Introducing bdclean: a user friendly biodiversity data cleaning pipeline

Tomer Gueta†, Vijay Barve§, Thiloshon Nagarajah¶, Ashwin Agrawal‖, Yohay Carmel‡

† Department of Civil and Environmental Engineering, The Technion – Israel Institute of Technology, Haifa, Israel
§ Florida Museum of Natural History, Gainesville, United States of America
¶ Informatics Institute of Technology, Colombo, Sri Lanka
‖ Indian Institute of Technology (IIT) - BHU, Varanasi, India

Corresponding author: Tomer Gueta (tomer.gu@gmail.com)

Received: 05 Apr 2018 | Published: 22 May 2018


Abstract

A new R package for biodiversity data cleaning, ‘bdclean’, was initiated in the Google Summer of Code (GSoC) 2017 and is available on github. Several R packages have great data validation and cleaning functions, but ‘bdclean’ provides features to manage a complete pipeline for biodiversity data cleaning; from data quality explorations, to cleaning procedures and reporting. Users are able go through the quality control process in a very structured, intuitive, and effective way. A modular approach to data cleaning functionality should make this package extensible for many biodiversity data cleaning needs. Under GSoC 2018, ‘bdclean’ will go through a comprehensive upgrade. New features will be highlighted in the demonstration.

Presenting author

Tomer Gueta
Grant title

ISF Grant No. 127/16

Google Summer of Code program