Tom Webb, Animal and Plant Sciences

Climate change and marine ecology

Tom Webb of the Department of Animal and Plant Sciences has benefited from open research practices over the course of his career. In his project on the effect of increased environmental temperatures on marine species, Tom looks at how he is improving his own research practices.

As a macroecologist and biodiversity scientist, I am dependent on other people's data in my quest to better understand how marine life is changing in the anthropocene. I have experienced the frustration of key datasets being unavailable or untidy, of broken links and code lacking documentation. I also understand the need to incentivise data providers, and follow best practice in acknowledging their efforts through citation and collaboration. As a data user, I also feel responsible for the ongoing process of improving my own research practices.

Applying open research in marine ecology

My progress towards open research is exemplified by my recent project on the thermal limits of marine species. It's a study built on open data, developed using open-source software, with results published in an open access journal. It also encapsulates my belief that we must make the best use of existing data in our efforts to address the biodiversity and climate crises: collating, linking, remixing and enriching openly available data to ask and - eventually - answer novel questions.

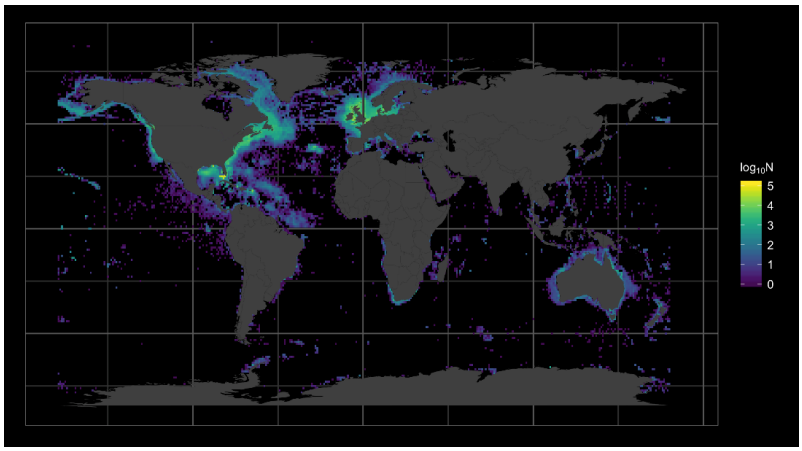
I started with the aim of quantifying the thermal tolerances of marine species. Sometimes these limits have been determined experimentally, but the logistics, expense and ethics of this approach mean only a few hundred marine species will ever be assessed in this way. However, we do have access to over 78 million occurrence records for more than 150,000 species through the Ocean Biodiversity Information System (OBIS), as well as large open datasets of sea temperature. We wanted to find out if these existing datasets could be linked together to summarise the temperatures in which different marine species have been recorded, so we matched large samples of this data with openly available sea temperature datasets. Our estimates of thermal tolerance were shown to match experimental results well, demonstrating that it is possible to use open data to obtain accurate assessments of thermal tolerances, a vital indicator for predicting changes in distribution under climate change.

"Adopting open research practices enables us to make the best possible use of existing data in our efforts to address the biodiversity and climate crises: collating, linking, remixing and enriching openly available data to ask and - eventually - answer novel questions."

As well as using openly available data, thereby reducing the need to expend resources on collecting new data, we wanted to make our own research as widely accessible as possible. A good illustration of this is the Data Availability Statement in our article published in *Ecology and Evolution* (2020):

'A major aim of this work is to make the tools required to replicate, adapt, and extend the methods presented freely available to the community. Our work uses existing publicly available data, and we show users how to access the same data from within the open source statistical environment R. Processed datasets and code for analysis and visualization are available via GitHub and are also deposited in Figshare [1] via the University of Sheffield's Online Research Data repository. [2]'

This model of open data has become my preferred workflow, with data processed and analysed using open-source tools, archived in ORDA, and documented with an extensive readme serving as both documentation and tutorial. (There are more examples of our open data in GitHub and ORDA [3].)



The global distribution of the >2M openly available occurrence records that we matched to global sea temperature datasets to calculate observed thermal limits for >500 marine species. The legend shows the total number of records mapped on a 1 degree grid.

Looking to the future

As part of our research, we created a data product [4] for the European Marine Observation and Data Network (EMODnet), showing how to derive, summarise and visualise thermal affinities for European marine species. Building on this, I am delighted to be leading the data products development team for Phase IV of EMODnet Biology (2021-23), a fantastic opportunity to work with scientists, data professionals, and research software engineers to make useful, accessible products to help shape international marine policy.

As someone of moderate technical ability, I have found the process of improving my research practice challenging, and it has taken me too long to adopt some aspects of good practice. But I am now embedding open research into the culture of my research group, and I am proud that the scientists I am training - including my early career co-authors [2] - are taking these principles with them wherever they go next.

Our open research

- Existing open biodiversity & climate data used extensively; newly generated data & code published
- Open data placed in repositories with documentation to maximise accessibility and usability
- Open research incorporated into research group training to help spread best practices

References

[1] Webb, T. (2020). Data and code for Occupancy-derived thermal affinities reflect known physiological thermal limits of marine species. [Dataset] University of Sheffield, Figshare. https://doi.org/10.15131/shef.data.12249686.v1
[2] Webb, T. et al. (2020). Occupancy-derived thermal affinities reflect known physiological thermal limits of marine species. *Ecology and Evolution* 10: 7050– 7061. https://doi.org/10.1002/ece3.6407
[3] Webb, T. (2020). Linking dimensions of data on global marine animal diversity. [Software] University of Sheffield, Figshare. https://doi.org/10.15131/shef.data.12833891.v1
[4] Webb, T. and Lines, A., (2018). Thermal affinities for European marine species. [Dataset] *Marine Data Archive*. https://doi.org/10.14284/378