

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

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, synthesis 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 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 `library("popler")` is a resource for data synthesis efforts 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, data synthesis, 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 (Metcalf and Pavard, 2007), community ecology (Levine and HilleRisLambers, 2009), and ecosystem ecology (Medvigy et al., 2009; Fisher et al., 2018). Given their central role, studies of population dynamics will be an essential component in the advances allowed by the flourishing culture of open access and data synthesis.

The increase in freely available data is poised to change ecological science (Laurance et al., 2016). The rising focus on open data is clear in changing publishing standards, in the design of observational networks (Schimel et al., 2007), and in the availability of previously proprietary data (Kratz et al., 2003; Bechtold et al., 2005). 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 (Lindenmayer et al., 2012), quantify temporal variance (Compagnoni et al., 2016), and identify endogenous (Knape and de Valpine, 2012) or exogenous (Hampton et al., 2013) 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, Inchausti and Halley, 2001), the Living Planet Index (LPI, Loh et al., 2005), and BioTIME (Dornelas et al., 2018). These databases are an important resource for population biologists (e.g., Knape and de Valpine, 2012), but their characteristics make them optimal for a specific set of analyses. For example, the GPDD could limit the flexibility and power of statistical analyses, because its time series contain only one observation of population size per temporal replicate. Moreover, when comparing the LPI with BioTIME, 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 (Dornelas et al., 2019). The fact that BioTIME contains assemblage (i.e.

multispecies) datasets might provide an advantage in assessing biodiversity trends. Finally, none of these three databases provides much experimental data.

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. Synthetic and comparative studies from the LTER network have made valuable contributions to ecological understanding (Knapp et al., 2012). However, the majority of LTER synthesis research has focused on ecological dynamics at the community (e.g. Wilcox et al. (2017)) and ecosystem (e.g. Knapp and Smith (2001)) scales. Nevertheless, every LTER site collects population abundance data as one of its five core areas of continuous observations (Callahan, 1984). In our opinion these data, which have been accumulating since 1980, are under-used.

LTER population data provides two distinct advantages compared to existing databases. First, both assemblage and single-species LTER dataset should be a resource in quantifying biodiversity. Assemblage datasets are expected to be an unbiased reflection of biodiversity trends (Dornelas et al., 2019), and LTER single-species studies are generally not focused on species of conservation concern. Second, many of the analyses on LTER experiments were published a few years after the start of manipulations. Hence, analysis of updated data from these LTER experiments could provide novel scientific insights.

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

To overcome the issues posed by heterogeneous data structures, we developed `popler` (POPulation dynamics in Long-term Ecological Research), an online database of LTER population studies. This database defines a common data structure that can accommodate in principle all population data, and its SQL environment allows updates whenever new data becomes available.

62 We also developed a companion R package to facilitate the identification, access, and manipula-
63 tion of raw and heterogeneous population data. Our goals here are to provide introductions to
64 the database and package. We focus on LTER time series, but expanding popler beyond the
65 LTER network is a priority for future development.

66 The popler database

67 To combine population data from the LTER network using a common structure, we identified
68 a set of relevant variables (Table 1) and organized them into a relational database. Here, we
69 present the structure of the database in Fig. 1, and we provide a simplified entity relationship
70 diagram (ERD) in the supplementary material (Fig. ??). In popler we stored “raw” data,
71 meaning that we have not modified, edited, or aggregated the original observations.

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

77 The popler database currently contains data from 305 studies (122 of which are experimen-
78 tal) representing 4377 cumulative years of observations. On average, studies in popler contain
79 10.5 years of data (median: 7), with the longest study containing 67. The sampling designs are
80 predominantly yearly (49%) and sub-yearly (44%), and only 6% of designs sampled populations
81 irregularly or less often than yearly. popler also contains abundant spatial replication, with
82 studies containing a mean of 295 (median: 72) unique spatial replicates distributed across an
83 average of 2.4 (median: 2) nested spatial replication levels. Finally, popler contains data from
84 665 plant species, 382 animal species, and 1 fungal species.

85 Population data

86 We define “population data” as time-series of observations on the size or density of a population
87 of a species or other taxonomic unit. Observations of population size are stored in a variable
88 called `abundance_observation` and can be measured as a count, biomass, density, or cover.

89 These four types of population data are stored in the namesake tables of the database (Fig. 1A).

90 The population datasets contained in `popler` are always replicated temporally. Temporal
91 replicates are identified with up to three variables: `year`, `month`, and `day`. Population data are
92 also almost always spatially replicated, and spatial replicates are often nested, where for example
93 a study might include separate sites, each of which contains intermediate spatial replicates (e.g.
94 a transect, a block), which in turn contain the smallest spatial replicate at which observations are
95 made (e.g. a plot, a quadrat). The hypothetical study described above would have three nested
96 levels of spatial replication, identified by three numbered `spatial_replication` variables.
97 In `popler`, we accommodate data sets with up to five spatial replication levels (Table 1). We
98 call the first and therefore largest spatial replicate “study site” (Fig. 1C). Note that this does
99 not refer to the LTER site, one of the 28 NSF-supported locations (Table ??).

100 `popler` contains both observational and experimental studies. Experimental datasets con-
101 tain information on one or more experimental treatments. `Popler` accommodates information on
102 up to three experimental treatments, identified by three numbered `treatment_type` variables
103 (Table 1).

104 Most datasets also contain one or more variables in addition to the ones described above which
105 we store in a character variable called `covariates` (Table 1). These are variables that do not
106 conform to our data model. `covariates` stores in each row, the content of an arbitrary number
107 of such non-conforming variables. `covariates` can be useful, for example, for time series that
108 contain information on population structure. In these datasets, observations on population size
109 are grouped based on subdivisions of the entire population, such as males and females, large and
110 small individuals, etc. We identify these datasets through a variable in the metadata table called
111 `structured_data` (Table ??).

112 Finally, in addition to time series of abundance, `popler` contains individual-level data. This
113 data provides information on the attributes of the individuals, or a subset thereof, that make up a
114 population. We store this information in a dedicated table (“Individual”, Fig. 1A). As individual
115 attributes we consider variables that describe identity, size, sex, life stage or status (e.g. repro-
116 ductive or non-reproductive). We refer to these individual attributes with the term “structure”:
117 `popler` accommodates data sets that measure up to four types of structure simultaneously. We
118 store these data in up to four numbered `structure_type` variables. While these data are not

119 population time series, we chose to include them in `popler` because they provide information on
120 demographic transitions that can be used to derive estimates of population growth. Moreover,
121 in the cases of datasets that sample all of the individuals in a population, individuals can be
122 aggregated (i.e. summed) as a measure of population size.

123 Taxonomic information

124 Each observation corresponds to a taxonomic unit (Fig. 1B), typically a species or a genus,
125 but also include data that refer to a higher taxonomic rank, such as family, or order. `popler`
126 provides 15 taxonomic ranks, and two additional variables that refer to how taxonomic infor-
127 mation is recorded in the original datasets. The additional variables are `sppcode`, which are
128 taxon-specific alphanumeric codes, and `common_name`, the common name of each taxonomic
129 unit (Table S1). `popler` also can store accepted taxonomic information in an additional ta-
130 ble (Fig. 1B). This table accounts for ambiguities contained in the raw taxonomic data, which
131 originate by the dynamic changes in species classifications (Chamberlain and Szöcs, 2013). Fur-
132 ther versions of `popler` will populate this second table with the accepted taxonomic units (which
133 include taxonomic information above the level of genus) provided by the R package `taxize`
134 (Chamberlain and Szöcs, 2013).

135 Study site

136 We stored the locations of datasets by recording the latitude (`lat_study_site`) and longitude
137 (`lng_study_site`) of study sites (Fig. 1C). Storing this information in a separate table allows
138 for explicit connections between independent data sets collected at the same locations within
139 LTER sites.

140 Metadata

141 The metadata table (Table ??) provides information on temporal and spatial replication, and
142 study design (Fig. 1D), including title, link to online metadata, contact information for data
143 originators, and the type of data provided by the dataset (i.e., which of the five tables in Fig.
144 1A the data is stored in). All remaining metadata is related to the variables stored in the tables

145 of 1A and 1B. First, some population datasets subdivide the population in groups that share
146 the same characteristic (e.g. sex, developmental stage, age). These datasets, however, are not
147 individual data (Fig. 1D): we flag them through the variable `structured_data`. Second, we
148 provide the years elapsed between the first and last observation (`duration_years`), and the
149 sampling frequency (`samplefreq`). Third, we provide the number of levels of nested spatial
150 replicates, and the number of replicates for each spatially nested level. Fourth, we show whether
151 studies focus on a single species or on multiple species through the `community` variable. Fifth,
152 we identify studies as observational or experimental (`studytype`). If a study is experimental,
153 we provide information on the type of treatments imposed by the study (`treatment_type_n`)
154 and, when available, which one is the control treatment (`control_group`). Finally, we report
155 information on the data stored in the `abundance_observation` variable: its units of measure
156 (`samplingunits`), the area over which this abundance data was observed
157 (`spatial_replication_level_n_extent` and
158 `spatial_replication_level_n_extent_units`), and in case the data was aggregated
159 across space or time we flag these data as derived (`derived`).

160 **The `popler` package**

161 The `popler` R package consists of three core functions that allow users to browse and retrieve
162 data from the database (Fig. 2). In order of intended use, these functions are: `pplr_dictionary()`,
163 `pplr_browse()`, and `pplr_getdata()`.

164 **The `pplr_dictionary()` function**

165 The dictionary function is a good place for new users to begin working with `popler` (Fig.
166 2). With no arguments provided, this function returns a subset of the most useful metadata
167 variables associated with each dataset (Fig. 1). Providing argument `full_tbl = TRUE` returns
168 all 77 metadata variables. Each one of these variable names can be provided as an argument
169 to `pplr_dictionary()`, which then returns the possible unique values of the variable. For
170 example, `pplr_dictionary(lterid)` returns the three letter codes of the LTER network sites
171 included in `popler`. For numeric variables such as `duration_years`, `pplr_dictionary()`

172 returns a summary including quantiles, mean, and median.

173 The `pplr_browse()` function

174 Once the user is familiar with the meaning and content of the variables that define popler
175 datasets, they are ready to dig deeper using `pplr_browse()` (Fig. 2). Running `pplr_browse()`
176 without arguments provides the metadata from the entire contents of the database. This will be
177 a 305by20 data frame, with each row corresponding to a study and each column corresponding
178 to a variable defined by `pplr_dictionary()`.

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

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

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

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

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

190 The `pplr_get_data()` function

191 Once data sets of interest have been identified, `pplr_get_data()` downloads the data from a
192 server that hosts the database. This function can take as its first argument a browse object, a
193 logical expression, or both. The data downloaded from popler are in “long” form, meaning that

each row of data reports a single measure of population size, and separate variables indicate the temporal and spatial replicate, taxa, etc. This format makes it easy to further subset downloaded datasets with the aim of visualization and analysis.

Ancillary functions

`popler` also provides three additional functions to open the url of the original dataset, unpack covariates, and provide a citation for each dataset. First, the function `metadata_url()` launches the online study description in a web browser. Second, the `cov_unpack()` function transforms the `covariates` variable into a data frame (which `pplr_get_data()` does not provide by default). Third, `pplr_citation()` generates a citation for the originators or each data set.

Limitations and opportunities for development

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

In the future, there are opportunities to increase the size of `popler` and expand its scope. First, because many of the studies included in `popler` are ongoing, there will be opportunities to run regular updates aimed at including new observations in `popler`. Second, because our schema (Fig. 1) is very general, the database could be expanded to include population datasets outside of the LTER network. Third, it would be valuable to explicitly associate `popler`'s population-level data with environmental drivers, especially climate. Thus, it is our intention and hope that the resources provided by `popler` will advance ecological understanding of population dynamics within the LTER network, and more generally.

220 Acknowledgements

221 We thank Trevor Drees and Michael Saucedo for assistance in database development, Maurizio
222 Compagnoni for assistance in database management, and Scott Chamberlain for developing the
223 API to query the online database. Support for database and package development was provided
224 by the US National Science Foundation to TEXM (DEB-1543651). This research was additionally
225 supported by a Julian Huxley Faculty Fellowship from Rice University and a Faculty Research
226 Grant awarded by the Committee on Research from the University of California, Santa Cruz
227 (KZ). The LTER network is supported by the US National Science Foundation.

228 Authors' contributions

229 AC, AJB, KZ, BMO, TEXM designed and built the database. AC, AJB, KZ, BMD, SL, and
230 TEXM designed and built the R package. AC and TEXM led the writing of the manuscript. All
231 authors contributed to manuscript drafts and gave final approval for publication.

232 Data Availability

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

234 References

- 235 W. A. Bechtold, P. L. Patterson, et al. *The enhanced forest inventory and analysis program:*
236 *national sampling design and estimation procedures*, volume 80. US Department of Agriculture
237 Forest Service, Southern Research Station Asheville, North Carolina, 2005.
- 238 J. T. Callahan. Long-term ecological research. *BioScience*, 34(6):363–367, 1984.
- 239 S. A. Chamberlain and E. Szöcs. taxize: taxonomic search and retrieval in r. *F1000Research*, 2,
240 2013.
- 241 A. Compagnoni, A. J. Bibian, B. M. Ochocki, H. S. Rogers, E. L. Schultz, M. E. Sneck, B. D.
242 Elder, A. M. Iler, D. W. Inouye, H. Jacquemyn, and T. E. X. Miller. The effect of demographic
243 correlations on the stochastic population dynamics of perennial plants. *Ecological Monographs*,
244 86(4):480–494, 2016. ISSN 0012-9615. doi: 10.1002/ecm.1228.
- 245 M. Dornelas, L. H. Antao, F. Moyes, A. E. Bates, A. Magorrra, D. Adam, et al. Biotime: a
246 database of biodiversity time series for the anthropocene. *Global Ecology and Biogeography*,
247 2018.
- 248 M. Dornelas, N. J. Gotelli, H. Shimadzu, F. Moyes, A. E. Magurran, and B. J. McGill. A balance
249 of winners and losers in the anthropocene. *Ecology letters*, 22(5):847–854, 2019.
- 250 R. A. Fisher, C. D. Koven, W. R. Anderegg, B. O. Christoffersen, M. C. Dietze, C. E. Farrior,
251 J. A. Holm, G. C. Hurtt, R. G. Knox, P. J. Lawrence, et al. Vegetation demographics in earth
252 system models: A review of progress and priorities. *Global change biology*, 24(1):35–54, 2018.
- 253 S. E. Hampton, E. E. Holmes, L. P. Scheef, M. D. Scheuerell, S. L. Katz, D. E. Pendleton, and
254 E. J. Ward. Quantifying effects of abiotic and biotic drivers on community dynamics with
255 multivariate autoregressive (mar) models. *Ecology*, 94(12):2663–2669, 2013.
- 256 P. Inchausti and J. Halley. Investigating long-term ecological variability using the global popu-
257 lation dynamics database. *Science*, 293(5530):655–657, 2001.

258 J. Knape and P. de Valpine. Are patterns of density dependence in the global population dy-
259 namics database driven by uncertainty about population abundance? *Ecology letters*, 15(1):
260 17–23, 2012.

261 A. K. Knapp and M. D. Smith. Variation among biomes in temporal dynamics of aboveground
262 primary production. *Science*, 291(5503):481–484, 2001.

263 A. K. Knapp, M. D. Smith, S. E. Hobbie, S. L. Collins, T. J. Fahey, G. J. Hansen, D. A. Landis,
264 K. J. La Pierre, J. M. Melillo, T. R. Seastedt, et al. Past, present, and future roles of long-term
265 experiments in the lter network. *BioScience*, 62(4):377–389, 2012.

266 T. K. Kratz, L. A. Deegan, M. E. Harmon, and W. K. Lauenroth. Ecological variability in space
267 and time: Insights gained from the us lter program. *AIBS Bulletin*, 53(1):57–67, 2003.

268 W. F. Laurance, F. Achard, S. Peedell, and S. Schmitt. Big data, big opportunities. *Frontiers*
269 *in Ecology and the Environment*, 14(7):347–347, 2016.

270 J. M. Levine and J. HilleRisLambers. The importance of niches for the maintenance of species
271 diversity. *Nature*, 461(7261):254, 2009.

272 D. B. Lindenmayer, G. E. Likens, A. Andersen, D. Bowman, C. M. Bull, E. Burns, C. R.
273 Dickman, A. A. Hoffmann, D. A. Keith, M. J. Liddell, et al. Value of long-term ecological
274 studies. *Austral Ecology*, 37(7):745–757, 2012.

275 J. Loh, R. E. Green, T. Ricketts, J. Lamoreux, M. Jenkins, V. Kapos, and J. Randers. The living
276 planet index: using species population time series to track trends in biodiversity. *Philosophical*
277 *Transactions of the Royal Society of London B: Biological Sciences*, 360(1454):289–295, 2005.

278 D. Medvigy, S. Wofsy, J. Munger, D. Hollinger, and P. Moorcroft. Mechanistic scaling of ecosys-
279 tem function and dynamics in space and time: Ecosystem demography model version 2. *Journal*
280 *of Geophysical Research: Biogeosciences*, 114(G1), 2009.

281 C. J. E. Metcalf and S. Pavard. Why evolutionary biologists should be demographers. *Trends in*
282 *Ecology & Evolution*, 22(4):205–212, 2007.

283 D. Schimel, W. Hargrove, F. Hoffman, and J. MacMahon. Neon: A hierarchically designed
 284 national ecological network. *Frontiers in Ecology and the Environment*, 5(2):59–59, 2007.

285 K. R. Wilcox, A. T. Tredennick, S. E. Koerner, E. Grman, L. M. Hallett, M. L. Avolio, K. J.
 286 La Pierre, G. R. Houseman, F. Isbell, D. S. Johnson, J. M. Alatalo, A. H. Baldwin, E. W.
 287 Bork, E. H. Boughton, W. D. Bowman, A. J. Britton, J. F. Cahill, S. L. Collins, G. Du,
 288 A. Eskelinen, L. Gough, A. Jentsch, C. Kern, K. Klanderud, A. K. Knapp, J. Kreyling, Y. Luo,
 289 J. R. McLaren, P. Magonigal, V. Onipchenko, J. Prevéy, J. N. Price, C. H. Robinson, O. E.
 290 Sala, M. D. Smith, N. A. Soudzilovskaia, L. Souza, D. Tilman, S. R. White, Z. Xu, L. Yahdjian,
 291 Q. Yu, P. Zhang, and Y. Zhang. Asynchrony among local communities stabilises ecosystem
 292 function of metacommunities. *Ecology letters*, 20:1534–1545, Dec. 2017. ISSN 1461-0248. doi:
 293 10.1111/ele.12861.

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

Variable	Description
abundance_observation	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.
day	Day of observation
month	Month of observation
year	Year of observation
spatial_replicate_n	The n^{th} level of spatial replication, where spatial_replicate_1 is the study site. popler accommodates up to five levels of spatial replication.
treatment_type_n	For datasets originating from an experimental study, the n^{th} treatment. popler accommodates up to three treatments.
covariates	Ancillary observations that do not fall into the standard schema of popler.
structure_type_n	For individual data, these variables measure the n^{th} attribute of individuals (identity, size, sex, status, stage). popler accommodates up to four structure types per dataset.

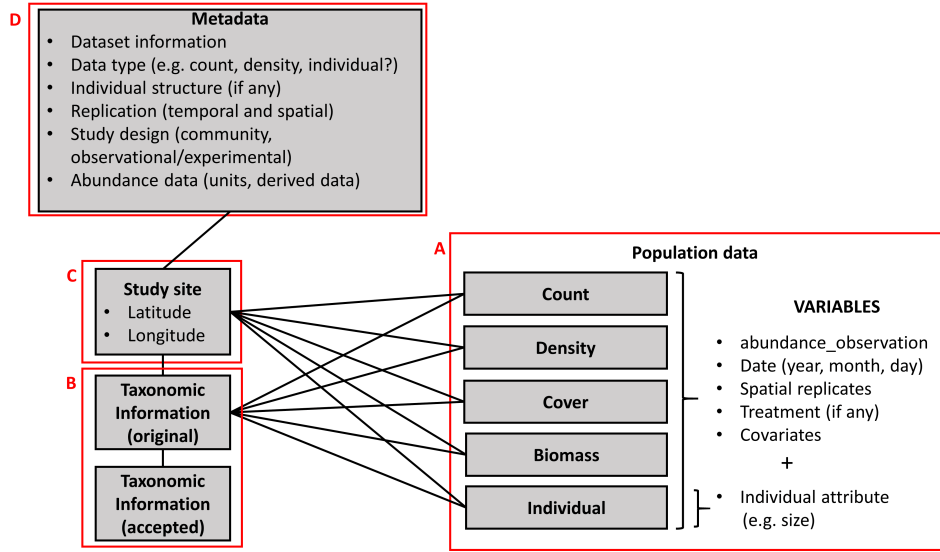


Figure 1: Schematic representation of the entity relationship diagram of the `popler` database. `popler` provides metadata on the studies that originated abundance data points (D). This metadata contains information on the unique identifiers of each study, on its design (observational or experimental), 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).

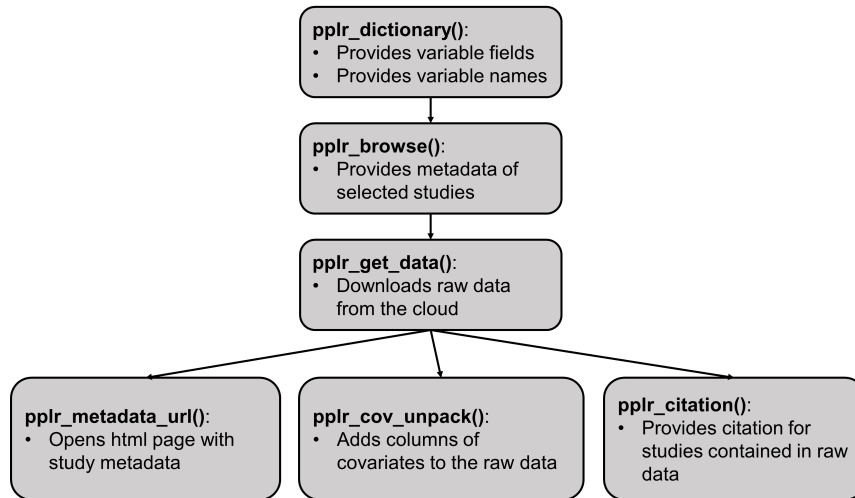


Figure 2: Suggested workflow when using the `popler` R package to interface with the `popler` on-line database. The function `pplr_dictionary()` refers to the variables of the metadata that describe the data sets contained in `popler`. `pplr_dictionary()` describes these variables and returns their possible 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).