

Saving Our Marine Archives

A concerted effort has begun to gather and preserve archives of marine samples and descriptive data, giving scientists ready access to insights on ancient environments.



A scientist collects a core sample from a coral colony on a western Australian reef. The MARPA project is working to establish standards and capabilities for preserving the metadata related to physical samples like this one in a format and centralized location in a form that facilitates future access and research. Credit: Eric Matson, Australian Institute of Marine Science

By [Emilie Dassié](#), Kristine DeLong, Hali Kilbourne, Branwen Williams, Nerilie Abram, Logan Brenner, Chloé Brahmi, Kim M. Cobb, Thierry Corrège, Delphine Dissard, Julien Emile-Geay, Heitor Evangelista, Michael N. Evans, Jesse Farmer, Thomas Felis, Michael Gagan, David P. Gillikin, Nathalie Goodkin, Myriam Khodri, Ana Carolina Lavagnino, Michèle LaVigne, Claire Lazareth, Braddock Linsley, Janice Lough, Helen McGregor, Intan S. Nurhati, Gilman Ouellette, Laura Perrin,

Maureen Raymo, Brad Rosenheim, Michael Sandstrom, Bernd R. Schöne, Abdelfettah Sifeddine, Samantha Stevenson, Diane M. Thompson, Amanda Waite, Alan Wanamaker, and Henry Wu  24 February 2017

Some scientific communities, including paleontologists and archaeologists, curate their physical samples in museums. In the paleoceanography and paleoclimatology communities, however, scientists generally store their physical samples in their own facilities, unless they are affiliated with international core drilling programs, which maintain centralized archives.

Data derived from the physical samples are usually (but not always) uploaded to an electronic database, but individual curation of the associated samples and metadata imposes limits on future retrieval and further research. This also places these resources at risk from hazards like floods and fires. Likewise, storing metadata in an individual lab, personal computer files, or field notebooks limits the ability of other scientists to make use of these samples.

Historically, one way to encourage voluntary data submission to the existing data archives was to keep the formatting and metadata requirements to a minimum. However, regularized data standards could make future data synthesis much easier, and these standards are potentially viable now that funding agencies, including the National Science Foundation, have begun requesting that proposals include detailed data management plans for archiving both physical samples and electronic data. However, these agencies provide no guidance on what practices are acceptable for a given scientific community.

Paleoceanographers and paleoclimatologists are beginning to define the standards they need to properly store physical samples, metadata, and derived proxy data.

To solve these data archival and hosting problems, paleoceanographers and paleoclimatologists are beginning to define the standards they need to properly store physical samples, metadata, and derived proxy data. The [Marine Annually Resolved Proxy Archives](https://marpsite.wordpress.com/about/) (MARPA) project is a grassroots effort by the scientific community to build consensus on data and [sample archiving procedures](https://eos.org/opinions/embracing-open-data-in-field-driven-sciences) (<https://eos.org/opinions/embracing-open-data-in-field-driven-sciences>), working with existing and new data repositories to ensure that the needs of the community are represented.

Scattered Archives

Paleoceanographers and paleoclimatologists collect, archive, and study marine materials, which serve as important sources of [information on past climates](https://eos.org/meeting-reports/high-resolution-tools-advance-study-of-paleoclimate-archives) (<https://eos.org/meeting-reports/high-resolution-tools-advance-study-of-paleoclimate-archives>). In cases where changes in the properties of these specimens are dependent on climate variables, these changes serve as an indirect indicator (or proxy) for changes in Earth's climate through time.

Within these communities, the need for standards and metadata is particularly important for scientists who use marine specimens with relatively high accretion rates—corals, mollusks, coralline algae, sclerosponges (hard sponges), varved sediments, and the like—as samples, metadata, and data are scattered in individual labs and select repositories.

Geochemical data and their related metadata extracted from physical samples are often stored on a researcher's and/or lab's computers until they are published and stored in public repositories, including those at the National Oceanic and Atmospheric Administration's National Centers for Environmental Information Paleoclimate Program ([NOAA-Paleo \(https://www.ncdc.noaa.gov/data-access/paleoclimatology-data\)](https://www.ncdc.noaa.gov/data-access/paleoclimatology-data)) and [PANGAEA \(http://www.pangaea.de/\)](http://www.pangaea.de/), which is hosted by the German organizations Alfred Wegener Institute Helmholtz Center for Polar and Marine Research (AWI) and Center for Marine Environmental Sciences (MARUM) at the University of Bremen.

The data shared with these repositories include basic metadata, such as location (latitude, longitude, and elevation or water depth), time interval, title, investigators, publication, taxonomy, and reconstruction variables. However, these metadata are not necessarily linked to the physical samples and other important sample metadata such as sample size, field site pictures, and imaging of the samples (e.g., X-rays, thin sections, etc).

Paleoclimate research has produced a critical mass of data that is now comprehensive enough to be used for large-scale research syntheses.

From the ever-increasing array of marine annually resolved proxy archives, the paleoclimate research endeavor has produced a critical mass of data that is now comprehensive enough to be used for large-scale research syntheses. Communities in addition to paleoclimatologists use these data, including statisticians, modelers, data assimilators, anthropologists, ecologists, historians, and policy makers. Conducting more and more data-intensive research requires the ability to probe large sets of diverse data efficiently. This, in turn, requires standard data-archiving practices, along with the computer storage facilities and the [means of accessing them \(https://eos.org/editors-vox/your-science-is-your-openly-shared-data\)](https://eos.org/editors-vox/your-science-is-your-openly-shared-data). Ultimately, this would enable use and reuse of valuable physical samples and the data derived from them.

The MARPA Story

Early in 2011, the [EarthCube program \(http://earthcube.org/\)](http://earthcube.org/) instigated an effort to develop a community-driven [cyberinfrastructure \(https://kb.iu.edu/d/auhf\)](https://kb.iu.edu/d/auhf) that supports standards for interoperability, promotes advanced technologies to improve and facilitate interdisciplinary research, and helps educate scientists in the emerging practices of digital scholarship, data and software stewardship, and open science.

Among the communities targeted by the EarthCube program, some communities, including scientists using marine annually resolved proxy archives, have yet to establish a cyberinfrastructure with

improved standards for storage and sharing of paleoclimate data and archive-specific metadata on the physical samples.

To address this need, the MARPA project started in 2013 under the EarthCube umbrella and has been growing ever since. MARPA aims to do the following:

advocate to establish cyberinfrastructures for the annually resolved proxy community through discussions among community members.

deliver guidance on creating community-generated standards for data sharing.

identify tools that facilitate the archiving of metadata for physical materials and data (geochemical or other derived data) for their group and the wider community, which will increase accessibility to data and specimens while creating a lasting legacy for future paleoclimate research beyond the careers of individual researchers.

The MARPA group is working with other initiatives to create new practices for storing and sharing data.

The MARPA group is working with other initiatives to create new practices for storing and sharing data. Some examples include generating standard names for measured variables in conjunction with NOAA's National Centers for Environmental Information ([NOAA NCEI \(https://www.ncei.noaa.gov/\)](https://www.ncei.noaa.gov/)) project. We are working with the Past Global Changes consortium community ([PAGES \(http://www.futureearth.org/projects/pages-past-global-changes\)](http://www.futureearth.org/projects/pages-past-global-changes)) to make sure that our standards for data sharing integrate with other archive standards. We are involved with EarthCube's Internet of Samples in the Earth Sciences (iSamples), the Cyberinfrastructure for Paleogeosciences ([C4P \(http://earthcube.org/group/c4p\)](http://earthcube.org/group/c4p)), and the [LinkedEarth \(http://linked.earth/\)](http://linked.earth/) projects. These groups work to advance the use of innovative cyberinfrastructure to connect physical sample collections across the Earth sciences, advance the role of cyberinfrastructure in unraveling large-scale data sets, and better organize and share Earth science data, especially paleoclimatic data. Our involvement in those initiatives will make sure that MARPA's voice is heard in shaping the future of data archiving and sharing.

Registering Metadata on Physical Samples

Our first recommendation is to use the System for Earth Sample Registration ([SESAR \(http://www.geosamples.org/\)](http://www.geosamples.org/)). SESAR, a registry for physical samples taken from the natural environment, has the capacity needed to store physical sample metadata for the annually resolved proxy community. SESAR catalogs and archives sample metadata, and it offers tools and services for users to manage their sample metadata and to obtain an International Geo Sample Number ([IGSN \(http://www.igsn.org/\)](http://www.igsn.org/)) that ensures unique identification and unambiguous citation of samples. SESAR developed the IGSN, but the system has become an international standard for sample identification. Several scholarly journal publishers, including the American Geophysical Union (AGU) and Elsevier,

recently recommended (<https://eos.org/agu-news/agu-opens-its-journals-to-author-identifiers>) that IGSNs be used for referencing physical samples.

Registering samples in SESAR is easy: investigators can download a template as an Excel file, save it on a personal computer, and upload data to SESAR using this template at their convenience. Investigators are able to control the privacy of their sample metadata, meaning that they can store and manage their physical sample metadata in SESAR, obtain IGSNs for each sample, and make that information public in SESAR only when they are ready: upon publication release, for example.

Physical sample IGSNs can be inserted into publications, saving space and time by providing a compact reference to all the information on the physical samples that is freely accessible online, rather than cluttering journal articles with metadata that are important but perhaps peripheral to the particular scientific findings of the article. The IGSN can link to the catalog of an institutional repository, where physical samples are stored and curated.

SESAR facilitates metadata management and may make sample management easier, for example, by providing the ability to easily print labels. The Lamont Doherty Earth Observatory Core Repository has taken advantage of these capabilities to organize and catalog their physical samples (Figure 1). The [MARPA website](https://marpasite.wordpress.com/) (<https://marpasite.wordpress.com/>) provides an example template and a video tutorial on how to register physical samples in SESAR, in particular those from annually resolved proxies.



Fig. 1. Lamont-Doherty Coral Core Repository at Columbia University before and after organizing coral samples in boxes. Each sample was given an IGSN and Quick Response (QR) code for the

storage box so that core pictures, X-ray images, and other available information can be linked to the samples. Credit: Emilie P. Dassié and Michael Sandstrom

Going Forward

MARPA's goal is to continue engaging with the marine (as well as the emerging freshwater) annually resolved proxy archive community to ensure that our data storage and retrieval needs are represented in the existing cyberinfrastructures and those that are currently being developed. We are actively seeking feedback from the community to address these concerns.

We are asking the MARPA community to describe their needs and desires regarding the storage and retrieval of their geochemical data.

In addition to promoting the use of SESAR to store physical sample metadata and training investigators to use SESAR, the MARPA project has defined several goals for the near future. We are asking the MARPA community to describe their needs and desires regarding the storage and retrieval of their geochemical data. We have and will continue to incorporate our community's input into existing projects such as the NOAA NCEI paleoclimatology program and LinkedEarth as they evolve and change to meet the needs of researchers.

We first presented this initiative at the 2015 AGU Fall Meeting, then at the 2016 Ocean2K and 2016 Sclerochronology meetings, and we will present further developments in the work and recommendations of MARPA at the PAGES Open Science Meeting in May 2017 to help other communities move forward with their specific archival issues. We will use the MARPA website forum section to discuss various ongoing aspects of our community needs, including how to store physical samples that have already been analyzed once their metadata have been registered in SESAR.

More information on the MARPA project, updates on recent activities, and channels for providing feedback are available on the [MARPA website](https://marpsite.wordpress.com/) (<https://marpsite.wordpress.com/>).

Acknowledgments

We thank the current and former NOAA Paleoclimate team for helpful discussions, David Anderson, Eugene Wahl, Carrie Morill, and Bridget Thrasher, as well as Kerstin Lehnert and Megan Carter of the Interdisciplinary Earth Data Alliance, the data facility that operates SESAR. We also thank the attendees of the first MARPA workshop during the 2014 Ocean Sciences Meeting and the informal MARPA town hall meeting at the 2015 AGU Fall Meeting. This is UMCES contribution number 5251.

Author Information

Emilie Dassié (email: edassie@ldeo.columbia.edu (<mailto:edassie@ldeo.columbia.edu>)), Lamont-Doherty Earth Observatory, Columbia University, New York, N.Y.; and Laboratoire d'Océanographie et du Climat: Expérimentations et Approches Numériques (LOCEAN), Paris, France; Kristine DeLong, Department of Geography and Anthropology, Louisiana State University, Baton Rouge; Hali

Kilbourne, University of Maryland Center for Environmental Science, Solomons; Branwen Williams, Claremont McKenna-Pitzer-Scripps Colleges, Claremont, Calif.; Nerilie Abram, Research School of Earth Sciences, Australian National University (ANU), Canberra; Logan Brenner, Lamont-Doherty Earth Observatory, Columbia University, New York, N.Y.; Chloé Brahmi, Université de la Polynésie Française, Faa'a, Tahiti; Kim M. Cobb, Georgia Tech University, Atlanta; Thierry Corrège, Université Bordeaux, Pessac, France; Delphine Dissard, Institut de Recherche pour le Développement (IRD), Sorbonne Universités (Université Pierre et Marie Curie (UPMC), Université Paris 6), Centre National de la Recherche Scientifique (CNRS), Muséum National d'Histoire Naturelle (MNHN), LOCEAN, Paris, France; Julien Emile-Geay, University of Southern California, Los Angeles; Heitor Evangelista, Laboratory of Radioecology and Global Changes, Roberto Alcântara Gomes Institute of Biology, Rio de Janeiro State University (LARAMG/IBRAG/UERJ), Rio de Janeiro, Brazil; Michael N. Evans, University of Maryland, College Park; Jesse Farmer, Lamont-Doherty Earth Observatory, Columbia University, New York, N.Y.; Thomas Felis, MARUM-Center for Marine Environmental Sciences, University of Bremen, Germany; Michael Gagan, Research School of Earth Sciences, ANU, Canberra; David P. Gillikin, Union College, Schenectady, N.Y.; Nathalie Goodkin, Earth Observatory of Singapore, Singapore; Myriam Khodri, IRD, Sorbonne Universités (UPMC, Université Paris 6), CNRS, MNHN, LOCEAN, Paris, France; Ana Carolina Lavagnino, LARAMG/IBRAG/UERJ, Rio de Janeiro, Brazil; Michèle LaVigne, Bowdoin College, Brunswick, Maine; Claire Lazareth, IRD, Sorbonne Universités (UPMC, Université Paris 6), CNRS, MNHN, LOCEAN, Bondy, France; Braddock Linsley, Lamont-Doherty Earth Observatory, Columbia University, New York, N.Y.; Janice Lough, Australian Institute of Marine Science, Townsville, Qld., Australia; Helen McGregor, University of Wollongong, NSW, Australia; Intan S. Nurhati, Indonesian Institute of Sciences, Jakarta, Indonesia; Gilman Ouellette, Louisiana State University, Baton Rouge; Laura Perrin, LOCEAN, Paris, France; Maureen Raymo, Lamont-Doherty Earth Observatory, Columbia University, New York, N.Y.; Brad Rosenheim, University of South Florida, St. Petersburg; Michael Sandstrom, Lamont-Doherty Earth Observatory, Columbia University, New York, N.Y.; Bernd R. Schöne, Institute of Geosciences, University of Mainz, Germany; Abdelfettah Sifeddine, IRD, Sorbonne Universités (UPMC, Université Paris 6), CNRS, MNHN, LOCEAN, Bondy, France; Samantha Stevenson, National Center for Atmospheric Research, Boulder, Colo.; Diane M. Thompson, Boston University, Mass.; Amanda Waite, University of Miami, Coral Gables, Fla.; Alan Wanamaker, Iowa State University, Ames; and Henry Wu, IRD, Sorbonne Universités (UPMC, Université Paris 6), CNRS, MNHN, LOCEAN, Bondy, France

Citation: Dassié, Emilie, et al. (2017), Saving our marine archives, *Eos*, 98, <https://doi.org/10.1029/2017EO068159>.
Published on 24 February 2017.

Text © 2017. The authors. [CC BY-NC-ND 3.0](https://creativecommons.org/licenses/by-nc-nd/3.0/)

Except where otherwise noted, images are subject to copyright. Any reuse without express permission from the copyright owner is prohibited.