

concordance=TRUE

popler: an R package for extraction of population time series from long-term ecological research (LTER) data

Aldo Compagnoni^{*a,b,c}, Andrew J. Bibian^a, Brad M. Ochocki^a, Sam Levin^{b,c}, Kai Zhu^d and Tom E.X. Miller^a

^aDepartment of BioSciences, Program in Ecology and Evolutionary Biology, Rice University, 6100 Main St, MS-170, Houston, TX 77005

^bInstitute of Biology, Martin Luther University Halle-Wittenberg, Am Kirchtor 1, 06108 Halle (Saale), Germany

^cGerman Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany

^dDepartment of Environmental Studies, University of California, Santa Cruz, CA 95064, USA

Running headline: The popler database and R package

^{*}aldo.compagnoni@gmail.com

Abstract

1. Population dynamics play a central role in the historical and current development of fundamental and applied ecological science. The nascent culture of open data promises to increase the value of population dynamics studies to the field of ecology. However, the access and use of population data is constrained by the difficulty in identifying relevant datasets, by the heterogeneity of available data, and by access to raw (as opposed to aggregated or derived) observations.
2. To obviate these issues, we built a relational database, `popler`, and its R client, `library("popler")`. `popler` accommodates the vast majority of population data under a common structure, and without the need for aggregating raw observations. `library("popler")` is designed for users unfamiliar with the structure of the database and with the SQL language. This R library allows users to identify, download, explore, and cite datasets salient to their needs.
3. We implemented `popler` as a PostgreSQL instance, where we stored population data originated by the United States Long Term Ecological Research (LTER) Network. Our focus on the US LTER data aims to leverage the untapped potential of this vast open data resource. The database currently contains 305 datasets from 25 LTER sites. `popler` is designed to accommodate automatic updates of existing datasets, and to accommodate additional datasets from LTER as well as non-LTER studies.
4. The combination of the online database and the R library(`"popler"`) is a unique data resource in population ecology. The common structure of `popler` simplifies comparative analyses, and the availability of raw data confers flexibility in data analysis. `library("popler")` maximizes these opportunities by providing a user-friendly interface to the online database.

Keywords

- 1 open long-term population data, US Long Term Ecological Research Network data, online
- 2 database, database structure, PostgreSQL, R package, comparative analysis

3 Introduction

Population dynamics – changes in species’ abundance and composition through time and space – are central to ecology for both applied and fundamental reasons. Populations are the building blocks of ecological dynamics at higher scales of organization, and examples abound showing how the study of population ecology improves understanding in evolution (?), community ecology (?), and ecosystem ecology (??). Given their central role, studies of population dynamics will be an essential component in the advances allowed by the flourishing culture of open access.

The increase in freely available data is poised to change ecological science (?). The rising focus on open data is clear in changing publishing standards, in the design of observational networks (?), and in the availability of previously proprietary data (??). This deluge of open data holds promise to facilitate comparative analyses and to test the generality of ecological hypotheses. For population dynamics in particular, it is the increasing availability of long-term data that will likely yield the most substantial scientific advances, as long time series are required to detect trends in abundance (?), quantify temporal variance (?), and identify endogenous (?) or exogenous (?) drivers of population fluctuations.

There are currently three public databases that provide time series of population data. These are the Global Population Dynamics Database (GPDD, ?), the Living Planet Index (?), and BioTIME (?). These databases are an important resource for population biologists (e.g., ?), but their characteristics make them optimal for a specific set of analyses. For example, the GPDD time series contain only one observation of population size or density per temporal replicate, BioTIME focuses on assemblage (i.e. multispecies) datasets, and the Living Planet Index contains information on single populations of conservation concern. These differences can be decisive in scientific inference. For example, LPI data indicate worldwide biodiversity declines, while BioTIME data indicate stable biodiversity due to higher species turnover. This is likely due to the focus of the LPI on species of conservation concern (?). Finally, none of these three databases provides much data from experiments.

One of the best sources of publicly available long-term data is the Long-Term Ecological Research (LTER) network. The LTER was founded in 1980 and grew from the original six sites to, as of 2016, 28 sites throughout North America, Puerto Rico, French Polynesia, and Antarctica.

32 Synthetic and comparative studies from the LTER network have made valuable contributions
33 to ecological understanding (?). However, the majority of LTER synthesis research has focused
34 on ecological dynamics at the community (e.g. ?) and ecosystem (e.g. ?) scales. Nevertheless,
35 every LTER site collects population abundance data as one of its five core areas of continuous
36 observations (?). In our opinion these data, which have been accumulating since 1980, are
37 under-used.

38 LTER population data provides two distinct advantages compared to existing databases.
39 First, LTER data contains both single-species and assemblage datasets that might be free from
40 the biases suggested for the LPI database. Assemblage datasets are expected to be an unbiased
41 reflection of biodiversity trends (?), and LTER single-species studies are generally not focused
42 on species of conservation concern. Second, many of the analyses on LTER experiments were
43 published a few years after the start of manipulations. Hence, analysis of updated data from
44 these LTER experiments could provide unique scientific insights.

45 One issue that may limit the use of LTER population data is their heterogeneity. The struc-
46 ture of LTER data sets may be widely different, employing a variety of data types (counts of
47 individuals, biomass estimates, percent cover, etc.), experimental designs driven by the priori-
48 ties of particular PIs, and diverse replication schemes – idiosyncrasies that may be difficult to
49 accommodate in a one-size-fits-all database. However, these challenges also present valuable
50 opportunities. For example, the hierarchical replication structure of many LTER studies (e.g.,
51 subplots within plots within transects) can facilitate more sophisticated statistical investigation
52 than would be possible with simpler, aggregated, or unreplicated data.

53 To overcome the issues posed by heterogeneous data structures, we developed `popler` (POP-
54 ulation dynamics in Long-term Ecological Research), an online database of LTER population
55 studies. This database defines a common data structure that can accommodate in principle all
56 population data, and its SQL environment allows updates whenever new data becomes available.
57 We also developed a companion R package to facilitate the identification, access, and manipula-
58 tion of raw and heterogeneous population data. Our goals here are to provide introductions to
59 the database and package. We focus on LTER time series, but expanding `popler` beyond the
60 LTER network is a priority for future development.

61 The `popler` database

62 To combine population data from the LTER network using a common structure, we identified
63 a set of relevant variables (Table ??) and organized them into a relational database. Here, we
64 present the structure of the database in Fig. ??, and we provide a simplified entity relationship
65 diagram (ERD) in the supplementary material (Fig. ??). In `popler` we stored “raw” data,
66 meaning that we have not modified, edited, or aggregated the original observations.

67 For inclusion in `popler`, we only considered studies that included (1) repeated observations
68 of populations or individuals through time, (2) at least five population censuses (as of database
69 creation in 2017), and (3) taxonomic information associated with abundance observations (e.g.,
70 we excluded time series of functional groups). We provide technical details of database creation
71 in Appendix S1.

72 The `popler` database currently contains data from 305 studies (122 of which are experimen-
73 tal) representing 4377 cumulative years of observations. On average, studies in `popler` contain
74 10.5 years of data (median: 7), with the longest study containing 67. The sampling designs are
75 predominantly yearly (49%) and sub-yearly (44%), and only 6% of designs sampled populations
76 irregularly or less often than yearly. `popler` also contains abundant spatial replication, with
77 studies containing a mean of 295 (median: 72) unique spatial replicates distributed across an
78 average of 2.4 (median: 2) nested spatial replication levels. Finally, `popler` contains data from
79 665 plant species, 382 animal species, and 1 fungal species.

80 Population data

81 We define “population data” as time-series of observations on the size or density of a population
82 of a species or other taxonomic unit. Observations of population size are stored in a variable
83 called `abundance_observation` and can be measured as a count, biomass, density, or cover.
84 These four types of population data are stored in the homonymous tables of the database (Fig.
85 ??A).

86 The population datasets contained in `popler` are always replicated temporally. Temporal
87 replicates are identified with up to three variables: `year`, `month`, and `day`. Population data are
88 also almost always spatially replicated, and spatial replicates are often nested, where for example

89 a study might include separate sites, each of which contains intermediate spatial replicates (e.g.
 90 a transect, a block), which in turn contain the smallest spatial replicate at which observations are
 91 made (e.g. a plot, a quadrat). The hypothetical study described above would have three nested
 92 levels of spatial replication, identified by three numbered `spatial_replication` variables.
 93 In `popler`, we accommodate data sets with up to five spatial replication levels (Table ??). We
 94 call the first and therefore largest spatial replicate “study site” (Fig. ??C). Note that this does
 95 not refer to the LTER site, one of the 28 NSF-supported locations (Table ??).

96 `popler` contains both observational and experimental studies. Experimental datasets con-
 97 tain information on one or more experimental treatments. `Popler` accommodates information on
 98 up to three experimental treatments, identified by three numbered `treatment_type` variables
 99 (Table ??).

100 Most datasets also contain one or more variables in addition to the ones described above
 101 which we store in a character variable called `covariates` (Table ??). These are variables that
 102 do not conform to our data model. `covariates` stores in each row, the content an arbitrary
 103 number of such non-conforming variables. `covariates` can be useful, for example, for time
 104 series that contain information on population structure. In these datasets, observations on
 105 population size are grouped based on subdivisions of the entire population, such as males and
 106 females, large and small individuals, etc. We identify these datasets through a variable in the
 107 metadata `structured_data` (Table ??).

108 Finally, in addition to time series of abundance, `popler` contains individual-level data. This
 109 data provides information on the attributes of the individuals, or a subset thereof, that make
 110 up a population. We store this information in a dedicated table (“Individual”, Fig. ??A). As
 111 individual attributes we consider variables that describe identity, size, sex, life stage or status
 112 (e.g. reproductive or non-reproductive). We refer to these individual attributes with the term
 113 “structure”: `popler` accommodates data sets that measure up to four types of structure simul-
 114 taneously. We store these data in up to four numbered `structure_type` variables. While
 115 these data are not population time series; we chose to include them in `popler` because they
 116 provide information on demographic transitions that can be used to derive estimates of popula-
 117 tion growth. Moreover, in the cases of datasets that sample all of the individuals in a population,
 118 individuals can be aggregated (i.e. summed) as a measure of population size.

119 Taxonomic information

120 Each observation corresponds to a taxonomic unit (Fig. ??B), typically a species or a genus, but
121 also include data that refer to a higher taxonomic rank, such as family, or order. `popler` provides
122 15 taxonomic ranks, and two additional variables that refer to how taxonomic information is
123 recorded in the original datasets. The additional variables are `sppcode`, which are taxon-specific
124 alphanumeric codes, and `common_name`, the common name of each taxonomic unit (Table S1).
125 `popler` also allows to store accepted taxonomic information in an additional table (Fig. ??B).
126 This table accounts for ambiguities contained in the raw taxonomic data, which originate by
127 the dynamic changes in species classifications (?). Further versions of `popler` will populate this
128 second table with the accepted taxonomic units (which include taxonomic information above the
129 level of genus) provided by the R package `taxize` (?).

130 Study site

131 We stored the locations of datasets by recording the latitude (`lat_study_site`) and longitude
132 (`lng_study_site`) of study sites (Fig. ??C). Storing this information in a separate table
133 allows for explicit connections between independent data sets collected at the same locations
134 within LTER sites.

135 Metadata

136 The metadata table (Table ??) provides information on temporal and spatial replication, and
137 study design (Fig. ??D), including title, link to online metadata, contact information for data
138 originators, and the type of data provided by the dataset (i.e., which of the five tables in
139 Fig. ??A the data is stored in). All remaining metadata is related to the variables stored
140 in the tables of ??A and ??B. First, some population datasets subdivide the population in
141 groups that share the same characteristic (e.g. sex, developmental stage, age). These datasets,
142 however, are not individual data (Fig. ??D). We flag these datasets through the variable
143 `structured_data`. Second, we provide the years elapsed between the first and last obser-
144 vation (`duration_years`), and the sampling frequency (`samplefreq`). Third, we provide
145 the number of levels of nested spatial replicates, and the number of replicates for each spa-

146 tially nested level. Fourth, we show whether studies focus on a single species or on multiple
147 species through the `community` variable. Fifth, we identify studies as observational or ex-
148 perimental (`studytype`). If a study is experimental, we provide information on the type of
149 treatments imposed by the study (`treatment_type_n`) and, when available, which one is the
150 control treatment (`control_group`). Finally, we report information on the data stored in the
151 `abundance_observation` variable: its units of measure (`samplingunits`), the area over
152 which this abundance data was observed (`spatial_replication_level_n_extent` and
153 `spatial_replication_level_n_extent_units`), and in case the data was aggregated
154 across space or time we flag these data as derived (`derived`).

155 The **popler** package

156 The `popler` R package consists of three core functions that allow users to browse and retrieve
157 data from the database (Fig. ??). In order of intended use, these functions are: `pplr_dictionary()`,
158 `pplr_browse()`, and `pplr_getdata()`.

159 The **pplr_dictionary()** function

160 The dictionary function is a good place for new users to begin working with `popler` (Fig. ??).
161 With no arguments provided, this function returns a subset of the most useful metadata variables
162 associated with each dataset (Fig. ??). Providing argument `full_tbl = TRUE` returns all
163 77 metadata variables. Each one of these variable names can be provided as an argument
164 to `pplr_dictionary()`, which then returns the possible unique values of the variable. For
165 example, `pplr_dictionary(lterid)` returns the three letter codes of the LTER network sites
166 included in `popler`. For numeric variables such as `duration_years`, `pplr_dictionary()`
167 returns a summary including quantiles, mean, and median.

168 The **pplr_browse()** function

169 Once the user is familiar with the meaning and content of the variables that define `popler`
170 datasets, they are ready to dig deeper using `pplr_browse()` (Fig. ??). Running `pplr_browse()`
171 without arguments provides the metadata from the entire contents of the database. This will be

172 a 305by20 data frame, with each row corresponding to a study and each column corresponding
173 to a variable defined by `pplr_dictionary()`.

174 The full strength of `pplr_browse()` is achieved by subsetting studies according to desired
175 criteria using logical expressions. For example, the user might want to consider only studies
176 whose duration is 30 years or greater, which can be subsetted with:

```
LTER_30 <- pplr_browse( duration_years > 29)
```

177 This operation will create the object `LTER_30`, which provides metadata for the data sets
178 that satisfy the specified criterion. Multiple criteria may be combined. For example, 30+ year
179 studies of plants can be browsed with

```
LTER_30_plants <- pplr_browse( duration_years > 29 &  
                               kingdom == "Plantae")
```

180 To facilitate data exploration, `pplr_browse()` output can be printed in a more readable
181 settings by providing `report = TRUE` as an argument, which opens up a formatted html doc-
182 ument. The metadata provided by `pplr_browse()` not only contains information on the
183 characteristics of a study but also information on how to cite the study, its unique identifiers,
184 including digital object identifier (DOI), and the contact information of study PIs.

185 The `pplr_get_data()` function

186 Once data sets of interest have been identified, `pplr_get_data()` downloads the data from a
187 server that hosts the database. This function can take as its first argument a browse object, a
188 logical expression, or both. The data downloaded from `popler` are in “long” form, meaning that
189 each row of data reports a single measure of population size, and separate variables indicate the
190 temporal and spatial replicate, taxa, etc. This format makes it easy to further subset downloaded
191 datasets with the aim of visualization and analysis.

192 Ancillary functions

193 `popler` also provides three additional functions to open the url of the original dataset, un-
194 pack covariates, and provide a citation for each dataset. First, the function `metadata_url()`

195 launches the online study description in a web browser. Second, the `cov_unpack()` function
196 transforms the `covariates` variable into a data frame (which `pplr_get_data()` does not
197 provide by default). Third, `pplr_citation()` generates a citation for the originators of each
198 data set.

199 Limitations and opportunities for development

200 Working with raw, spatially replicated, and non-aggregated data provides key advantages in
201 quantitative analyses of population dynamics which were a driving force behind the development
202 of `popler`. However, users need to examine individual datasets and the associated online study
203 descriptions to understand their peculiarities. Single datasets have unique idiosyncrasies that
204 require vetting. For example, many datasets have gaps or changes in the sampling design during
205 the length of the study, or the `covariates` variable can hold key information. Hence, we urge
206 authors to consult the online documentation of the original datasets.

207 In the future, there are opportunities to increase the size of `popler` and expand its scope.
208 First, because many of the studies included in `popler` are ongoing, there will be opportunities to
209 run regular updates aimed at including new observations in `popler`. Second, because our schema
210 (Fig. ??) is very general, the database could be expanded to include population datasets outside
211 of the LTER network. Third, it would be valuable to explicitly associate `popler`'s population-
212 level data with environmental drivers, especially climate. Thus, it is our intention and hope that
213 the resources provided by `popler` will advance ecological understanding of population dynamics
214 within the LTER network, and more generally.

215 Acknowledgements

216 We thank Trevor Drees and Michael Saucedo for assistance in database development, Maurizio
217 Compagnoni for assistance in database management, and Scott Chamberlain for developing the
218 API to query the online database. Support for database and package development was provided
219 by the US National Science Foundation to TEXM (DEB-1543651). This research was additionally
220 supported by a Julian Huxley Faculty Fellowship from Rice University and a Faculty Research

221 Grant awarded by the Committee on Research from the University of California, Santa Cruz
222 (KZ). The LTER network is supported by the US National Science Foundation.

223 **Authors' contributions**

224 AC, AB, KZ, MO, TEXTM designed and built the database. AC AB, KZ, BD, SM, and TEXTM
225 designed and built the R package. AC and TEXTM led the writing of the manuscript. All authors
226 contributed to manuscript drafts and gave final approval for publication.

227 **Data Availability**

228 The `popler` R package is publicly available at <https://github.com/ropensci/popler>.

Table 1: Variables used to store population or individual data in `popler`.

Variable	Description
<code>abundance_observation</code>	Measure of population abundance at a specific time and location. This variable measures abundance as a count, biomass, density, or cover. For individual data sets this variable is always equal to 1, because each attribute or set of attributes refer to a single individual.
<code>day</code>	Day of observation
<code>month</code>	Month of observation
<code>year</code>	Year of observation
<code>spatial_replicate_n</code>	The n^{th} level of spatial replication, where <code>spatial_replicate_1</code> is the study site. <code>popler</code> accommodates up to five levels of spatial replication.
<code>treatment_type_n</code>	For datasets originating from an experimental study, the n^{th} treatment. <code>popler</code> accommodates up to three treatments.
<code>covariates</code>	Ancillary observations that do not fall into the standard schema of <code>popler</code> .
<code>structure_type_n</code>	For individual data, these variables measure the n^{th} attribute of individuals (identity, size, sex, status, stage). <code>popler</code> accommodates up to four structure types per dataset.

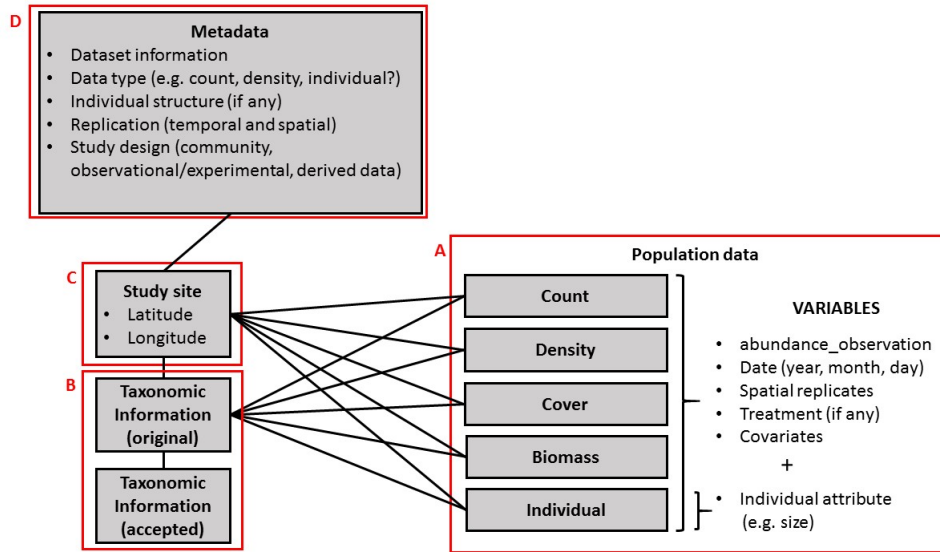


Figure 1: Schematic representation of the entity relationship diagram of the `popler` database. `popler` provides metadata on the studies that originated abundance data points (D). This metadata contains information on the unique identifiers of each study, on its design (observational or experimental), temporal and spatial replication. `popler` stores the latitude and longitude of the study site (C). Each abundance data point corresponds to a specific taxonomic unit (B). Finally, the time series of population data collected in a study can be of four different types (count, density, biomass, cover), or they may be individual data with attributes such as size or sex (A).

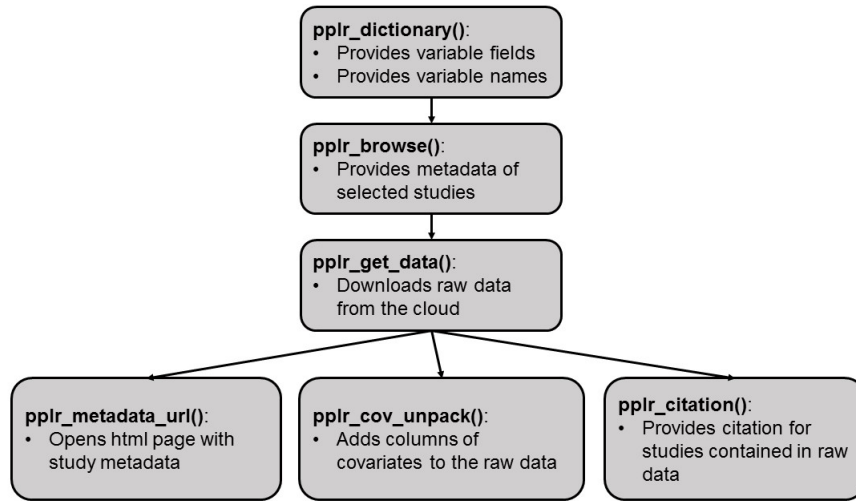


Figure 2: Suggested workflow when using the `popler` R package to interface with the homonymous online database. The function `pplr_dictionary()` refers to the variables of the metadata that describe the data sets contained in `popler`. `pplr_dictionary()` describes these variables and returns their possible values. This information advises which criteria to use when subsetting `popler`. The user can provide a criterion (that is, a logical statement) to browse the metadata, using `pplr_browse()`, or to download data using `pplr_get_data()`. Moreover, the output of `pplr_get_data()` (a data frame) can be the argument of three ancillary functions: `pplr_metadata_url()` opens the webpage containing the original dataset and their associated online metadata. `pplr_cov_unpack()` can be used to format the covariates contained in a raw data object into separate columns of a data frame. Finally, `pplr_citation()` provides a citation for the downloaded data set(s).

229 Appendix S1: Pre-processing **popler** data

230 Before uploading datasets into the online **popler** database, we combined datasets, transformed
231 datasets from wide to long form, converted non-ASCII characters, and modified ambiguous study
232 site names.

233 The variables of many datasets were contained in two or more separate files, which we com-
234 bined in a single file. When the original dataset provided data in wide form, we transformed
235 it into long form. In wide form datasets, abundance data associated with different species was
236 stored in separate columns. **popler** stores these datasets in long form, whereby each row of
237 abundance data is related to a specific taxonomic unit in the table containing taxonomic infor-
238 mation (Fig. ??B). We converted all data in ASCII format, because the encoding of the database
239 is the UTF-8. We often re-defined study site names to unambiguously associate them with one
240 of the 26 LTER sites. Many site names are alphanumeric codes (e.g. “U1”) which can overlap
241 across several LTER sites. Hence, we changed site names following a standard formula (namely,
242 from “U1” to “site_sbc_U1”, where “sbc” refers to the Santa Barbara coastal LTER site).

243 In a handful of cases, we removed single data rows from the original dataset. These data
244 rows were associated with two types of typos in the original dataset. First, some abundance
245 observations were not associated with a time of observation. We removed this data because
246 **popler** can only accommodate population information associated with a time of observation.
247 Second, a handful of abundance data points were clear typos (e.g. the letter “l” instead of a
248 numeric value). We substituted these data points with a missing value. We uploaded these
249 pre-processed datasets in the **popler** database through a Graphic User Interface developed in
250 Python using libraries **panda** and **pyqt5**.

Table S1: Taxonomic variables contained in the popler table on original taxonomic information.

Variable
sppcode
kingdom
subkingdom
infrakingdom
superdivision
division
subdivision
superphylum
phylum
subphylum
class
subclass
order
family
genus
species
common_name

Table S2: Metadata variables used to describe the datasets stored in popler.

Variable	Description
proj_metadat_key	Unique ID
lter_project_key	ID of LTER site
lter_project_key	ID of LTER site
title	Title of study
samplingunits	Unit of measure (if any) referred to population data.
datatype	Data type: count, biomass, cover, density, and individual. These correspond to the tables in Fig. 1A.
structured_data	If data type is not individual, but the abundance observations refer to sub-groups of the population based on, for example, sex, developmental stage, or age)
structured_type_n	If individual data, this shows what type of structure is stored. A study can contain up to $n = 4$ types of structure.
structured_type_n_units	Unit of measure (if any) referred to structure data.
studystartyr	Start year of the study
studyendyr	End year of the study
duration_years	Duration of the study in years
samplefreq	Frequency of population census

<code>studytype</code>	Whether study is observational or experimental
<code>community</code>	Whether study includes single taxon (<code>community = F</code>) or multiple taxa (<code>community = T</code>)
<code>spatial_replication_level_n_extent</code>	Extent of spatial replication level number n . A dataset can have up to to 5 replication levels.
<code>spatial_replication_level_n_extent_units</code>	Unit of spatial extent of the n spatial replication level.
<code>spatial_replication_level_n_label</code>	Label of the spatial replication level (e.g. transect, plot, quadrat, ect.). The label of spatial replication level 1 is "site".
<code>spatial_replication_level_n_number_of_unique_reps</code>	The number of unique replicates for the n th level of spatial replication.
<code>treatment_type_n</code>	The type of treatment (e.g. resource manipulation). A study can contain up to $n = 3$ treatments.
<code>control_group</code>	If study is experimental, this shows the field(s) that identify the control replicate.
<code>derived</code>	Is population size data raw, or is it derived (e.g. it is aggregated)?
<code>authors</code>	Author(s) of the original dataset

authors_contact	Email address(es) of the author(s) associated with the original dataset.
metalink	url of the original dataset
knbid	Knowledge Network for Biocomplexity identifier.

Table S3: LTER identification acronyms and their meaning as used in the popler database.

Variable	LTER name
AND	Andrew Forest LTER
ARC	Arctic LTER
BES	Baltimore Ecosystem Study
BNZ	Bonanza Creek LTER
CAP	Central Arizona - Phoenix LTER
CCE	California Current Ecosystem LTER
CDR	Cedar Creek Ecosystem Science Reserve LTER
CWT	Coweeta LTER
FCE	Florida Coastal Everglades LTER
GCE	Georgia Coastal Ecosystems LTER
HBR	Hubbard Brook LTER
HFR	Harvard Forest LTER
JRN	Jornada Basin LTER
KBS	Kellogg Biological Station LTER
KNZ	Konza Prairie LTER
LNO	LTER Network Office
LUQ	Luquillo LTER
MCM	McMurdo Dry Valleys LTER
MCR	Moorea Coral Reef LTER
NCO	LTER Network Communications Office
NTL	North Temperate Lakes LTER
NWT	Niwot Ridge LTER
PAL	Palmer Antarctica LTER
PIE	Plum Island Ecosystems LTER
SBC	Santa Barbara Coastal LTER
SEV	Sevilleta LTER
SGS	Shortgrass Steppe LTER
VCR	Virginia Coastal Reserve LTER

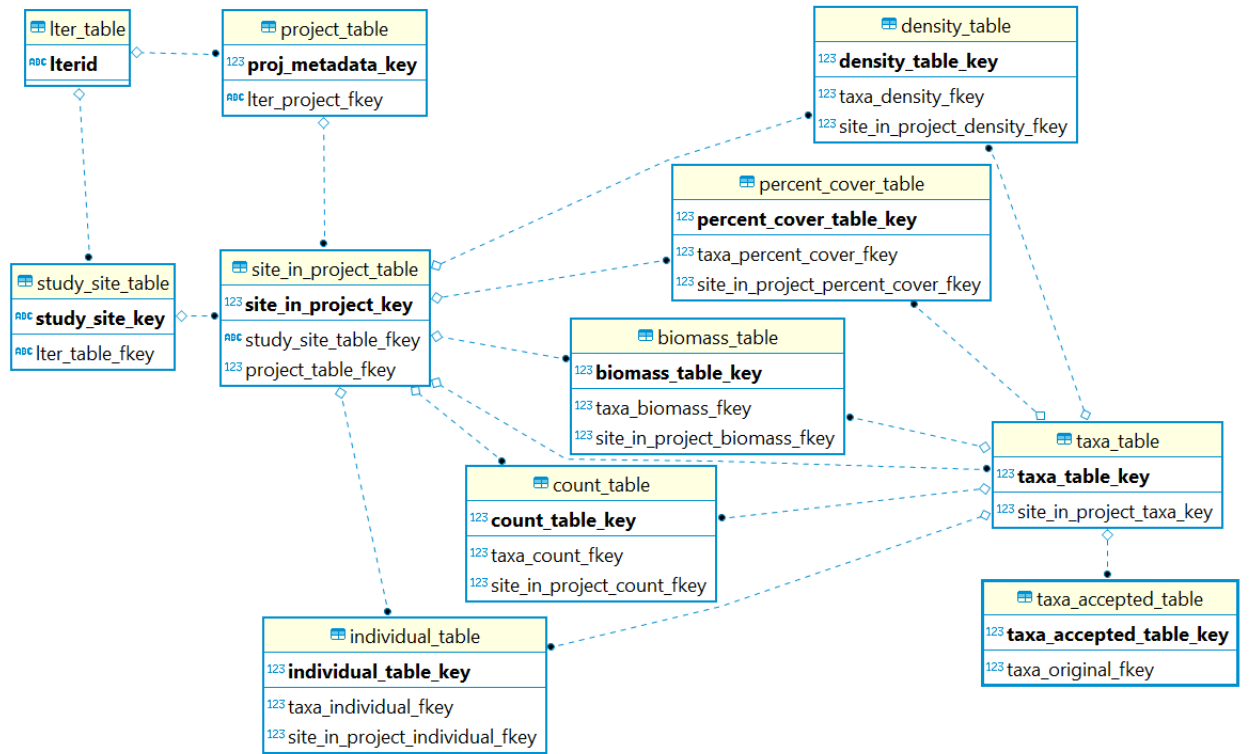


Figure S1: Simplified entity relationship diagram of the popler database. This figure shows table names, primary keys, and foreign keys of the popler database.