portalr: an R package for summarizing and using the Portal Project Data

Erica M. Christensen[1,2]**, Glenda M. Yenni**[1]**, Hao Ye**[1]**, Juniper L. Simonis**[1]**, Ellen K. Bledsoe**[1]**, Renata Diaz**[1]**, Shawn D. Taylor**[1]**, Ethan P. White**[1]**, and S. K. Morgan Ernest**[1]

1 University of Florida **2** Jornada Experimental Range, New Mexico State University

Summary

Data sharing, the practice of making data widely available for use by others, benefits the ecological research community in ensuring study reproducibility and enabling many data sets to be combined to address broad-scale questions (e.g. meta-analyses). However, there are few standards or rules about how ecological data are collected and stored, and so it can be difficult and/or time-consuming to understand how to use unfamiliar data appropriately. `portalr` is an R package (R Core Team, 2016) to facilitate working with the data collected from a long-term ecological study called The Portal Project. All data is freely available in a GitHub repository (https://github.com/weecology/PortalData) and archived on Zenodo (Ernest et al., 2018a). The purpose of `portalr` is to allow new users to quickly and easily gain access to the data and start to use it, and to allow experienced users to produce analyses more efficiently and consistently. The functions contained in this package were developed by the researchers most familiar with the data and are designed to reflect best use practices.

The Portal Project is a long-term ecological study site located in the Chihuahuan Desert, near the town of Portal, Arizona, USA. Data collection spans 1977 to the present, and includes data on small mammals, ants, plants, as well as weather data from an automated weather station. For a complete description of the data and data collection process see (Ernest et al., 2018b). These data have been used in a wide variety of ecological research areas, from competition studies (E. J. Heske, Brown, & Mistry, 1994) to ecological forecasting (White et al., 2018).

The `portalr` package contains functions to download the most up-to-date version of the data from GitHub and Zenodo, and to summarize or manipulate the data in several common ways. For example, plant abundance data can be obtained at the plot level or aggregated to site level; small mammal data can be obtained as time series of abundances, biomass, or metabolic energy; and records for "unknown" species or incomplete records (containing "NAs") can be stripped from the data if desired. See https://weecology.github.io/portalr/index.html for documentation.

Acknowledgements

References

Ernest, S. K. M., Yenni, G. M., Allington, G., Bledsoe, E. K., Christensen, E. M., Diaz, R., Geluso, K., et al. (2018a, October). Weecology/portaldata 1.64.0. *Zenodo*. doi:10.5281/zenodo.1475410

Ernest, S. K. M., Yenni, G. M., Allington, G., Bledsoe, E., Christensen, E., Diaz, R., Goheen, J., et al. (2018b). The portal project: A long-term study of a chihuahuan desert ecosystem. *bioRxiv*. doi:10.1101/332783

Heske, E. J., Brown, J. H., & Mistry, S. (1994). Long-term experimental study of a chihuahuan desert rodent community: 13 years of competition. *Ecology*, *75*, 438–445.

R Core Team. (2016). R: A language and environment for statistical computing.

White, E. P., Yenni, G. M., Taylor, S. D., Christensen, E. M., Bledsoe, E. K., Simonis, J. L., & Ernest, S. K. M. (2018). Developing an automated iterative near-term forecasting system for an ecological study. *Methods in Ecology and Evolution*.