creosotebush demographic data occurred during early June of every year from 2013-2017, at the Sevilleta National Wildlife Refuge LTER site in central New Mexico. Four diﬀerent sampling sites in the eastern part of the reserve were designated, with each of the sites containing 3 diﬀerent transects. Lengths of these transects varied from 200 to 600 m, and no two sites had identical compositions of transect lengths. Transect length was determined by the strength of vegetation transition, as areas where shrubland more quickly transitions to grassland do not need as long of a transect to capture the gradient of densities as a more gradual transition does. All transects were placed longitudinally along the shrubland-grassland ecotone so a full range of shrub densities could be captured; each transect spanned shrub-dense ”core” areas as well as grasslands with few shrubs and the transition zones in between.

At 50-meter intervals, a subset of the shrubs within 1 m on each side of the transect were tagged, with each plant given a unique identiﬁer that allowed it to be recognised based on sampling site, transect number, and location within 50-m and 5-m subsections. These tagged shrubs then had various demographic measurements recorded on an annual basis. Height, maximum width, and width perpendicular to the maximum on each shrub were measured in order to estimate size. Survival status of the shrubs was also recorded, with dead individuals being noted and excluded from measurements in subsequent years. Counts of ﬂowers and fruits on each shrub were recorded as well. In instances where shrubs had large numbers of reproductive structures that would prove diﬃcult to reliably count, estimates were made, with a more accurate count on a fraction of the shrub being extrapolated to the entire individual. The position of each shrub along the transect was noted to a resolution of 5 m so that it could be matched with the baseline density of its corresponding subsection. For shrubs in which a given 5-m subsection was not recorded, their position was estimated to the nearest 50 m; however, compared to the number of ﬁner-resolution 5-m subsections, this occurred relatively infrequently. Establishment of recruits was also accounted for. Each year we searched for new recruits within 1-m on either side of the entire transect length. New recruits were tagged and added to the demographic census.

In addition to the observational census, we also conducted a transplant experiment to assess survival of shrub recruits at different densities along the ecotone transects. Seeds were collected in 2014, germinated in Rice University greenhouses, and seedlings were brought back to the SEV for transplanting July 25 – August 2, 2015. We planted 48 transplants along each transect, distributed across 12 1m x 1m plots with 4 transplants per 0.5m x 0.5m subplot. Plots were placed unevenly along transects to sample as much variation in shrub density as possible. Transplant survival and size were recorded in October 2015 and again in June 2016, at the time of the observational census.

## 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 description:** Add a brief description for each table

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

**LT\_DemographyData**: Demographic data from annual surveys on tagged individuals.

|  |  |  |  |
| --- | --- | --- | --- |
| Column name | Description | Unit or  code explanation or date format | Empty value code |
| site | One of four different sites | PDC = Palo Duro Canyon  MOD = Middle of Desert  FPS = Five Points Square  SLP = Southern Los Piños |  |
| transect | One of 3 transects at a given site |  |  |
| designated.window | 50-m resolution location on transect | meters |  |
| actual.window | 5-m resolution location on transect | meters |  |
| plant | Unique plant number in a given window |  |  |
| ID | Unique ID consisting of transect, window, and plant |  |  |
| year\_t | Year of measurement, before one year has elapsed |  |  |
| max.ht\_t | Height of the shrub at the tallest point, at time t | centimeters |  |
| max.w\_t | Width of the shrub at the widest point, at time t | centimeters |  |
| perp.w\_t | Width of the shrub in the direction perpendicular to the maximum width, at time t | centimeters |  |
| max.ht\_dead | Height of the shrub at the tallest point, at time t, including defoliated branches | centimeters |  |
| max.w\_dead | Width of the shrub at the widest point, at time t, including defoliated branches | centimeters |  |
| perp.w\_dead | Width of the shrub in the direction perpendicular to the maximum width, at time t, including defoliated branches | centimeters |  |
| flowers\_t | Number of flowers counted, at time t |  |  |
| fruits\_t | Number of fruits counted, at time t |  |  |
| reproductive\_fraction\_t | Fraction of the plant on which flowers and fruits were counted, at time t |  |  |
| node.galls\_t |  |  |  |
| stem.galls\_t |  |  |  |
| bagworms\_t |  |  |  |
| notes\_t | Additional observations at time t |  |  |
| year\_t1 | Year of measurement, after one year has elapsed |  |  |
| new.plant\_t1 | Boolean stating if plant was added to the study | 0 = False, 1 = True |  |
| seedling\_t1 | Boolean stating if plant was a seedling at t1, based on the observer’s judgment | 0 = False, 1 = True |  |
| newplant.meter | Location of new plant on the transect | meters |  |
| survival\_t1 | Boolean stating if plant was alive at t1 | 0 = False, 1 = True |  |
| max.ht\_t1 | Height of the shrub at the tallest point, at time t1 | centimeters |  |
| max.w\_t1 | Width of the shrub at the widest point, at time t1 | centimeters |  |
| perp.w\_t1 | Width of the shrub in the direction perpendicular to the maximum width, at time t1 | centimeters |  |
| flowers\_t1 | Number of flowers counted, at time t1 |  |  |
| fruits\_t1 | Number of fruits counted, at time t1 |  |  |
| reproductive\_fraction | Fraction of the plant on which flowers and fruits were counted, at time t1 |  |  |
| notes\_t1 | Additional observations at time t1 |  |  |

**LT\_TransectData**: Initial size and location data for all individuals within 1 m of the transect, including both tagged and untagged individuals. Data is from the first survey only.

|  |  |  |  |
| --- | --- | --- | --- |
| Column name | Description | Unit or  code explanation or date format | Empty value code |
| site | One of four different sites | PDC = Palo Duro Canyon  MOD = Middle of Desert  FPS = Five Points Square  SLP = Southern Los Piños |  |
| transect | One of 3 transects at a given site |  |  |
| GIS\_id |  |  |  |
| demog\_id |  |  |  |
| window | Location on transect (5-m resolution) | meters |  |
| max.ht | Height of the shrub at the tallest point, at beginning of study | centimeters |  |
| max.w | Width of the shrub at the widest point, at beginning of study | centimeters |  |
| perp.w | Width of the shrub in the direction perpendicular to the maximum width, at beginning of study | centimeters |  |
| max.ht.dead | Height of the shrub at the tallest point, at beginning of study, including defoliated branches | centimeters |  |
| max.w.dead | Width of the shrub at the widest point, at beginning of study, including defoliated branches | centimeters |  |
| perp.w.dead | Width of the shrub in the direction perpendicular to the maximum width, at beginning of study, including defoliated branches | centimeters |  |
| notes | Additional observations at beginning of study |  |  |
| outwindow | Location on transect (5-m resolution), but shrub is > 1 m from transect | meters |  |

**LT\_TransectLengths**: List of all transects and their corresponding lengths.

|  |  |  |  |
| --- | --- | --- | --- |
| Column name | Description | Unit or  code explanation or date format | Empty value code |
| Site | One of four different sites | PDC = Palo Duro Canyon  MOD = Middle of Desert  FPS = Five Points Square  SLP = Southern Los Piños |  |
| Transect | One of 3 transects at a given site |  |  |
| Length\_m | Transect length | meters |  |

**LT\_TransplantExp:** Data from the transplant experiment; includes cover data as well as survival and other demographic data on transplants only.

|  |  |  |  |
| --- | --- | --- | --- |
| Column name | Description | Unit or  code explanation or date format | Empty value code |
| Site | One of four different sites | PDC = Palo Duro Canyon  MOD = Middle of Desert  FPS = Five Points Square  SLP = Southern Los Piños |  |
| Transect | One of 3 transects at a given site |  |  |
| plot | One of 12 plots in a given transect |  |  |
| subplot | One of four subplots within each plot; there were nine 16.67 mm x 16.67 mm squares in each subplot that were used to record cover data |  |  |
| plot\_location | Location of the plot on the transect | meters |  |
| year\_t | Year of measurement, before one year has elapsed |  |  |
| num\_bare\_t | Number of the nine 16.67 mm x 16.67 mm squares in each subplot that were bare |  |  |
| num\_shrub\_t | Number of the nine 16.67 mm x 16.67 mm squares in each subplot that were creosotebush |  |  |
| num\_black\_gramma\_t | Number of the nine 16.67 mm x 16.67 mm squares in each subplot that were black gramma |  |  |
| num\_blue\_gramma\_t | Number of the nine 16.67 mm x 16.67 mm squares in each subplot that were blue gramma |  |  |
| num\_other\_grass\_t | Number of the nine 16.67 mm x 16.67 mm squares in each subplot that were other grasses |  |  |
| num\_other\_t | Number of the nine 16.67 mm x 16.67 mm squares in each subplot that were other plants or ground cover |  |  |
| Cover\_notes | Notes pertaining to ground cover records |  |  |
| max.ht\_t | Height of the shrub at the tallest point, at time t | centimeters |  |
| max.w\_t | Width of the shrub at the widest point, at time t | centimeters |  |
| perp.w\_t | Width of the shrub in the direction perpendicular to the maximum width, at time t | centimeters |  |
| name | Name if the plant received one, as a joke |  |  |
| notes | Notes pertaining to the plot |  |  |
| falll\_survival\_t | Boolean stating if plant was alive at the fall census |  |  |
| fall\_survival\_notes | Notes regarding the fall survival census |  |  |
| spring\_survival\_t1 | Boolean stating if plant was alive at the spring census |  |  |
| max.ht\_t1 | Height of the shrub at the tallest point, at time t1 | centimeters |  |
| max.w\_t1 | Width of the shrub at the widest point, at time t1 | centimeters |  |
| perp.w\_t1 | Width of the shrub in the direction perpendicular to the maximum width, at time t1 | centimeters |  |
| flowers\_t1 | Number of flowers counted, at time t1 |  |  |
| fruits\_t1 | Number of fruits counted, at time t1 |  |  |
| reproductive\_fraction | Fraction of the plant on which flowers and fruits were counted, at time t1 |  |  |

**SD\_Summary**: Contains seed mass and drop height information for the seed drop trials.

|  |  |  |  |
| --- | --- | --- | --- |
| Column name | Description | Unit or  code explanation or date format | Empty value code |
| Trial | Trial number |  |  |
| SMBP | Seed mass before addition of fluorescent powder | grams |  |
| SMAP | Seed mass after addition of fluorescent powder | grams |  |
| Notes | Additional notes about a particular entry |  |  |
| DH | Drop height above ground | meters |  |

**SD\_Trials**: Contains seed position vs time data for each trial.

|  |  |  |  |
| --- | --- | --- | --- |
| Column name | Description | Unit or  code explanation or date format | Empty value code |
| Frame | Frame number in video |  |  |
| Time | Real time corresponding to frame and object position; in increments of 1/240 since camera framerate is 240 FPS | seconds |  |
| TXX.y | Distance fallen as a function of time for trial XX | meters |  |

## Articles

List articles that cite this dataset (optional)

|  |  |  |
| --- | --- | --- |
| Article DOI or URL (DOI is preferred) | Article title | Journal title |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

## 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 |
| 00\_RunAll | Runs scripts 01-07 | R |
| 01\_SeedVelocities | Estimates seed terminal velocities from seed drop data | R |
| 02\_WindSpeeds | Generates probability distributions of wind speeds from weather data | R |
| 03\_Dispersal | Constructs dispersal kernel functions | R |
| 04\_CDataPrep | Prepares demography, transect, and transplant data for analysis | R |
| 05\_CDataAnalysis | Analyses demography data using generalised mixed-effect models | R |
| 06\_SIPM | Constructs spatial integral projection model functions | R |
| 07\_MainFigures | Generates figures to be used in the manuscript | 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.

N/A

## QA/QC Procedures

Describe how the data were checked for accuracy. (e.g., "Data were range checked using a SAS program." *or* "Data were entered twice and then compared to find errors.")

## Notes and Comments

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

We currently plan to have the data released by 01 January 2021 or when the manuscript is submitted, whichever occurs later.