Mikrotax: Developing a Genuinely Effective Platform for Palaeontological Geoinformatics

Jeremy R. YOUNG[1, *], Paul R. BOWN[1], Bridget S. WADE[1], Brian E. PEDDER[1], Brian T. HUBER[2] and David B. LAZARUS[3]

[1] *Department of Earth Sciences, University College London, London UK*
[2] *Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, Washington D.C., USA*
[3] *Museum fürNaturkunde, Invalidenstrasse 43, 10115 Berlin, Germany*

Citation: Young et al., 2019. Mikrotax: Developing a Genuinely Effective Platform for Palaeontological Geoinformatics. *Acta Geologica Sinica* (English Edition), 93(supp. 1): 70-72

Planktonic microfossils have played an enormously important role in modern earth science, notably including industrial hydrocarbon exploration and academic Deep Sea Drilling, providing essential data for age models and rich suites of palaeoceanographic data. Conversely this broad application has resulted in the development of detailed knowledge of planktonic microfossils and in stimulating study of them. Over 50 years of drilling a vast archive of data has accumulated but it has been difficult to access not least due to the complexities and inconsistencies of naming of species. There is also growing concern about sustaining the expertise necessary for micropalaeontology as a generation of expert scientists nears retirement. New generations of talented micropalaeontologists are being trained, but they need to have multidisciplinary skills and rarely have the time to develop the depth of taxonomic expertise of the founders of the field. Moreover, new centres of microfossil study in countries such as Vietnam or Tanzania often have limited access to specialist research libraries. These factors contributed to make online documentation of taxonomy an obvious priority for the field.

With funding-support from the UK Natural Environment Research Council (NERC), we have now produced large-scale web-syntheses of taxonomic data on both calcareous nannofossils (Nannotax) and planktonic foraminifera (pforams@mikrotax), and companion sites for acritarchs and radiolarians are being developed. These websites provide authoritative identification guides and data on taxonomic ranges accessible to scientists across the world (Huber et al., 2017). They currently cover approximately 5000 described taxa and are illustrated by ca 30,000 images. They are accessed daily by hundreds of users both in industry and academia and are intensively used by scientists on IODP legs. In addition to acting as taxonomic reference sites we are developing them as data portals for distribution data and specifically for displaying data from the Neptune database assembled by David Lazarus and co-workers. Charts of occurrence frequency through time are provided on species pages using this data, and additional tools allow plotting of comparative range charts, distribution charts for individual sites, distribution maps, and age-latitude plots.

The success of the sites can be demonstrated by the high level of usage as documented by Google Analytics (Fig. 1). For the year Jan 2018 to Jan 2019 the records indicate that we had: 24,161 total users; 104,672 sessions; an average session duration of 12:23 mins; and 1,008,234 page views. The geographic spread of users is global and also reflects the distribution of known centres of microfossil research

The recorded level of usage is mirrored by personal feedback from our colleagues, by anecdotal evidence and by responses to a user survey (http://www.mikrotax.org/Nannotax3/pages/ntax-survey.php).Many users have reported that the sites are essential tools for their day-to-day work or that they played a key role in their learning of taxonomy of the groups.

We believe that fundamental to the success of the project has been that the system was designed by working micropalaeontologists in order to solve the immediate needs of micropalaeontologists and that it has evolved in response to feedback from end users. In addition, we have recognised that the system will be used by a wide range of users ad designed to be accessible tothem - in particular we have in mind (1) non-experts who simply need to be able to look up data on a individual species; (2) students learning taxonomy; (3) expert specialists needing a tool to explore the accumulated knowledge on these fossils. In practical terms some of the more important ways we have succeeded in meeting these objectives have been:

Image-based, browseable interface - the system is organised following the classical Linnean taxonomy with pages for each taxon but it also provides on every higher taxonomic page an illustrated "taxon-table" summarizing the taxa within it (Fig. 2). This allows users to browse through the taxonomy, and guide their knowledge of the group. This is far more effective and user friendly than the common alternative of simply producing cryptic lists of names, or direct database search. Likewise, on species pages image display is prioritised and for many taxa there are large collections of images - this reflects the prime requirement of micropalaeontologists for images.

Proper use of geological time - geological time handling is built into the system. Age data is encoded in terms of age units and position within them then dynamically reinterpreted as numerical Ma ages using look-up tables. This system provides long term stability since only the look-up tables need to be revised as new time scale calibrations are produced (Baker et la., 2011). This data is used to display age ranges, to allow search by age and to allow selective display of taxa within a time interval of interest ("Time Control").

* Corresponding author. E-mail: jeremy.young@ucl.ac.uk

http://www.geojournals.cn/dzxbcn/ch/index.aspx
https://onlinelibrary.wiley.com/journal/17556724

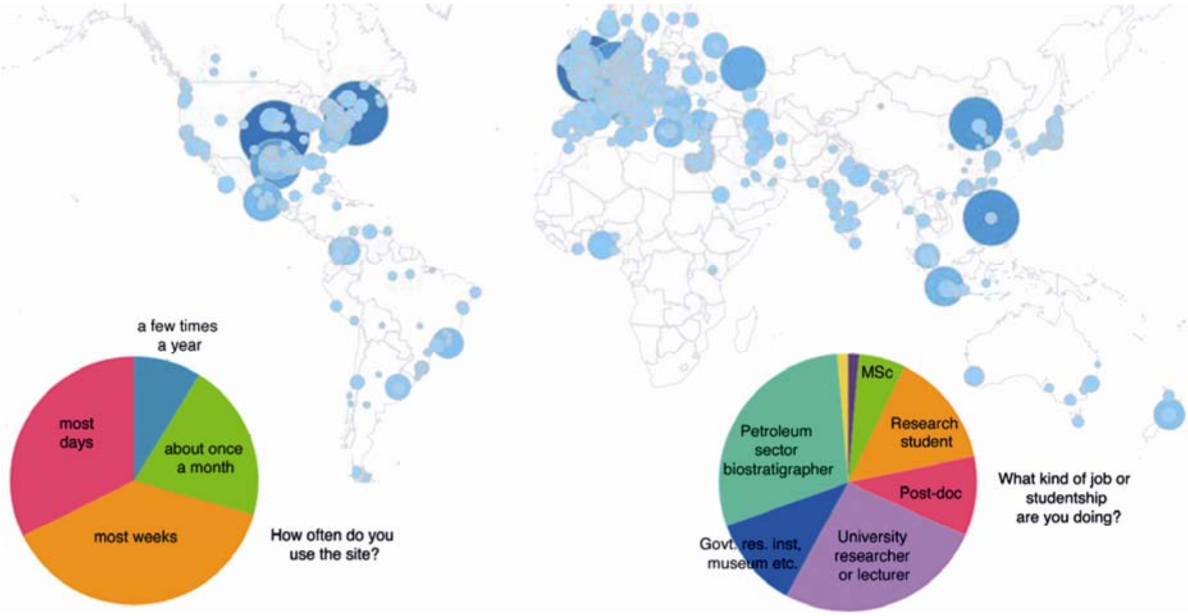

Fig. 1. Site use.
Map - geographic spread of users (during April 2018). Pie Charts - responses of users to survey questions.

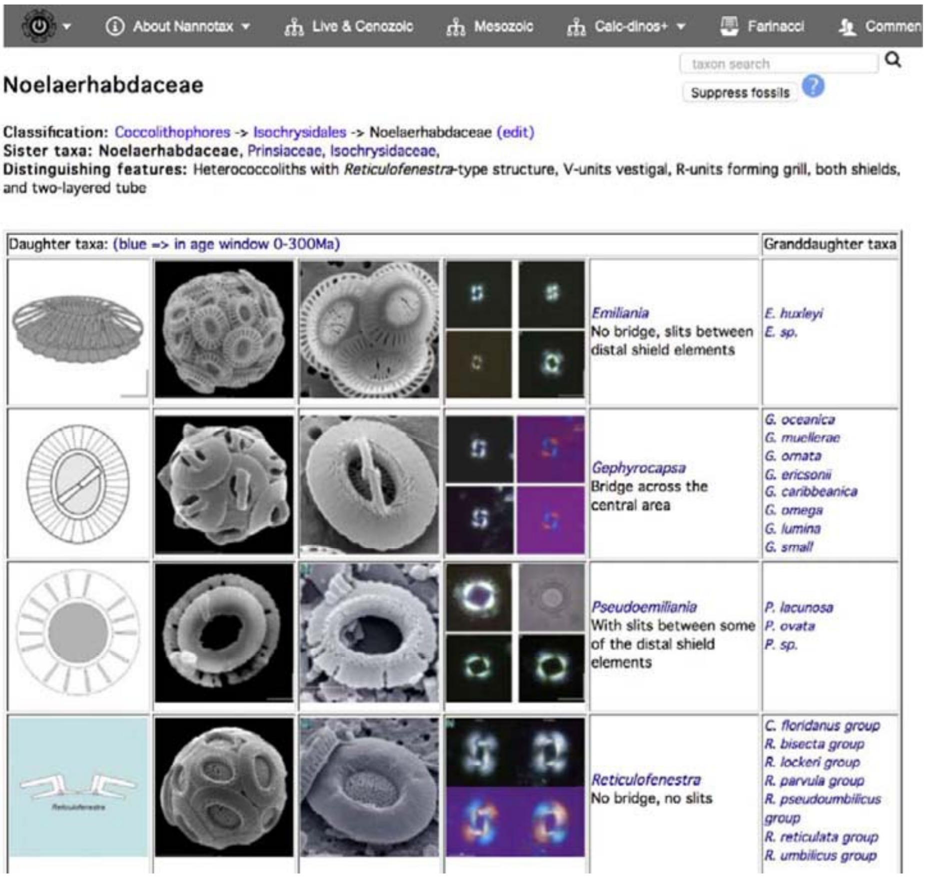

Fig. 2. Taxon table.
These taxon tables are autogenerated on every higher taxon page and provide concise illustrated summaries of the classification and so a visual guide to learning the taxonomy.

Parallel databases for all described taxa and for working taxonomy. Microfossil taxonomy is still in a state of continuous evolution, refinement and revision. Consequently, there is a need both to document the full range of taxa described and to synthesis the working taxonomy in use. Separating these two aspects has produced a powerful data structure which suits the needs of a wide range of users and clearly separates objective and subjective aspects of our synthesis.

Effective search. Rapid access to data is key for a database. We have two search systems. Simple search allows search for a taxon name and is optimized by a number of features -the searchbox appears at the top right of every page, if there is only one search hit the user is taken straight to the relevant page, only a fragment of a name is needed for search, alternate Latin endings are automatically searched in parallel (e.g. *pulchra*, *pulchrum* and *pulcher*), etc. Advanced search by contrast has multiple options in terms of both, search constraints (age, taxonomic, group, author, description date, geological age, morphological feature, and search output (citation, images, morphology summary, character matrix).

Integration with occurrence data - with the assistance of collaborator David Lazarus (U. Berlin) we have integrated querying of the Neptune database of occurrence data (Lazarus, 1994; Spencer-Cervato, 1999) from scientific ocean drilling (DSDP, ODP, IODP) into the mikrotax system. This provides quantitative data on the occurrence frequency of taxa through time (Fig. 3). The data is provided as summary histograms on every species page and can be further queried through a range of custom plotting tools. This both provides novice users with highly graphical summaries of the abundance and distribution of different taxa and provides expert users with a rich source of research data.

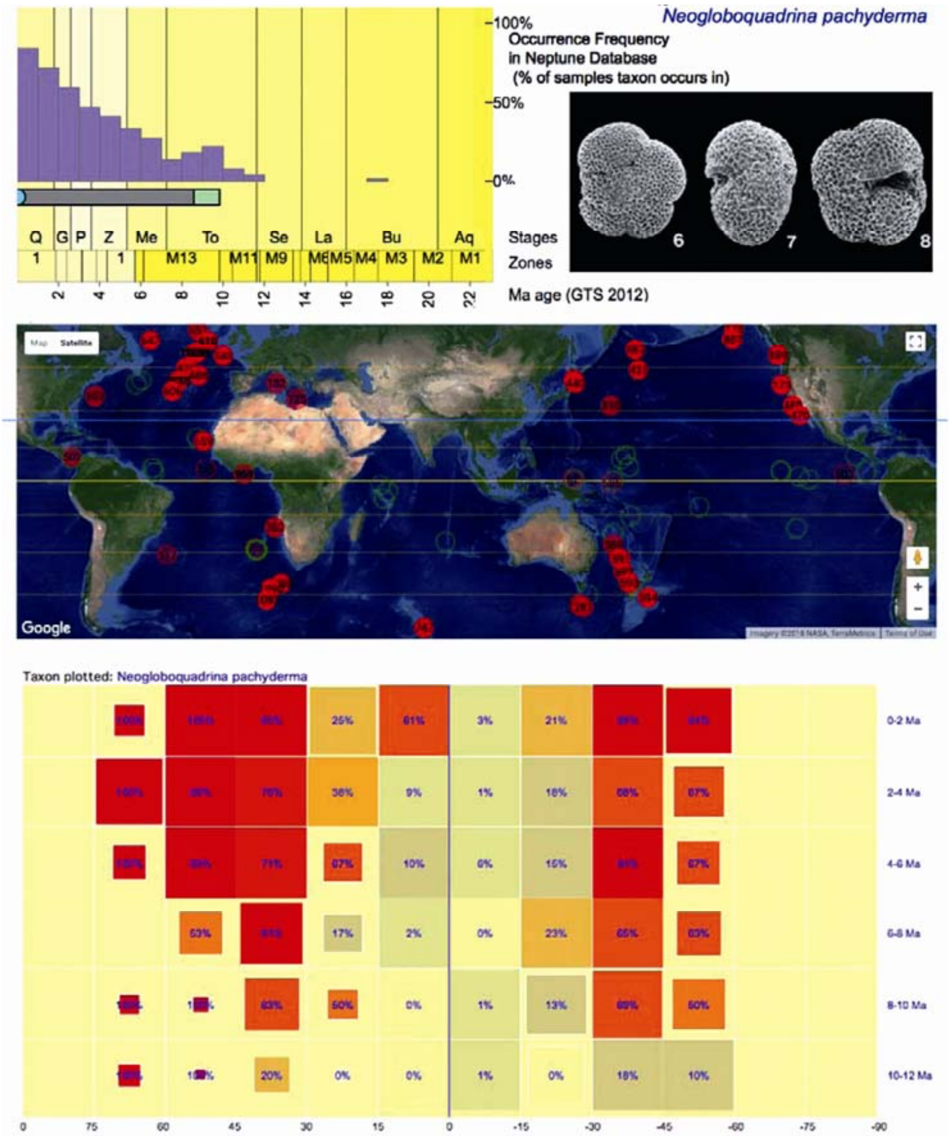

Fig. 3. Display of occurrence data from Neptune database for a single species *Neogloboquadrinapachyderma*.
Top Left standard plot at base of taxon page. Bottom left map of recorded occurrences for a time slice, top right summary plot of occurrence frequency by geological age and paleolatitude.

Key words: web-taxonomy, micripaolaeontology, nannofossills, cocolitophores

Acknowledgments: This work has been supported grants from the UK Natural Environment Research Council (Grant Nos. NE/N017900/1 &NE/J006416/1)and from the State Key Laboratory of Paleobiology and Stratigraphy (Nanjing).

References

Baker, E., Johnson, K.G., and Young, J.R., 2011. The future of the past in the present: biodiversity informatics and geological time. *ZooKeys*, 150: 397–405.

Huber, B.T., Petrizzo, M.R., Young, J.R., Falzoni, F., Gilardoni, S.E., Bown, P.R., and Wade, B.S., 2017. Pforams@microtax: A new online taxonomic database for planktonic foraminifera. *Micropaleontology*, 62(6): 429–438.

Lazarus, D.B., 1994. Neptune: A marine Micropaleontology Database. *Mathematical Geology*, 26(7): 817–832.

Spencer-Cervato, C., 1999. The Cenozoic Deep Sea Microfossil Record: Explorations of the DSDP/ODP Sample Set Using the Neptune Database. *Palaeontologia Electronica*, 2(2): 1–268.

About the first author (also corresponding author)

Jeremy R. YOUNG, male, born in 1960 in London, England; PhD, from Imperial College London, MSc from University College London, MA from Oxford University. Research Associate at the Department of Earth Sciences, University College London. He is a micropalaeontologist with research expertise in fossil and living coccolithophores. He has developed the *mikrotax* content management system and lead creation of the Nannotax and pforams@mikrotax websites. Email: jeremy.young@ucl.ac.uk; phone: +44 7813 114208.