Supporting Information for The Modern Ocean Sediment Archive and Inventory of Carbon (MOSAIC): version 2.0

Sarah Paradis[1], Kai Nakajima[1], Tessa S. Van der Voort[2], Hannah Gies[1], Aline Wildberger[1], Thomas Blattmann[1], Lisa Bröder[1], Tim Eglinton[1]

[1]Department of Earth Sciences, Geological Institute, ETH Zürich, Sonneggstrasse 5, 8092 Zürich, Switzerland
[2]Nutrient Management Institute, 6709PA Wageningen, Netherlands

Correspondence to: Sarah Paradis (sparadis@ethz.ch)

Contents

This document holds supplementary tables and figures that further describe the contents of the database in more detail. Tables S1-S2 explains the metadata stored in key tables: geopoints (Table S1), sample analyses (Table S2). Table S3 describes the different fractions of material analyzed that are currently supported by MOSAIC 2.0. Table S4 shows the equations employed to harmonize the data stored in MOSAIC. Figure S1 distinguishes the cores that have been included in MOSAIC thanks to georeferencing tools. Figure S2 is an example of the hierarchy of the methods stored in the metadata, to ensure comparability between datapoints (more details in Dataset S2). Finally, Figure S3 shows a general comparison between the variables stored in the previous iteration and in this new MOSAIC v.2.0.

The following additional files can also be found in the supplementary information:

- Dataset S1 describing the quality check control for each variable, including the different classifications of the method techniques and Dataset S2 describing all the analyses that are stored in MOSAIC 2.0
- SQL file to reproduce the PostgreSQL database

Tables

Table S1. Metadata reported in the geopoints table in the MOSAIC v2.0 database.

Column name	Data type	Description
core_id	Integer	Core identifier (primary key).
core_name	Variable character	Free-text of the core name.
latitude	Float	Decimal latitude (WGS 84) of the geopoint.
longitude	Float	Decimal longitude (WGS 84) of the geopoint.
georeferenced_coordinates	Integer	True (1) or false (0) assignment of whether the data was georeferenced (1) or obtained from the manuscript/dataset (0).
sampling_date	Datetime	Date-time format of the sampling date, if available.
sampling_year	Integer	Numerical value of the sampling year
sampling_month	Integer	Numerical value of the sampling month.
sampling_day	Integer	Day of the month of the sampling.
sampling_method_id	Integer	Identifier of the sampling method (foreign key), which is linked to the *sampling_method* table that stores the description of the different sampling gear types.
sampling_campaign_id	Integer	Identifier of the sampling campaign (foreign key), which is linked to the *sampling_campaign* table that stores the metadata associated to the cruise (i.e. vessel name, cruise name, cruise dates, equipment used).
core_length_cm	Integer	Core length (cm) of the geopoint extracted.
geomorphological_site	Variable character	Geomorphological site where the sample was collected from (i.e. shelf, slope, abyss, rise, canyon, trench)
core_comment	Variable character	Free-text of any important additional information of the sampled core.
entry_user_name	Variable character	Free-text of the user that input the data

Table S2. Metadata reported in the sample analysis table in the MOSAIC v2.0 database.

Column name	Data type	Description
sample_id	Integer	Sample identifier (foreign key), which is linked to the *samples* table that stores all the metadata associated to the sample (i.e. depth in the sediment core, core identifier)
material_analyzed_id	Integer	Identifier of the material analyzed (foreign key), which is linked to the *sample_material_analyzed* table that stores the description of the material analyzed (i.e. bulk, TOC, grain size fraction)
sample_analysis_id	Integer	Identifier of the variable analyzed (foreign key), which is linked to the *sample_analysis_description* table that stores the description of each analysis (i.e. total_organic_carbon_%, Delta_14C)
sample_exclusivity_clause	Integer	True (1) or false (0) assignment of whether the data is confidential or not (published).
sample_core_id	Integer	Core identifier (foreign key), which is linked to the *geopoints* table that stores all the metadata associated to the sampled geopoint (i.e. coordinates, sampling date, sampling method).
sample_analysis_calculated	Integer	True (1) or false (0) assignment of whether the data was calculated during the harmonization process.
sample_article_id	Integer	Article identifier (foreign key), which is linked to the *articles* table that stores the metadata associated to the published manuscript, thesis, or dataset (i.e. title, year of publication, title, DOI). If the data is not published, it is assigned a value of 0.
sample_contact_person_id	Integer	Identifier of the contact person of those analyses (foreign key), either the first author of the published manuscript where the data came from, or the owner of the unpublished data, which is linked to the *authors* table that stores the first and last name of the person.
method	Variable character	High-level classification of the method used to obtain the analysis.
method_details	Variable character	Specific method used to obtain the analysis (i.e. instrument model, sample preparation method, measured chain lengths of biomarkers)
entry_user_name	Variable character	Free-text of the user that input the data
replicate	Integer	Incremental replicate number. If replicate analyses are conducted for the same sample, the replicate number increments.

Table S3. Material analysed fractions stored in MOSAIC v2.0 database.

	Material analyzed	Description
	Bulk	Analyses performed on bulk sediment, without any physical or chemical separation. This is the default material analyzed.
	TOC	Analyses only performed on organic carbon fraction, after removing the inorganic carbon fraction, such as ^{14}C and ^{13}C analyses.
	TIC	Analyses only performed on inorganic carbon fraction, after removing the organic carbon fraction.
	< 2mm	Analyses performed only on the sediment fraction smaller than 2 mm, removing any gravel or coarser sediment.
Grain size and density fractions	> 0.5 mm	Analyses performed only on the sediment fraction larger than 0.5 mm.
	> 0.250 mm	Analyses performed only on the sediment fraction larger than 0.25 mm.
	0.25-0.5 mm	Analyses performed only on the sediment fraction between 0.25-0.5 mm.
	0.125-0.5 mm	Analyses performed only on the sediment fraction between 0.125-0.5 mm.
	> 0.063 mm	Analyses performed only on the sediment fraction larger than 63 µm.
	0.063-0.125 mm	Analyses performed only on the sediment fraction between 63 µm – 0.125 mm.
	0.010-0.063 mm	Analyses performed only on the sediment fraction between 10-63 µm.
	0.032-0.063 mm	Analyses performed only on the sediment fraction between 32-63 µm.
	0.02-0.032 mm	Analyses performed only on the sediment fraction between 20-32 µm.
	0.002-0.010 mm	Analyses performed only on the sediment fraction between 2-10 µm.
	< 0.020 mm	Analyses performed only on the sediment fraction smaller than 20 µm.
	< 0.002 mm	Analyses performed only on the sediment fraction smaller than 2 µm.
	> 2.5 g/cm³	Analyses performed only on the sediment denser than 2.5 g/cm³.
	2-2.5 g/cm³	Analyses performed only on the sediment with density between 2-2.5 g/cm³.
	1.6-2 g/cm³	Analyses performed only on the sediment with density between 1.6-2 g/cm³.
	<1.6 g/cm³	Analyses performed only on the sediment less dense than 1.6 g/cm³.
	Alkenones (C$_{35}$-C$_{39}$)	Analyses performed on alkenones
Biomarkers and related	Fatty acids (C$_{14}$-C$_{32}$)[1]	Compound-specific isotopic or radiocarbon analyses on isolated fatty acids (alkanoic acids) with carbon lengths ranging from 14 to 32.
	Alkanes (C$_{17}$-C$_{35}$)[1]	Compound specific isotopic or radiocarbon analyses on isolated alkanes with carbon lengths ranging from 17 to 35.
	Alcohols (C$_{16}$-C$_{30}$)[1]	Compound specific isotopic or radiocarbon analyses on isolated alcohols with carbon lengths ranging from 16 to 30.
	Vanillyl phenols	Compound specific isotopic or radiocarbon analyses on isolated vanillyl phenols (vanillin, vanillic acid, acetovanillone).
	Syringyl phenols	Compound specific isotopic or radiocarbon analyses on isolated syringyl phenols (syringealdehyde, syringic acid, acetosyringone).
	Lignin phenols	Compound specific isotopic or radiocarbon analyses on isolated lignin phenols (vanillyl phenols, syringyl phenols, cinnamyl phenols).
	Macro-fossils	Analyses performed on vegetal macro-fossils (e.g. roots, leaves, bark) separated from the bulk sediment sample.
	Foraminifera	Analyses performed on pelagic foraminifera (e.g. *Globigerina bulloides*, *Neogloboquadrina incompta*, *Globorotalia inflata*) separated from the bulk sediment sample.

[1]The full list of compounds are not listed for simplicity.

Table S4. Data harmonization equations.

Variable to harmonize	Required variables	Equation
$\Delta^{14}C$ [‰]	Sampling year (yr), Fraction modern ^{14}C (Fm)	$Fm * e^{\left(\frac{1}{8267}*(1950-yr)-1\right)} * 1000$
Fraction modern ^{14}C	Age ^{14}C (age)	$e^{\left(-\frac{age}{8033}\right)}$
Fraction modern ^{14}C	Sampling year (yr), $\Delta^{14}C$	$\left(\frac{\Delta^{14}C}{1000}+1\right) * e^{\left(-\frac{1}{8267}*(1950-yr)\right)}$
Age ^{14}C [ybp]	Fraction modern ^{14}C (Fm)	$-8033 * \ln(Fm)$
Total organic carbon [%]	Total carbon (TC), Total inorganic carbon (TIC)	$TC - TIC$
Total inorganic carbon [%]	Total carbon (TC), Total organic carbon (TOC)	$TC - TOC$
Total carbon [%]	Total organic carbon (TOC), Total inorganic carbon (TIC)	$TOC + TIC$
OC:TN ratio	Total organic carbon (TOC), Total nitrogen (TN)	$\frac{TOC}{TN} * \frac{14}{12}$
Porosity	Water content [%] (w)	$w / 100$
Water content [%]	Porosity (φ)	$\varphi \cdot 100$

Figures

Map-derived coordinates (n = 1264)
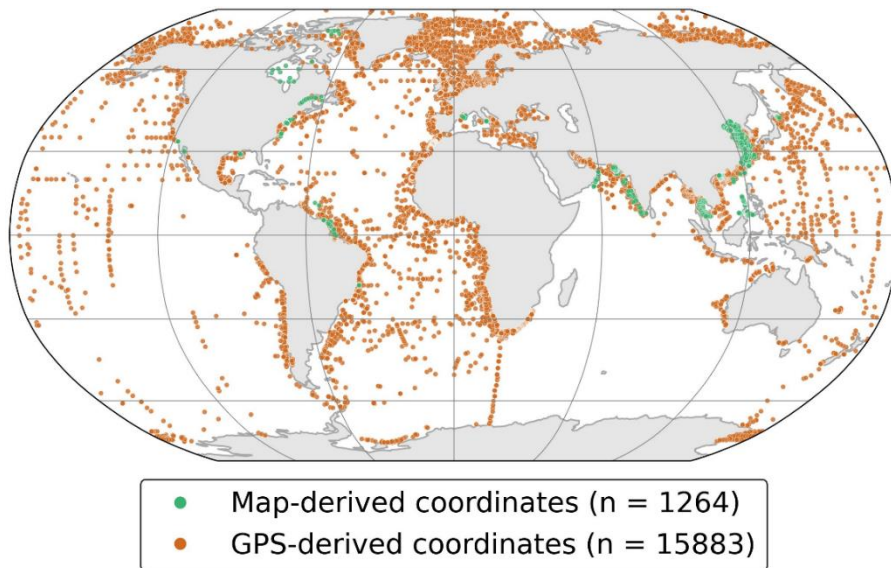GPS-derived coordinates (n = 15883)

Figure S1. Spatial distribution of sediment cores stored in MOSAIC v2.0 that have been georeferenced or whose coordinates were provided in the manuscript.
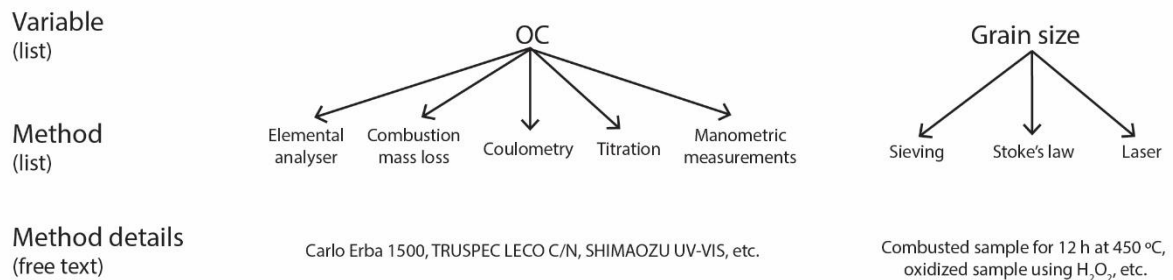

Figure S2: Example of reported metadata regarding the method employed, to ensure data comparability.
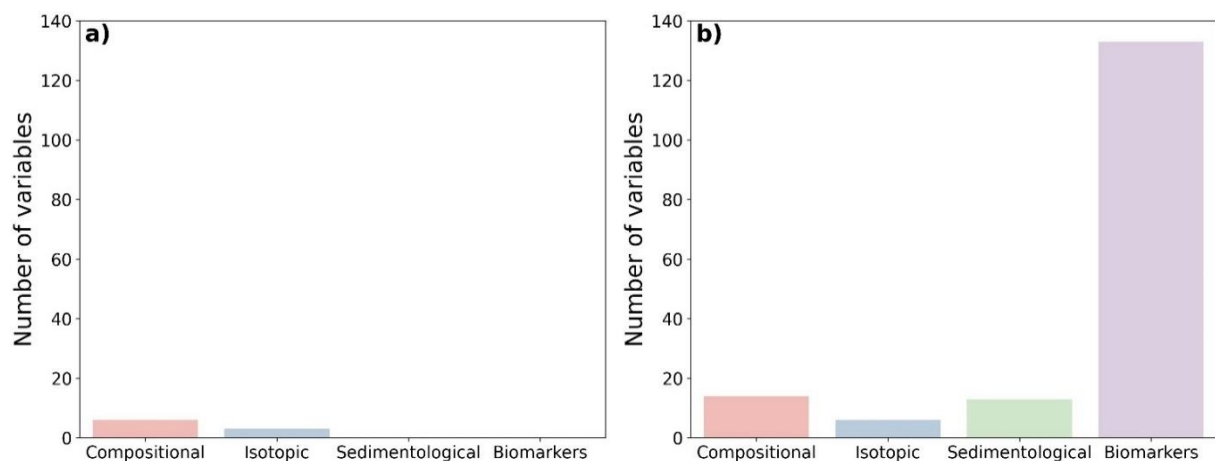

Figure S3. Number of variables stored in the first iteration of MOSAIC (a) and in MOSAIC v.2.0 (b), grouped by categories: compositional (e.g., OC, TN, CaCO3, biogenic opal), isotopic (e.g., δ13C, δ15N, Δ14C), sedimentological (e.g., grain size fractions, mineral surface area, dry bulk density), and biomarkers (e.g., lignin phenols, alkanes, fatty acids, alcohols). Note that biomarkers have a large number of variables due to the different molecules that can be analysed for each biomarker class (see section 3.2 for more details).