Plan Overview

A Data Management Plan created using DMPonline

Title: MindTheGap: Quantifying the completeness of the stratigraphic record and its role in reconstructing the tempo and mode of evolution

Creator: Emilia Jarochowska

Principal Investigator: Emilia Jarochowska

Data Manager: Emilia Jarochowska, Niklas Hohmann, Aristotelis Kandylas, Xianyi Liu

Project Administrator: Emilia Jarochowska

Affiliation: Utrecht University

Funder: European Research Council (ERC)

Template: ERC DMP

ORCID iD: 0000-0001-8937-9405

Project abstract:

Darwin identified the incompleteness of the geological record as a major concern in our ability to reconstruct the evolutionary history from fossils. Geological strata are not, however, a random selection of the past, but are deposited and preserved in processes which can be modelled mathematically. This project employs numerical forward modelling, a technique from sedimentary basin analysis, to quantify how stratigraphic gaps affect evolutionary patterns in the fossil record. It focuses on tropical carbonate platforms, because they are the richest archive of biodiversity through the Phanerozoic and are directly formed by organisms, thus preserving a positive feedback loop between the environment, evolution, and sediment formation. Completeness of the record depends on the time scale of measurement. Previous work demonstrated that long gaps in the record, in the range of 106-107 years, are reflected in biodiversity reconstructions. Here we address shorter gaps (103-105 years), which are more frequent, but harder to detect. They can be predicted using models of stochastic sedimentation, astronomical forcing, and sedimentary and diagenetic self-organization. Forward modelling will be used to simulate the effect of gaps in the record produced by each of these processes on diversity and character evolution. This will allow us to answer the questions: What part of diversity do we miss owing to the missing rock record? Are the preserved intervals exceptional or representative? Is there a systematic part of biodiversity and environmental records which will always fall into gaps? By applying the findings of the forward modelling to fossil successions, we will be able to identify and correct for the missing parts of the record and formulate testable hypotheses on the original tempo and mode of evolution at the highest temporal resolution achievable in the geological record.

ID: 128211

Start date: 01-02-2023

End date: 31-01-2028

Last modified: 02-07-2024

Grant number / URL: 101041077

Copyright information:

MindTheGap: Quantifying the completeness of the stratigraphic record and its role in reconstructing the tempo and mode of evolution

Summary

Project Acronym

MindTheGap

Project Number

101041077

Provide a dataset summary

Outputs from computer simulations, including parameters used to obtain them

Arrays of numbers obtained from model runs. Two models will be producing these outputs: a forward model of carbonate platforms and a diagenetic models simulating the composition and physical characteristics of carbonate sediment. The outputs will include numbers characterising the concentrations of ions in sediment, porosity of the sediment, type of sediment deposited in a model grid.

Empirical data

Photos of fossils and measurements extracted from these photos

Digitized photographs and processed spatial datasets of carbonate platforms obtained from project collaborators

Numerical data containing additional variables describing the photos and spatial datasets (such as geographic position, stratigraphic position, sample numbers).

FAIR data and resources

1. Making data findable

Datasets are stored in Open Science Framework (osf.io) using its metadata system and DOI. In parallel, it is automatically backed up to Yoda. Yoda provides a shared and secured data-storage. It allows to store data for a period of at least 10 years in a frozen state, together with a standardized set of metadata, and to publish the dataset with a DOI, making the dataset findable in the Yoda Catalogue via its metadata.

The code used to generate the datasets obtained from simulation allows to reproduce it fully, but datasets are stored to reduce the cost of computation. Each model run is stored together with the full set of parameters and parameters are described in software documentation (GitHub and Zenodo).

Photographs and associated numerical datasets are stored together with descriptions of how they

were obtained and the code used to process them allows to reproduce all analyses carried out on them (GitHub and Zenodo).

2. Making data openly accessible

All data will be shared in public repositories (Open Science Framework and Yoda) without restrictions. For simulation data, code required to reproduce the data is shared as well, together with tools allowing to obtain the same results on different architectures (using virtual environments).

3. Making data interoperable

The datasets created in simulations will be stored in the HDF5 format (https://portal.hdfgroup.org) which allows storing large and complex structured datasets and can be read using libraries for all major languages (Python, R etc). How the datasets were produced will be documented in associated GitHub repositories and the parameters used to obtain each run will be stored in the same file along with the results.

Photographs are stored as either .JPG or .TIFF and the associated numerical data as .csv files. Whenever these files had been used for analyses, code used to open and manipulate them (in Python, R, Julia and, occasionally, Matlab) will be provided.

4. Increase data reuse

Datasets will be described in scientific publications and made available in OSF and Yoda under a CC-By Attribution 4.0 International license.

The copyright owners are jointly Utrecht University and the Netherlands eScience Center.

Datasets will be shared upon the acceptance of the first publication based on them. There will be no restrictions on data sharing. Each dataset will have a DOI to allow tracking reuse and which will be used for citing by ourselves.

5. Allocation of resources and data security

No personal information will be collected nor stored. All output will be digital.

Data curation is carried out with the support of the Data Steward at Utrecht University.