

Extending Ocean Drilling Pursuits [eODP]: Making scientific ocean drilling data accessible through searchable databases

Leah LeVay¹, Andrew Fraass², Jocelyn Sessa³, and Shanan Peters⁴

¹Texas A&M University

²University of Bristol

³Drexel University

⁴University of Wisconsin, Madison

November 23, 2022

Abstract

Scientific ocean drilling through the International Ocean Discovery Program (IODP) and its predecessors, has a far-reaching legacy. They have produced vast quantities of marine data, the results of which have revolutionized many geoscience sub-disciplines. Meta-analytical studies from these efforts exist for micropaleontology, paleoclimate, and marine sedimentation, and several outstanding resources have curated and made available elements of offshore drilling data, but much of the data remain heterogeneous and dispersed. Each study, therefore, requires reassembling a synthesis of data from numerous sources; a slow, difficult process that limits reproducibility and slows the progress of hypothesis testing and generation. A computer programmatically-accessible repository of scientific ocean drilling data which spans the globe will allow for large-scale marine sedimentary geology and micropaleontologic studies and may help stimulate major advances in these fields. The eODP project, funded through the NSF's EarthCube program, seeks to facilitate access to and visualization of these large microfossil and stratigraphic datasets. To achieve these goals, eODP will be linking and enhancing three existing database structures: Open Core Data (OCD), the Paleobiology Database (PBDB), and Macrostrat. Over the next three years, eODP will be accomplishing the following goals: (1) enable construction of sediment-grounded and flexible age models in an environment that encompasses the deep-sea and continental records; (2) expand existing lithology and age model construction approaches in this integrated offshore-onshore stratigraphically-focused environment; (3) adapt key microfossil data into the PBDB data model from OCD; (4) develop new API-driven web user interfaces for easily discovering and acquiring data; and (5) establish user working groups for community input and feedback. This project is targeting shipboard drilling-derived data, but the infrastructure will be put in place to allow the addition of other shore-based information. The success of eODP hinges upon interaction, feedback, and contribution of the scientific ocean drilling community, and we invite anyone interested in participating in this project to join the eODP team.

Extending Ocean Drilling Pursuits [eODP]: Making Scientific Ocean Drilling Data Accessible Through Searchable Databases

Extending Ocean Drilling Pursuits [eODP]: Making Scientific Ocean Drilling Data Accessible Through Searchable Databases

Andrew J. Fraass (1,2) Leah J. LeVay (3) Jocelyn A. Sessa (2) Shanan E. Peters (4)

(1) School of Earth Sciences, University of Bristol
(2) Academy of Natural Sciences of Drexel University
(3) International Ocean Discovery Program, Texas A&M University
(4) Department of Geoscience, University of Wisconsin-Madison

eODP Overview and Outcomes

Scientific ocean drilling through the International Ocean Discovery Program (IODP) and its predecessors has a far-reaching legacy. They have produced vast quantities of marine data, the results of which have revolutionized many geoscience disciplines. Meta-analytical studies from these efforts exist for micropaleontology, paleoclimate, and marine sedimentation and several outstanding resources have curated and made available elements of offshore drilling data. However, much of the data remain heterogeneous and dispersed. Each study, therefore, requires reassembling a synthesis of data from numerous sources – a slow, difficult process that limits reproducibility and slows the progress of *biostratigraphic correlation and stratigraphic correlation*.

OPEN

Paleontology

Paleobiology Database

Created in 1998, the PBDB is an open resource of paleontological data on all fossil organisms through all time periods. Bibliographic references, taxonomic names, synonyms and classifications, primary collection data, taxonomic occurrences, re-identifications of occurrences, and geologic time scales are stored in a relational database. The PBDB contains over 67,000 references, 380,000 taxa and 1,395,000 occurrences in 199,000 collections. The PBDB has active partnerships with other international open data cyberinfrastructure projects, like the Global Biodiversity Information Facility and with DigBio through the ePANDORA project (epandora.org). Additionally, anyone can access the entire database via an API.

Ocean drilling microfossil occurrences and taxonomy will be housed in the PBDB, which can support quantitative and qualitative abundance data and multiple taxonomic concepts. The PBDB provides visualization tools that can be used for biogeographic reconstructions, global occurrences, or diversity. Data can be searched for by taxa name, taxa type, age, and region. Through building a robust, distributed set of microfossil occurrences, eODP will allow micropaleontologists to answer questions about regional or global diversity, group evolution, ecological responses to global-scale events, and long-term interactions between trophic levels. A new user interface will allow micropaleontologists to easily enter their own datasets.

OPEN

Stratigraphy

Macrostrat

Macrostrat, started in 2006, has assembled chronostratigraphic rock units and basic properties in order to enable quantitative analyses of large-scale sedimentation patterns around the globe. The database currently contains 1,474 regional rock columns, 33,903 rock units from North America, the Caribbean, New Zealand, and the marine realm and is integrated with more than 1.89 million geologic map polygons from over 100 separate geological maps sources spanning the globe at multiple different spatial resolutions. Macrostrat will house ocean drilling lithologic data along with site age models. These data will be searchable by lithology type, age, and region allowing scientists to study marine sedimentation trends through time. The program will host alternate depth scales to accommodate spliced sequences. Macrostrat will allow for multiple age models to be stored for each site so that researchers can choose the age model they prefer or input their own. Through an API, the age models in Macrostrat will be linked to the ocean drilling taxonomic data in the PBDB.

OPEN

Project Status and Outreach

Current Project Status

- Held the first PI-developer meeting in January of 2020
- Creating external database that compiles all of the microfossil and lithology data in a uniform structure that can be easily imported to Macrostrat and the PBDB. This database currently houses IODP data from the LIMS database. This will be extended to older datasets and the Neptune Sandbox.
- Entering microfossil taxonomy and biostratigraphic zonations into Macrostrat and the PBDB to prepare for ingestion of LIMS data.
- Begin planning for a Stratigraphy-focused workshop that will take place in the next year. Challenges associated with planning during COVID.

OPEN

eODP Ecosystem

Scientific ocean drilling data from over 3700 drill holes (see locality map) are stored in three separate locations: IODP LIMS database, Open Core Data, and the IODP-NCEI database. Each of these platforms stores lithology, stratigraphy, and micropaleontology data as flat files with limited search capabilities.

OPEN

Personnel and Acknowledgements

Associated Personnel

Brian Huber - Smithsonian National Museum of Natural History
 Doug Fife - Open Core Data, Consortium for Ocean Leadership
 Daven Quinn - Macrostrat, U. of Wisconsin
 Michael McClellenn - PBDB, U. of Wisconsin
 Seth Kaufman and Wai-Yin Kwan - PBDB, Whirl-i-Gig
 Shane Loeffler and Reed McEwan - Flyover Country, U. of Minnesota

OPEN

Andrew J. Fraass (1,2) Leah J. LeVay (3) Jocelyn A. Sessa (2) Shanan E. Peters (4)

(1) School of Earth Sciences, University of Bristol (2) Academy of Natural Sciences of Drexel University (3) International Ocean Discovery Program, Texas A&M University, (4) Department of Geoscience, University of Wisconsin-Madison





EODP OVERVIEW AND OUTCOMES

University of BRISTOL

Scientific ocean drilling through the International Ocean Discovery Program (IODP) and its predecessors has a far-reaching legacy. They have produced vast amounts of marine data, the results of which have revolutionized many geoscience subdisciplines. Meta-analytical studies from these efforts exist for micropaleontology, paleoclimate, and marine sedimentation, and several outstanding resources have curated and made available elements of offshore drilling data; however,

PRESENTED AT:

much of
the data



EarthCube

2020 EarthCube Annual Meeting
Virtual – June 18, 2020

heterogeneous and dispersed. Each study, therefore, requires reassembling a synthesis of data from numerous sources - a slow, difficult process that limits reproducibility and slows the progress of hypothesis testing and generation. A programmatically-accessible repository of scientific ocean drilling data that spans the globe will allow for large-scale marine sedimentary geology and micropaleontologic studies and may help stimulate major advances in these fields.

The eODP project, funded through the NSF's EarthCube program, seeks to facilitate access to and visualization of these large microfossil and stratigraphic datasets. To achieve these goals, eODP will be linking and enhancing three existing database structures: Open Core Data (OCD), the Paleobiology Database (PBDB), and Macrostrat.

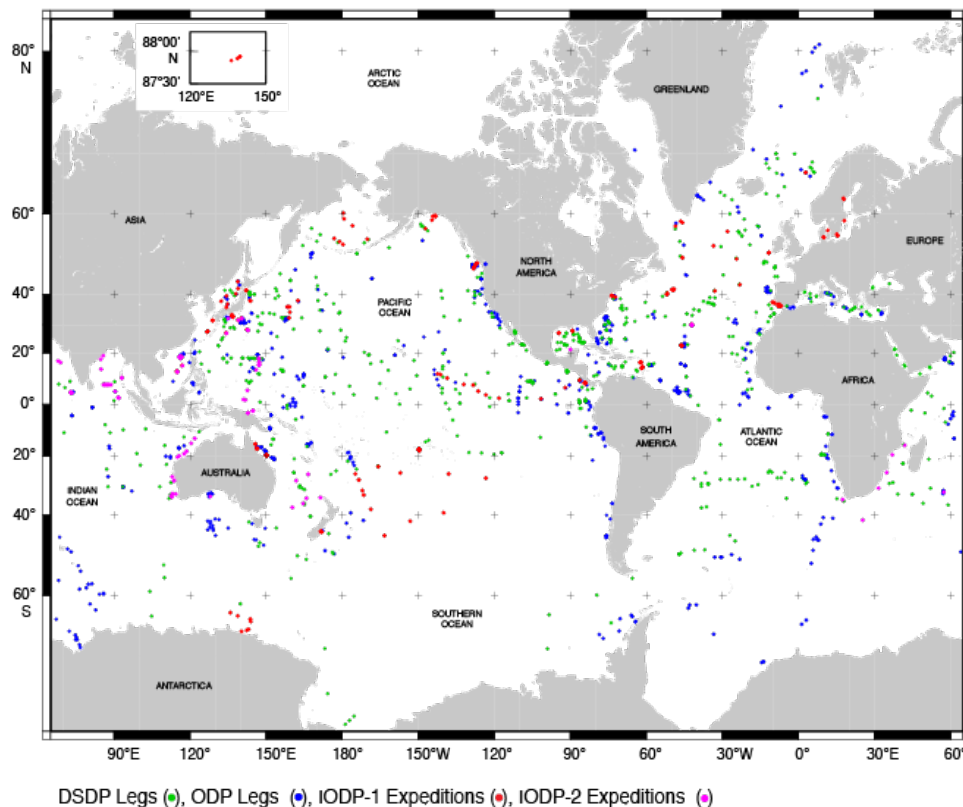
eODP will accomplish the following goals:

- (1) Enable construction of sediment-grounded and flexible age models in an environment that encompasses the deep-sea and continental records;
- (2) Expand existing lithology and age model construction approaches in this integrated offshore-onshore stratigraphically-focused environment;
- (3) Adapt key microfossil data into the PBDB data model from OCD, Neptune and other sources;
- (4) Develop new API-driven web user interfaces for easily discovering and acquiring data; and
- (5) Establish user working groups for community input and feedback.



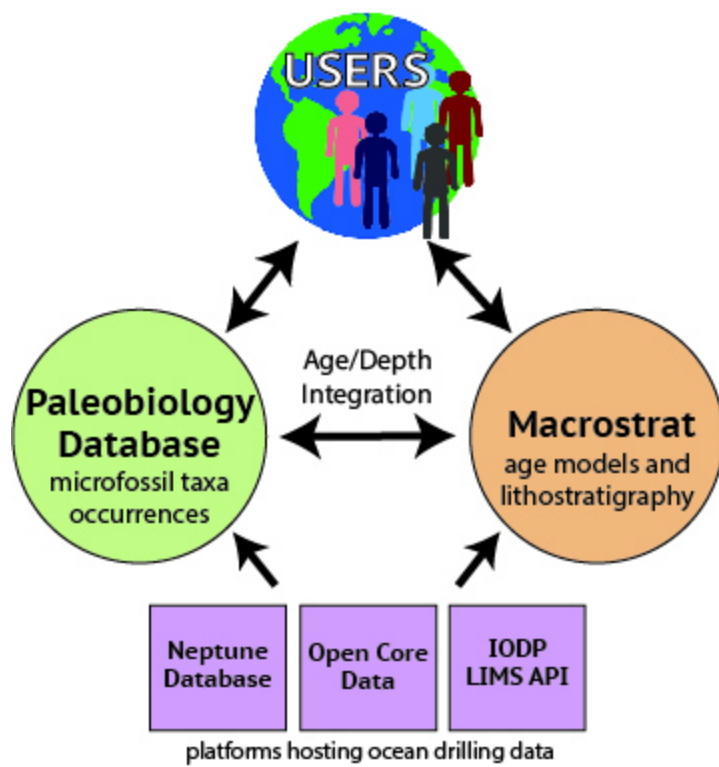
EODP ECOSYSTEM

Scientific ocean drilling data from over 3700 drill holes (see locality map) are stored in three separate locations: IODP LIMS database (<http://web.iodp.tamu.edu/LORE/>), Open Core Data (<http://opencoredata.org/>), and the NOAA-NCEI database (<https://www.ngdc.noaa.gov/mgg/geology/dsdp/dsdpcdv2.htm>). Each of these platforms stores lithology, stratigraphy, and micropaleontology data as flat files with limited search capabilities.



Much of the shipboard microfossil data is stored in the Neptune Sandbox Berlin (<http://www.nsb-mfn-berlin.de/index>)(NSB), the latest iteration of the Neptune Database. These data have been curated by experts in the field. This curation includes a standardized taxonomy and age model interpretations. NSB will be collaborating with eODP to bring their collections into the eODP Ecosystem.

Lithology and age data will be migrated to Macrostrat and microfossil data will be sent to the Paleobiology Database. These two platforms will transfer depth-age information via API.



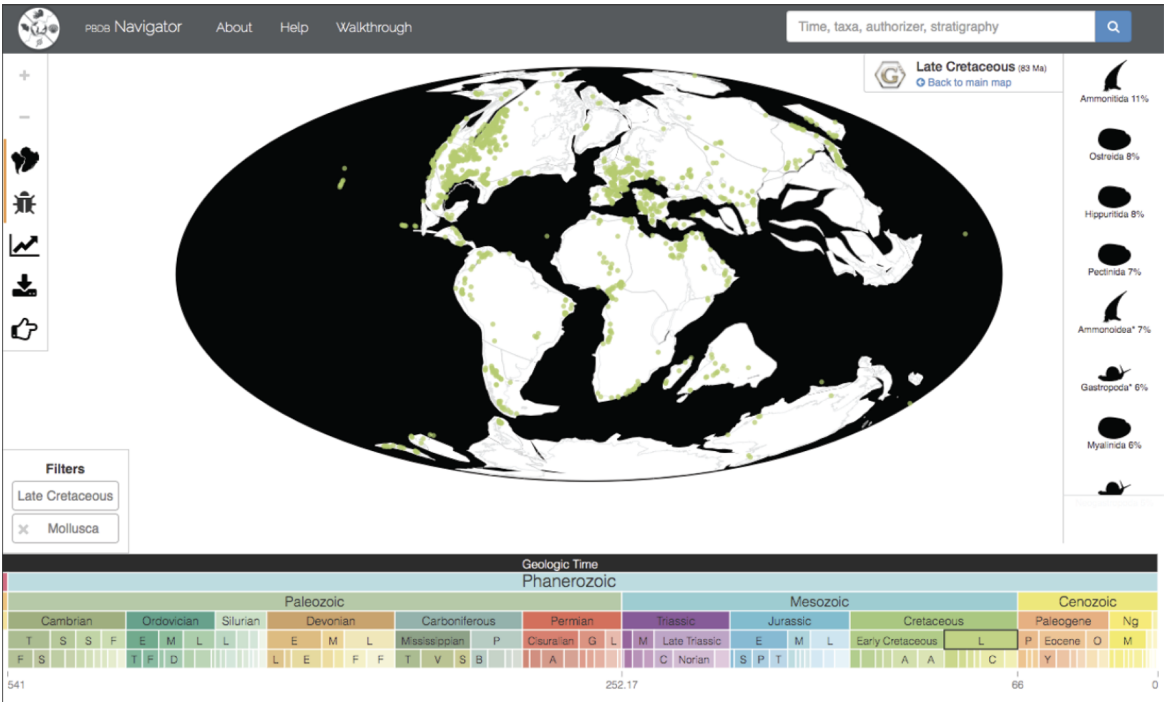
PALEONTOLOGY

Paleobiology Database



Created in 1998, the PBDB (<http://www.paleobiodb.org>) is an open resource of paleontological data on all fossil organisms through all time periods. Bibliographic references, taxonomic names, synonymies and classifications, primary collection data, taxonomic occurrences, re-identifications of occurrences, and geologic time scales are stored in a relational database. The PBDB contains over 67,000 references, 380,000 taxa and 1,395,000 occurrences in 199,000 collections. The PBDB has active partnerships with other international open data cyberinfrastructure projects, like the Global Biodiversity Information Facility and with iDigBio through the ePANDDA project (epandda.org). Additionally, anyone can access the entire database via an API.

Ocean drilling microfossil occurrences and taxonomy will be housed in the PBDB, which can support quantitative and qualitative abundance data and multiple taxonomic concepts. The PBDB provides visualization tools that can be used for biogeographic reconstructions, global occurrences, or diversity. Data can be searched for by taxa name, taxa type, age, and region. Through building a robust, distributed set of microfossil occurrences, eODP will allow micropaleontologists to answer questions about regional or global diversity, group evolution, ecological responses to global-scale events, and long-term interactions between trophic levels. A new user interface will allow micropaleontologists to easily enter their own datasets.



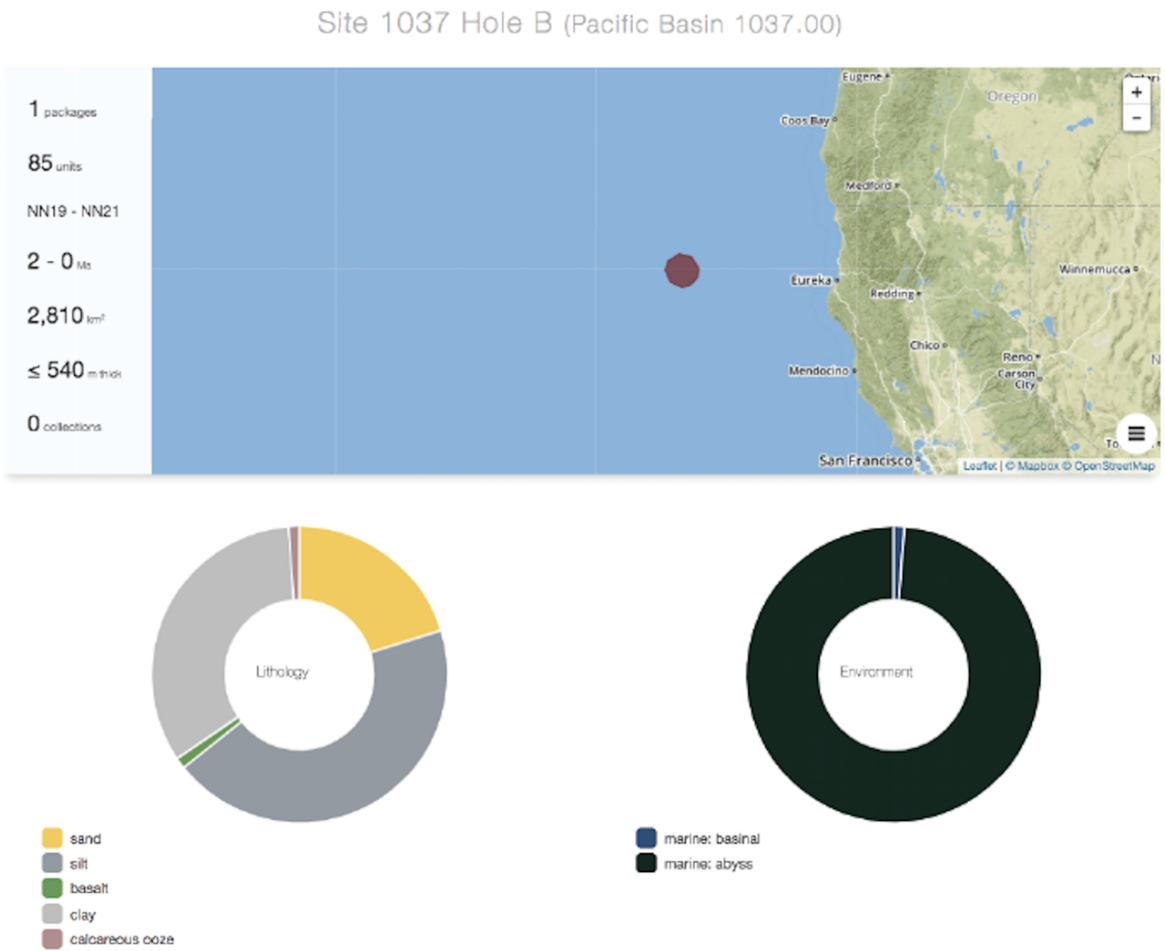
STRATIGRAPHY

Macrostrat



Macrostrat (<http://www.macrostrat.org>), started in 2005, has assembled chronostratigraphic rock units and basic properties in order to enable quantitative analyses of large-scale sedimentation patterns around the globe. The database currently contains 1,474 regional rock columns, 33,903 rock units from North America, the Caribbean, New Zealand, and the marine realm and is integrated with more than 1.89 million geologic map polygons from over 100 separate geological maps sources spanning the globe at multiple different spatial resolutions.

Macrostrat will house ocean drilling lithologic data along with site age models. These data will be searchable by lithology type, age, and region allowing scientists to study marine sedimentation trends through time. The program will host alternate depth scales to accommodate spliced sequences. Macrostrat will allow for multiple age models to be stored for each site so that researchers can choose the age model they prefer or input their own. Through an API, the age models in Macrostrat will be linked to the ocean drilling taxonomic data in the PBDB.



PROJECT STATUS AND OUTREACH

Current Project Status

- Held the first PI-developer meeting in January of 2020.
- Creating external database that compiles all of the microfossil and lithology data in a uniform structure that can be easily imported to Macrostrat and the PBDB. This database currently houses IODP data from the LIMS database. This will be extended to older datasets and the Neptune SandBox.
- Entering microfossil taxonomy and biostratigraphic zonations into Macrostrat and the PBDB to prepare for injection of LIMS data.
- Begin planning for a Stratigraphy-focused workshop that will take place in the next year. Challenges associated with planning during COVID.

Outreach and Broader Impacts

We are excited to partner with NSF Flyover Country (<https://flyovercountry.io/>) to bring ocean drilling science to a broader community. Flyover Country already presents information from the PBDB and Macrostrat on their platform. Now, eODP data will be accessible as well. We will be engaging undergraduate students to create pop-up content that illustrates major discoveries of ocean drilling.



PERSONNEL AND ACKNOWLEDGEMENTS

Associated Personnel

Brian Huber - *Smithsonian National Museum of Natural History*

Doug Fils - *Open Core Data, Consortium for Ocean Leadership*

Daven Quinn - *Macrostrat, U. of Wisconsin*

Michael McClellenn - *PBDB, U. of Wisconsin*

Seth Kaufman and Wai-Yin Kwan - *PBDB, Whirl-i-Gig*

Shane Loeffler and Reed McEwan - *Flyover Country, U. of Minnesota*

Acknowledgements

eODP is funded by NSF grant ICER-1928362; PBDB by NSF EAR- 0949416, NSF ICER-1540929, NSF ICER-1540997, and NSF DUE-1504718; Macrostrat by NSF EAR-1150082, NSF ICER-1440312, and U. Wisconsin, Madison Dept. of Geoscience; OCD by NSF EAR-1550787; Flyover Country by NSF EAR-1541800, NSF EAR-1550913, and NSF ICER-1643329. L.J.L. salary is supported by NSF OCE-1326927. A.F.J. is supported by the Vice Chancellor's Fellowship of the University of Bristol.