Data management workflows for experiment and model data

Hela Mehrtens
GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany
Agenda

• Data management: who, when and how
• Components of data management workflow
• Examples:
  • Ocean modeling
  • Mesocosm experiments
  • Samples
Data management: who, when and how

Who: scientists, data stewards, data manager

When: proposal, preparation of work, during expeditions/experiments/coding, exchange of results, publishing of results

How: agreement of group, set up of systems, using tools, monitoring and visibility
Central data management workflow @ GEOMAR

1. Agreement on one data policy or DMP (expected data, responsible persons, timeframes)
2. Information system with metadata and linked data for exchange and monitoring deliverables (OSIS)
3. Accessibility and reuseability of research data based on persistent identifiers (DOI + handle)
1. Data management plan (DMP)

- The **DMP** describes the expected data and how and when they will be handled, stored and made available.

- The information system **keeps track** on the **deliverables** and sends reminder.

- The result is an **overview on the outcome** of the project => data management record.
2. OSIS: Ocean Science Information System

• tool for search and share data for ongoing projects

• metadata are open access

• data can be login protected for project groups

• links to longterm repositories

→ Helps transparency (Open Science)
2. OSIS

**OSIS Ocean Science Information System for Expeditions, Numeric Models, Experiments...**

View Terms of Use

### My Deliverables

<table>
<thead>
<tr>
<th>Expedition</th>
<th>Name</th>
<th>Flag</th>
<th>Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>AL418</td>
<td>OFOS Data</td>
<td>OPEN</td>
<td>2016-07-02</td>
</tr>
<tr>
<td>M131</td>
<td>MSS</td>
<td>OSIS</td>
<td>2019-11-08</td>
</tr>
<tr>
<td>MSM72</td>
<td>LADCP</td>
<td>PUBLISHED</td>
<td>2020-09-30</td>
</tr>
<tr>
<td>MSM72</td>
<td>VMADDOP</td>
<td>PUBLISHED</td>
<td>2020-09-30</td>
</tr>
<tr>
<td>MSM72</td>
<td>underway CTD</td>
<td>PUBLISHED</td>
<td>2020-09-30</td>
</tr>
<tr>
<td>M119</td>
<td>Mooring data</td>
<td>PUBLISHED</td>
<td>2019-12-12</td>
</tr>
</tbody>
</table>

### Latest links

- **Type**: DSHIP, URL, Print Publication, Weekly Report
- **Description**: DSHIP Sensor data available at GEOMAR, Map with cruise track, Weekly reports
- **Linked to**: Leg: ALS62, Leg: ALS63, Leg: M176

### Files

- **Uploaded**: 2021/09/20
- **Creator**: Mehrten, Hela

### Expeditions

- **Label**: IODP Exp.396, SO293, MSM103
- **Departure - Return**: 2021/08/06 - 2021/10/06, 2022/07/17 - 2022/09/06, 2021/09/12 - 2021/11/07
- **Chief-Scientist**: Berndt, Christian, Brandt, Peter, Holz, Sebastian

### Models

- **Label**: VIKING20K.L46-K, NorESM-T4.4.3, NorESM-T4.4.2
- **Updated**: 2021/06/17, 2021/03/08, 2021/03/08
- **Classified as**: Person

### Experiments

- **Experiment**: KOSMOS 2021 Gran Canaria, KOSMOS 2022 Bergen, Dissolution Experiments
- **Updated**: 2021/05/27, 2021/01/14, 2021/05/12
- **Responsible Person**: Haunast, Matthias, Haunast, Matthias, Söhmer, Niels
3. Persistent identifier

DOI-Registry at GEOMAR: Workflows

Scientists

Data

Software Code in (public) GitLab repository

METADATA, URL & zip-file

GEOMAR Library/OceanRep

PANGAEA - Data Publisher for Earth & Environmental Science

DataCite - FIND, ACCESS, AND REUSE DATA
Example: Experiment workflow

- Cruise planning involves DMP components
- Data reminder in OSIS used for several projects
- Deliverable done: data are openly available as data publications with DOI at PANGAEA
Mesocosm experiments

- Information System used for 15 experiments (2009-2022)
- More than 130 data publications at PANGAEA

<table>
<thead>
<tr>
<th>Label</th>
<th>Subject</th>
<th>Species</th>
<th>Description</th>
<th>Duration of Experiment</th>
<th>PI</th>
<th>Files</th>
<th>Links</th>
<th>Action...</th>
</tr>
</thead>
<tbody>
<tr>
<td>KODOMOS 2008</td>
<td>Mesocosm</td>
<td>Eck</td>
<td>Borko Eck in Kiel Bight at about 54.538N, 19.038E</td>
<td>20.05.2009 - 05.07.2009</td>
<td>Rubesol, Uf</td>
<td>0</td>
<td>1</td>
<td>Action...</td>
</tr>
<tr>
<td>KODOMOS</td>
<td>Moscosm</td>
<td>Experiment</td>
<td>Kongfjorden, Ny-Alesund, on the north-west coast of Spitsbergen, Svalbard archipelago, Norway</td>
<td>31.05.2010 - 08.07.2010</td>
<td>Rubesol, Uf</td>
<td>1</td>
<td>24</td>
<td>Action...</td>
</tr>
<tr>
<td>KODOMOS 2011</td>
<td>Mesocosm</td>
<td>Experiment</td>
<td>Equipped Marine Biological Station of University of Bergen, Raunafjord, Norway</td>
<td>01.06.2011 - 31.07.2011</td>
<td>Rubesol, Uf</td>
<td>1</td>
<td>10</td>
<td>Action...</td>
</tr>
<tr>
<td>KODOMOS 2013</td>
<td>Mesocosm</td>
<td>Experiment</td>
<td>Zoologische Station Trømmene, Finfjord</td>
<td>13.06.2012 - 08.08.2012</td>
<td>Rubesol, Uf</td>
<td>8</td>
<td>17</td>
<td>Action...</td>
</tr>
<tr>
<td>KODOMOS</td>
<td>Mesocosm</td>
<td>Experiment</td>
<td>Kristineberg, Gullmarfjord, Sweden</td>
<td>29.01.2013 - 30.06.2013</td>
<td>Rubesol, Uf</td>
<td>8</td>
<td>9</td>
<td>Action...</td>
</tr>
<tr>
<td>KODOMOS 2013</td>
<td>Mesocosm</td>
<td>Experiment</td>
<td>Gran Canaria</td>
<td>23.09.2014 - 05.12.2014</td>
<td>Rubesol, Uf</td>
<td>2</td>
<td>7</td>
<td>Action...</td>
</tr>
<tr>
<td>KODOMOS</td>
<td>Mesocosm</td>
<td>Experiment</td>
<td>Equipped Marine Biological Station of University of Bergen, Raunafjord, Norway</td>
<td>03.05.2015 - 30.06.2015</td>
<td>Rubesol, Uf</td>
<td>1</td>
<td>6</td>
<td>Action...</td>
</tr>
<tr>
<td>KODOMOS 2014</td>
<td>Mesocosm</td>
<td>Gran Canaria</td>
<td>The study (KODOMOS Gran Canaria 2016) was carried out at the pair of Galapagos, Gran Canaria (Canary Islands), from 2nd March to 5th April 2016.</td>
<td>02.03.2016 - 05.04.2016</td>
<td>Rubesol, Uf</td>
<td>0</td>
<td>3</td>
<td>Action...</td>
</tr>
<tr>
<td>KODOMOS</td>
<td>Mesocosm</td>
<td>Gran Canaria</td>
<td>01.09.2017 - 01.10.2017</td>
<td>Rubesol, Uf</td>
<td>0</td>
<td>0</td>
<td></td>
<td>Action...</td>
</tr>
<tr>
<td>KODOMOS 2017</td>
<td>Mesocosm</td>
<td>Peru</td>
<td>31.01.2017 - 10.04.2017</td>
<td>Rubesol, Uf</td>
<td>36</td>
<td>4</td>
<td></td>
<td>Action...</td>
</tr>
<tr>
<td>KODOMOS</td>
<td>Mesocosm</td>
<td>Gran Canaria</td>
<td>01.09.2018 - 30.11.2018</td>
<td>Rubesol, Uf</td>
<td>36</td>
<td>4</td>
<td></td>
<td>Action...</td>
</tr>
<tr>
<td>KODOMOS 2018</td>
<td>Mesocosm</td>
<td>Gran Canaria</td>
<td>07.09.2019 - 09.10.2019</td>
<td>Rubesol, Uf</td>
<td>20</td>
<td>1</td>
<td></td>
<td>Action...</td>
</tr>
<tr>
<td>KODOMOS</td>
<td>Mesocosm</td>
<td>Peru</td>
<td>01.02.2020 - 15.04.2020</td>
<td>Rubesol, Uf</td>
<td>1</td>
<td>0</td>
<td></td>
<td>Action...</td>
</tr>
<tr>
<td>KODOMOS 2020</td>
<td>Mesocosm</td>
<td>Gran Canaria</td>
<td>D5-4: Comprehensive data set on ecological and biogeochemical responses of a line latitude oligotrophic ocean system to a gradient of acidification intensities</td>
<td>06.09.2021 - 07.10.2021</td>
<td>Rubesol, Uf</td>
<td>0</td>
<td>0</td>
<td>Action...</td>
</tr>
</tbody>
</table>
EXAMPLE I: Experiments

in OSIS

KüNO Workshop 'Data sharing in marine research', 29.9.2021

Mehrtens, Hela (hmehrtens@geomar.de) / Version 4.2r1061 - Logout

Experiment: KOSMOS 2017 Peru
Subject: Mesocosm Experiment
Organisation: Helmholtz Centre for Ocean Research Kiel
Experiment PI: Riebesell, Ulf
Start Date: 2017/02/31
End Date: 2017/04/10
Species:
Life stage:
Sample Origin:
Responsible Person: Bach, Lennart T.
Date of final data: 2017/06
Description: off Peru
Hypothesis:
Comment:
Project Context: [SFB754, SFB754-KOSMOS]
Created: Mehrtens, Hela / 2016-06-22 11:58
Last Updated: Springer, Pina / 2018-04-24 10:14

CREATE NEW RELATED LINK FOR EXPERIMENT KOSMOS 2017 PERU....

Type of Link | Link Address (Path) / Description | Date created | Action
--- | --- | --- | ---
Open Data Repository | https://www.pangaea.de/?q=event%3Alabel%3AKOSMOS_2017_Peru Data archived in PANGAEA | 2021-09-27 15:59 | Action...
Print Publication | https://oceanrep.geomar.de/475668 Overview of activities during the KOSMOS mesocosm study 2017 in the coastal upwelling system off Peru | 2019-05-10 12:07 | Action...
Print Publication | https://oceanrep.geomar.de/43596 Temporal dynamics of surface ocean carbonate chemistry in response to natural and simulated upwelling events during the 2017 coastal El Nino near Callao, Peru | 2019-05-12 11:04 | Action...
Print Publication | https://oceanrep.geomar.de/43529 Nutrient cycling in the Peru upwelling zone under coastal El Niño conditions, a mesocosm study | 2018-07-03 12:42 | Action...
Print Publication | https://oceanrep.geomar.de/41912 Influence of ocean acidification on elemental mass balances and particulate organic matter stoichiometry in natural plankton communities | 2018-03-16 11:34 | Action...

Mehrtens, Hela (hmehrtens@geomar.de) / Version 4.2r1061 - Logout
EXAMPLE I: Experiments in PANGAEA

Citation:
Baumann, Moritz; Sswat, Michael; Ortiz Cortes, Joaquin; Hernández-Hernández, Nauzet; Baños Cerón, Isabel; Vanharanta, Mari; Heinemann, Malte (2021): KOSMOS 2018 Gran Canaria mesocosm study: water column biogeochemistry. PANGAEA, https://doi.pangaea.de/10.1594/PANGAEA.933090 (dataset in review)

Abstract:
The data set contains biogeochemical water column data collected during a KOSMOS mesocosm experiment carried out in the framework of the Ocean Artificial Upwelling project. The experiment was performed in the North-East Atlantic Ocean off the coast of Gran Canaria in autumn 2018 and lasted for 39 days. In this study we investigated the effect of different intensities of artificial upwelling combined with two upwelling modes (recurring additions versus one singular addition) on POC export and its potential transfer efficiency to depth. The data set includes the amounts of surface water that were exchanged with nutrient-rich deep water (from ~300 m depth), primary production and chlorophyll a, elemental composition of suspended particulate matter (POC, PON, C:N) and prokaryotic heterotrophic production.

Keyword(s):
artificial upwelling, carbon sequestration, export flux, mesocosm study, particle properties, remineralization depth, remineralization rate, sinking velocity

Supplement to:
Baumann, Moritz; Taucher, Jan; Paul, Alannah; Heinemann, Malte; Vanharanta, Mari; Bach, Lennart Thomas; Spilling, Kristian; Ortiz Cortes, Joaquin; Aristegui, Javier; Hernández-Hernández, Nauzet; Riebesell, Ulf (in prep.): Effect of different rates and modes of artificial upwelling on particle flux and potential POC deep export.

Project(s):
Ocean Artificial Upwelling (Ocean-artUp)
EXAMPLE I: Experiments

Event(s):

- KOSMOS_2018_Mesocosm-M1: *Latitude: 27.927880, Longitude: -15.364330*, *Date/Time Start: 2018-11-06T00:00:00*, *Date/Time End: 2018-12-14T00:00:00*, *Location: Canarias Sea*, Campaign: KOSMOS_2018 (KOSMOS Gran Canaria), Method/Device: Mesocosm experiment (MESO)

- KOSMOS_2018_Mesocosm-M2: *Latitude: 27.927880, Longitude: -15.364330*, *Date/Time Start: 2018-11-06T00:00:00*, *Date/Time End: 2018-12-14T00:00:00*, *Location: Canarias Sea*, Campaign: KOSMOS_2018 (KOSMOS Gran Canaria), Method/Device: Mesocosm experiment (MESO)

- KOSMOS_2018_Mesocosm-M3: *Latitude: 27.927880, Longitude: -15.364330*, *Date/Time Start: 2018-11-06T00:00:00*, *Date/Time End: 2018-12-14T00:00:00*, *Location: Canarias Sea*, Campaign: KOSMOS_2018 (KOSMOS Gran Canaria), Method/Device: Mesocosm experiment (MESO)

Parameter(s):

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Short Name</th>
<th>Unit</th>
<th>Principal Investigator</th>
<th>Method/Device</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Event label</td>
<td>Event</td>
<td></td>
<td>Baumann, Mortiz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>DATE/TIME</td>
<td></td>
<td></td>
<td>Baumann, Mortiz</td>
<td></td>
<td>Geocode</td>
</tr>
<tr>
<td>3</td>
<td>Experiment day</td>
<td>Exp day</td>
<td>day</td>
<td>Baumann, Mortiz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Mesocosm label</td>
<td>Meso label</td>
<td></td>
<td>Baumann, Mortiz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Treatment</td>
<td>Treat</td>
<td></td>
<td>Baumann, Mortiz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Treatment</td>
<td></td>
<td></td>
<td>Baumann, Mortiz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Deep water exchange, total</td>
<td>Deep water ex</td>
<td>m³</td>
<td>Sowat, Michael</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>DEPTH, water, experiment</td>
<td>Depth water exp</td>
<td>m</td>
<td>Baumann, Mortiz</td>
<td></td>
<td>Geocode – top</td>
</tr>
<tr>
<td>9</td>
<td>DEPTH, water, experiment</td>
<td>Depth water exp</td>
<td>m</td>
<td>Baumann, Mortiz</td>
<td></td>
<td>Geocode – bottom</td>
</tr>
<tr>
<td>10</td>
<td>Primary production</td>
<td>PP</td>
<td>µmol/d</td>
<td>Ortiz Cortes, Joaquin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Primary production, cumulative</td>
<td>PP cum</td>
<td>µmol/d</td>
<td>Ortiz Cortes, Joaquin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Chlorophyll a</td>
<td>Chl a</td>
<td>µg/l</td>
<td>Ortiz Cortes, Joaquin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Carbon, organic, particulate</td>
<td>POC</td>
<td>µmol/l</td>
<td>Baumann, Mortiz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Nitrogen, organic, particulate</td>
<td>PON</td>
<td>µmol/l</td>
<td>Baumann, Mortiz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Carbon/Nitrogen ratio</td>
<td>CN</td>
<td>mol/mol</td>
<td>Baumann, Mortiz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Prokaryotic heterotrophic production</td>
<td>Ph het prod</td>
<td>µmol/d</td>
<td>Baños Cerón, Isabel</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Status: Curation Level: Enhanced curation (CurationLevelC)

Size: 2094 data points
Further Access: DAM Portal

KüNO Workshop 'Data sharing in marine research', 29.9.2021
Example II: Model results

Example: Model results

- Information System contains >300 descriptions
- Links to articles and storage
Example II: Model results

In OSIS

KüNO Workshop ‘Data sharing in marine research’, 29.9.2021
Example II: Model results

In OSIS

KüNO Workshop 'Data sharing in marine research', 29.9.2021
1. Preparation of model data for publication by data owner

- Dataset organisation
- File formatting
- Checksums

Metadata & Documentation

2. Dataset submission to the data management team

- Data transfer via GEOMAR cloud or (S)FTP server
- Data submission information to DM via web form
3. Data publication on data.geomar.de (OPeNDAP Service)

- Data integrity check
- Data dissemination via THREDDS-Server by data manager
OPeNDAP based server provides access to datasets of model experiments published by GEOMAR researchers either via web browser for full downloads or for OPeNDAP queries on data (sub)sets.
Example II: Access

GEOMAR THREDDS Server
THREDDS Data Server

Catalog https://data.geomar.de/thredds/20.500.12085/4e104bfe-aa9e-4d1e-99fa-5a7d1b9f9313/catalog.html

Dataset: Supplementary Data to "Decomposing barotropic transport variability in a model of the North Atlantic Ocean" (hdl:20.500.12085/4e104bfe-aa9e-4d1e-99fa-5a7d1b9f9313/psi_viking_1960_2009.nc)

- Data format: netCDF
- Data size: 202.5 Mbytes
- Data type: GRID
- Naming Authority: de.geomar
- ID: 20.500.12085/4e104bfe-aa9e-4d1e-99fa-5a7d1b9f9313/psi_viking_1960_2009.nc

Documentation:
- Decomposing barotropic transport variability in a high-resolution model of the North Atlantic Ocean (doc:10.1028/2010GC005156)
- Decomposing barotropic transport variability in a high-resolution model of the North Atlantic Ocean (http://oceanic.geomar.de/49274)
- Rights: cc_by_4.0

- Summary: The method uses a linear shallow water model to decompose the temporal variability of the barotropic streamfunction in a high-resolution model of the North Atlantic Ocean. The model configuration is VIKING200 for driving transport. The method is illustrated by examining how the recirculation region responds to the winter North Atlantic Oscillation (NAO). While no statistically significant response is found in the year overlapping the NAO cycle, the model configuration is found to be robust years 1 to 2 and when the northward momentum flux (EMF) dominates. After only 2 years, only the potential energy (PE) test play a role and is only after 5 years that the transport dependence on the NAO ceases to be significant. It is also shown that the contribution to the memory of up to 5 years in the Labrador and Icelandic seas. However, it is only around the northern rim of these seas that VIKING200 and the transport is due to mesoscale by the MFA and EMF contributions.

Access:
1. OPENDAI: https://data.geomar.de/thredds/20.500.12085/4e104bfe-aa9e-4d1e-99fa-5a7d1b9f9313/psi_viking_1960_2009.nc
2. HTTPServer: thredds/fileServer/20.500.12085/4e104bfe-aa9e-4d1e-99fa-5a7d1b9f9313/psi_viking_1960_2009.nc

Keywords:
- north atlantic
- transport variability
- high resolution model

NetCDF Attribute Convention for Dataset Discovery Report

Title: Supplementary Data to "Decomposing barotropic transport variability in a high-resolution model of the North Atlantic Ocean"

Total Score: 31/46

General File Characteristics

<table>
<thead>
<tr>
<th>Number of Global Attributes</th>
<th>Number of Variables</th>
<th>Number of Variable Attributes</th>
<th>Number of Standard Names</th>
<th>Number of Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>5</td>
<td>12</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

Access Key:

- Data Access
- Bibliographic Access
- Comprehensive Access

Other Attributes:

- Identification
- Text Search
- Other Metadata Information
- Citation Search
- Contributor Search
- Publisher Search
- Other Attributes
Example II: Further Access
Samples (in progress)

- Sample management systems for geosamples and biosamples
- Connection to OSIS for field samples as sediment cores and biological measurements and experiments
- Connection to Nagoya Protocol information necessary
- Metadata export for DAM Viewer
Example III: Samples

Access: DAM Portal

KüNO Workshop 'Data sharing in marine research', 29.9.2021
Questions and comments?

Data Management Team @GEOMAR

datamanagement@geomar.de