Scrubbing paleontological data

The winning entry of the Open Research Challenge (ORC) offers a solution for cleaning paleontological data.

The Open Research Challenge 2020, organised by Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), invited young researchers from around the world to submit innovative and creative ideas to remove incorrect and outdated entries in the largest database used in the field of paleontology. The winning submission, code developed by a paleontologist from Bristol in the UK, offers a solution for cleaning paleontological data.

How did organisms respond to climate change in the past? Why did some species survive mass extinctions? There are still many unanswered questions in paleontology that need big data to answer them. As paleontological knowledge grows, species names may change. Documenting these changes is a challenge. More often than not, new discoveries do not make it into the existing data repositories. Researchers spend a lot of their time making sure that their data is up to date before they can begin their work. Researchers at FAU, who use the database extensively, asked researchers to develop ideas to overcome this problem.

This year’s Open Research Challenge winner Joseph Flannery Sutherland from the School of Earth Sciences at the University of Bristol has developed code that will help paleontologists flag and correct inconsistencies in the data so that researchers can focus more on their actual research. Joe Flannery Sutherland has developed code that will automatically clean taxonomical errors in the Paleobiology Database (PBDB). The database, which is compiled by researchers from all around the world, is used extensively for quantitative analyses of diversification and extinction. It contains more than 1.2 million entries, many of which are erroneous or outdated. The code, developed in the statistics program R, will clean, and ideally replace, incorrect taxonomic and stratigraphic inconsistencies as well as temporal assignments of occurrence data.

Instead of having to fix every entry by hand, which is time-consuming, error prone, and requires expert knowledge, researchers can now rely on a straightforward and reproducible method. Wolfgang Kiessling, Professor of Paleobiology at FAU, believes the solution will result in higher quality publications and research. “The code developed by Joe Flannery Sutherland is a big step towards ridding the datasets of inconclusive and erroneous data. I am sure it will result in many highly interesting papers which, in turn, contribute to the overall quality of research in the field.”

wissenschaftliche Ansprechpartner:
More Information:
Blandina Mangelkramer
Stabstelle Presse und Kommunikation
09131/85-70210
blandina.mangelkramer@fau.de
Joe Flannery Sutherland is the winner of the Open Research Challenge 2020