Taxonomy has no shortcuts – why revisions are crucial for data analysis in Paleontology

MATHIAS HARZHAUSER[1] & BERNARD M. LANDAU[2]

[1]Natural History Museum Vienna, Burgring 7, 1010 Vienna, Austria

[2]Naturalis Biodiversity Center, P.O. Box 9517, 2300 RA Leiden, Netherlands; Instituto Dom Luiz da Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal; and International Health Centres, Av. Infante de Henrique 7, Areias São João, P-8200 Albufeira, Portugal

Great efforts have been devoted during the last decade to digitalize paleontological data. Flagship of these projects are the Paleobiology Database (https://paleobiodb.org/#/) and Fossilworks (http://fossilworks.org). These projects utilize literature data within a standardized taxonomic and stratigraphic frame and aim at detecting changes in diversity patterns on various temporal and spatial scales. Whilst this approach clearly has great potential, it carries the danger of producing "pseudo-revisions" by automated new combinations of higher systematic ranks. Moreover, we doubt that this approach can be used seriously for species-level analysis without careful and time-consuming revision of the input-data. An impressive example are our revisions of the Miocene gastropod families Cancellariidae, Columbellidae, Conidae, Costellariidae, Mitridae and Turritellidae from the Parathethys Sea. These revisions revealed major problems in previous generic placements resulting in up to 80% of new combinations. In some families, species level identifications of putatively wide-spread species, seemingly known also from the Mediterranean Sea and the northeastern Atlantic, turned out to be incorrect. Prior to these Parathethyan revisions, relationships with the northeastern Atlantic would have been considered high. However, following these revisions, such relationships were found to be almost non-existant. Similarly, the relationship with stratigraphically younger Pliocene faunas decreased drastically after revision. Furthermore, the literature-data would have completely failed to detect the faunistic relation between Parathethyan faunas and those of the tropical eastern Atlantic. Finally, a survey of species identifications in Parathethyan literature documented that up to 80% of the species identifications have been wrong in families such as the Conidae and Turritellidae. Thus, putting these data uncritically into databases will produce more noise than signal.