

# Data management workflows for campaigns and model data

**Linda Baldewein**

Helmholtz Coastal Data Center  
Institute of Carbon Cycles  
Helmholtz-Zentrum Hereon  
Geesthacht, 29.09.2021

**DAM**



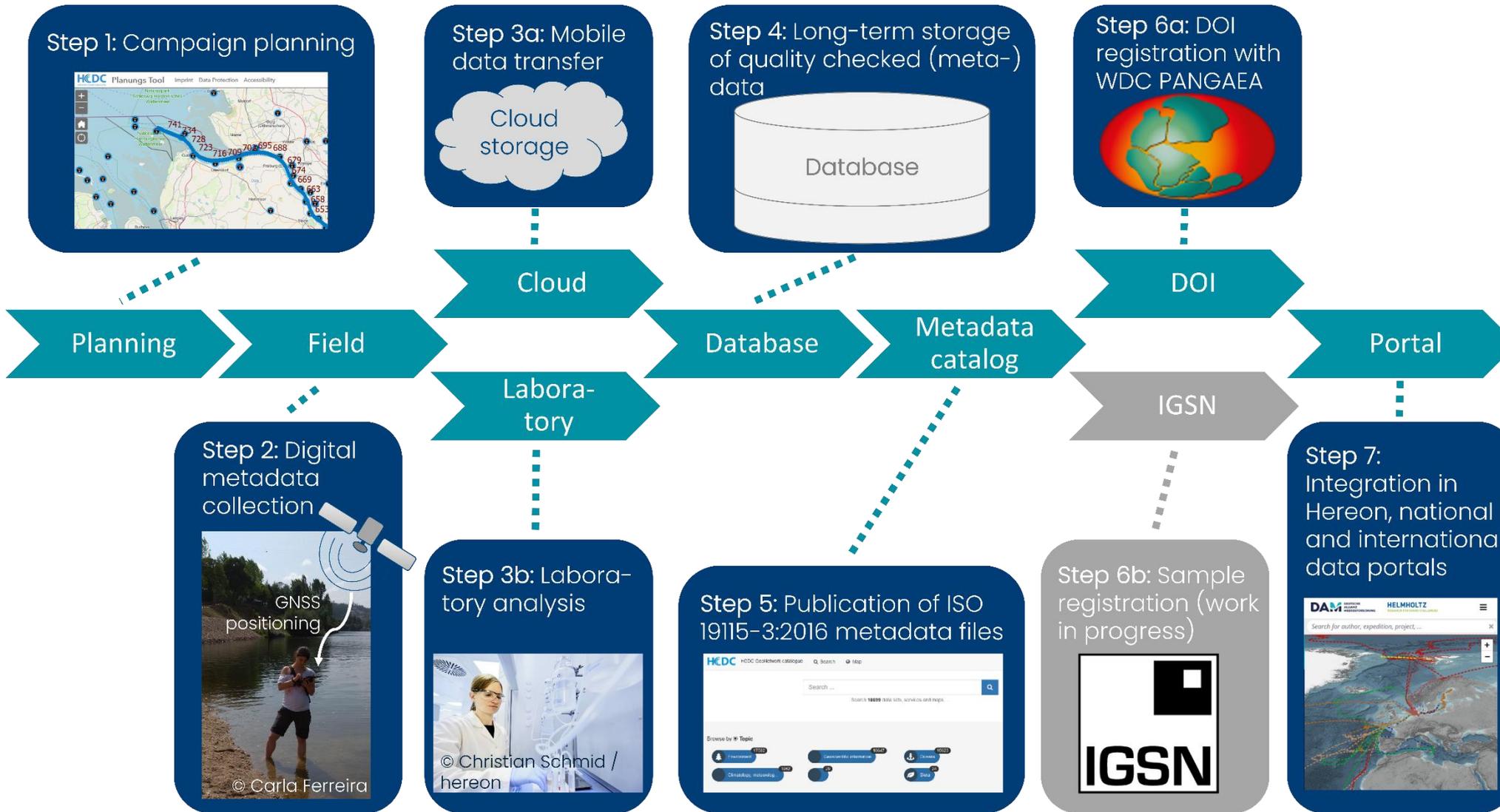
Helmholtz-Zentrum  
**hereon**

# Agenda

- Example workflow for campaigns
  - Hereon campaign data workflow
- Example workflow for models
  - Introduction to the Model Data Explorer



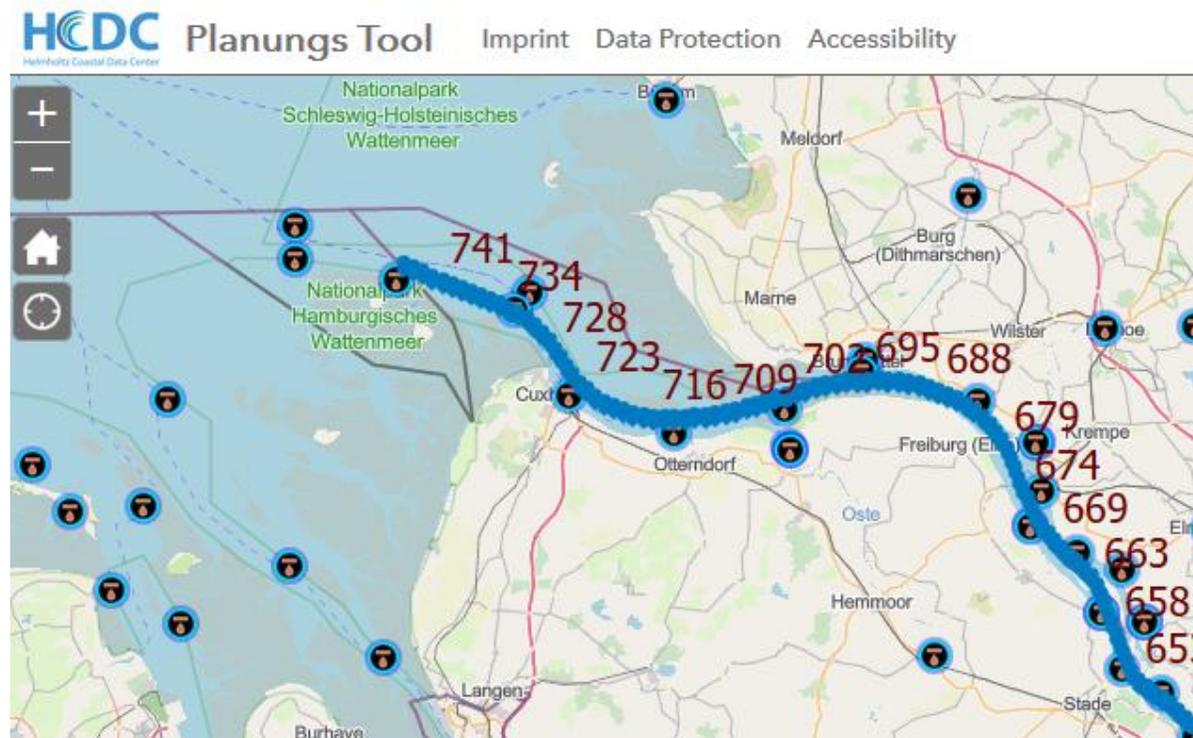
# Campaign data workflow



# Campaign data: Planning

Planning

- Where will the campaign take place?
- How many events / stations are planned?
- Create planned event / station list and / or web map
  - Most German research institutions have an ArcGIS Online subscription without you knowing.
  - QGIS as open-source Desktop alternative



# Campaign data: Planning



- How will metadata be recorded?
  - DSHIP on larger research vessels
  - Handwritten
  - Survey123 App
- Who will record metadata?

| Science Activity Number               | Science Activity Comment  | Science Activity Closed          | Science Activity Area |
|---------------------------------------|---|----------------------------------|-----------------------|
| Device Operation Number               | Device Operation  | Device Operation Comment         | Device Operation      |
| Closed Device Operation Label         | Device Shortname  | Device Comment                   | Underway              |
| Event Time                            | Event   | Latitude (deg)                   | Longitude             |
| 0                                     | f 1 AL557_0_Underway-1 t TSG TSG System permanently installed aboard RV |                                  |                       |
| ALKOR                                 | 04/06/2021 06:54 profile start  | "53° 57,841' N" "008° 37,109' E" | 14 0.7                |
| 312.6                                 | 53.964012 8.618485 101 2.8  |                                  |                       |
| 1                                     | t 1 AL557_1-1 t CTD water CTD water                                     | f 04/06/2021 07:09               | in                    |
| the water                             | "53° 57,749' N" "008° 37,225' E"  | 15 1 117.1 53.962488 8.620417    | 89.4 2.4              |
| 1                                     | t 1 AL557_1-1 t CTD water CTD water                                     | f 04/06/2021 07:19               | on                    |
| deck "53° 57,661' N" "008° 37,406' E" | 16 0.6 143.2 53.96101 8.623432  | 124.9 1.7                        |                       |
| 2                                     | t 1 AL557_2-1 t CTD water CTD water                                     | f 04/06/2021 09:08               | in                    |
| the water                             | "54° 00,007' N" "008° 05,928' E"  | 24 0.2 62.6 54.000112 8.098808   | 163.8 1.9             |
| 2                                     | t 1 AL557_2-1 t CTD water CTD water                                     | f 04/06/2021 09:16               | on                    |
| deck "54° 00,020' N" "008° 05,939' E" | 24 0.1 324.2 54.000335 8.098978   | 175.3 1.6                        |                       |
| 3                                     | t 1 AL557_3-1 t CTD water CTD water                                     | f 04/06/2021 10:28               | in                    |
| the water                             | "54° 03,665' N" "008° 01,021' E"  | 27 0.4 265.5 54.06108 8.01701    | 235.5 0.8             |
| 3                                     | t 1 AL557_3-1 t CTD water CTD water                                     | f 04/06/2021 10:37               | on                    |
| deck "54° 03,674' N" "008° 00,917' E" | 27 0.1 283.8 54.061235 8.01528  | 258 1.2                          |                       |
| 3                                     | t 2 AL557_3-2 MUC t GC GC Corer penetrating seabed driven by gravity    | f                                |                       |
| 04/06/2021 10:43                      | in the water  | "54° 03,673' N" "008° 00,922' E" | 27 0 283.8            |
| 54.061222                             | 8.015373 290.2 0.8  |                                  |                       |

# Campaign data: Collecting

## Required information

- Metadata of your campaign
  - Ship name
  - Station / Event name
  - Coordinates
  - Sample information
  - Participants

-> Ideally digitally collected using DSHIP or Survey123



# Campaign data: Processing and Analysing

- Metadata of your measurement
  - Parameter name
  - Unit
  - Method
  - Quality information
  - Origin of sample

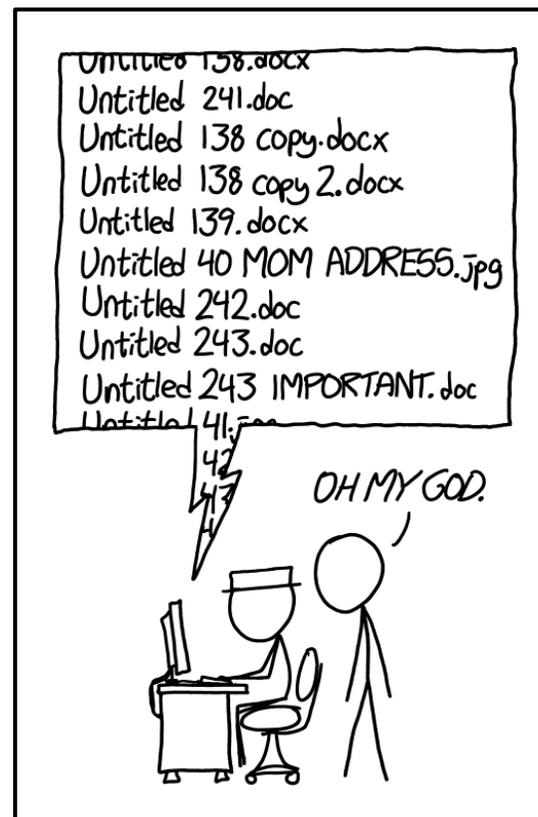
Processing  
and  
Analysing



© Christian Schmid / Hereon

# Campaign data: Preserving

- Store results locally
  - Meaningful file names
  - Shared storage with your working group
  - Include metadata
- Prepare for submission
  - Format based on requirements of repository
  - Use template, if available
  - Double check for typos, etc.
- Submit to institution repository (if available)



PRO TIP: NEVER LOOK IN SOMEONE ELSE'S DOCUMENTS FOLDER.

Randall Munroe, 2014, CC BY-NC 2.5

# Campaign data: Publishing and Sharing



- Publish your data
  - DOI registration with PANGAEA
  - See details in talk by Flavia Höring
- Publish your samples
  - IGSN
- Your local data curators may help you. Ask them!



Heinrich Heine  
POTSDAM



Physical Sample



Research Center Potsdam  
GFZ GERMAN RESEARCH CENTRE  
FOR GEOSCIENCES

| General Identifiers     |  |
|-------------------------|--|
| Project:                | Project 1 and 11   |
| Campaign:               | Litter harvest Exp 2017  |
| Type:                   | Specimen   |
| Name:                   | Senna cumingi  |
| IGSN:                   | GFRCM001G  |
| Parent IGSN:            | N/A  |
| Release Date:           | N/A  |
| Sampling Location       |  |
| Latitude:               | -30.053975   |
| Longitude:              | -71.0950333  |
| Coordinate System:      | WGS84  |
| Elevation:              | 629.0  |
| Location Type:          | N/A  |
| Location Name:          | QT Plot 1  |
| Location Description:   | 10x10 m Plot, established by Rafaela Canessa in April 2016. The study site is called Quebrada de Talca, and corresponds to a private conservation project, with past cattle grazing use and subsequent animal exclusion (for 7 years in 2016). Quebrada de Talca presents similar vegetation, parent material and climatic conditions to Santa Grecia, but thanks to the exclusion project, the vegetation grows better and has no current animal influence. |
| Country:                | Chile  |
| Province:               | Coquimbo Region  |
| County:                 | N/A  |
| City:                   | N/A  |
| Acquisition             |  |
| Material:               | Biology  |
| Biology Classification: | vegetation:leaf litter   |
| Biology Description:    | 5 glasses with milled litter Nr: 725, 726, 727, 728, 729.  |
| Collection Method:      | manual   |
| Funding Agency:         |  |

**Sample Family**

▶ Senna cumingi

Specimen

The Sample Family shows a sub-sampling graph. Select entries to navigate samples.

**Location Map**



Drilling Start/End: 2017-01-17 / 2017-01-17 \*  
Latitude: -30.05398 \* Longitude: -71.09503 \*  
QT Plot 1



**PANGAEA.**  
Data Publisher for Earth & Environmental Science

SEARCH SUBMIT ABOUT CONTACT

**Citation:**

**Apel, Christina; Joerss, Hanna; Ebinghaus, Ralf (2018):** Organic UV stabilizers and UV filters in the sediment of European North and Baltic Seas in 2016/2017. *PANGAEA*, <https://doi.org/10.1594/PANGAEA.895397>,

*Supplement to:* Apel, C et al. (2018): Environmental occurrence and hazard of organic UV stabilizers and UV filters in the sediment of European North and Baltic Seas. *Chemosphere*, **212**, 254-261, <https://doi.org/10.1016/j.chemosphere.2018.08.105>

Always quote above citation when using data! You can download the citation in several formats below.

[RIS Citation](#) [Bibtex Citation](#) [Copy Citation](#) [Facebook](#) [Twitter](#) [Show Map](#) [Google Earth](#)

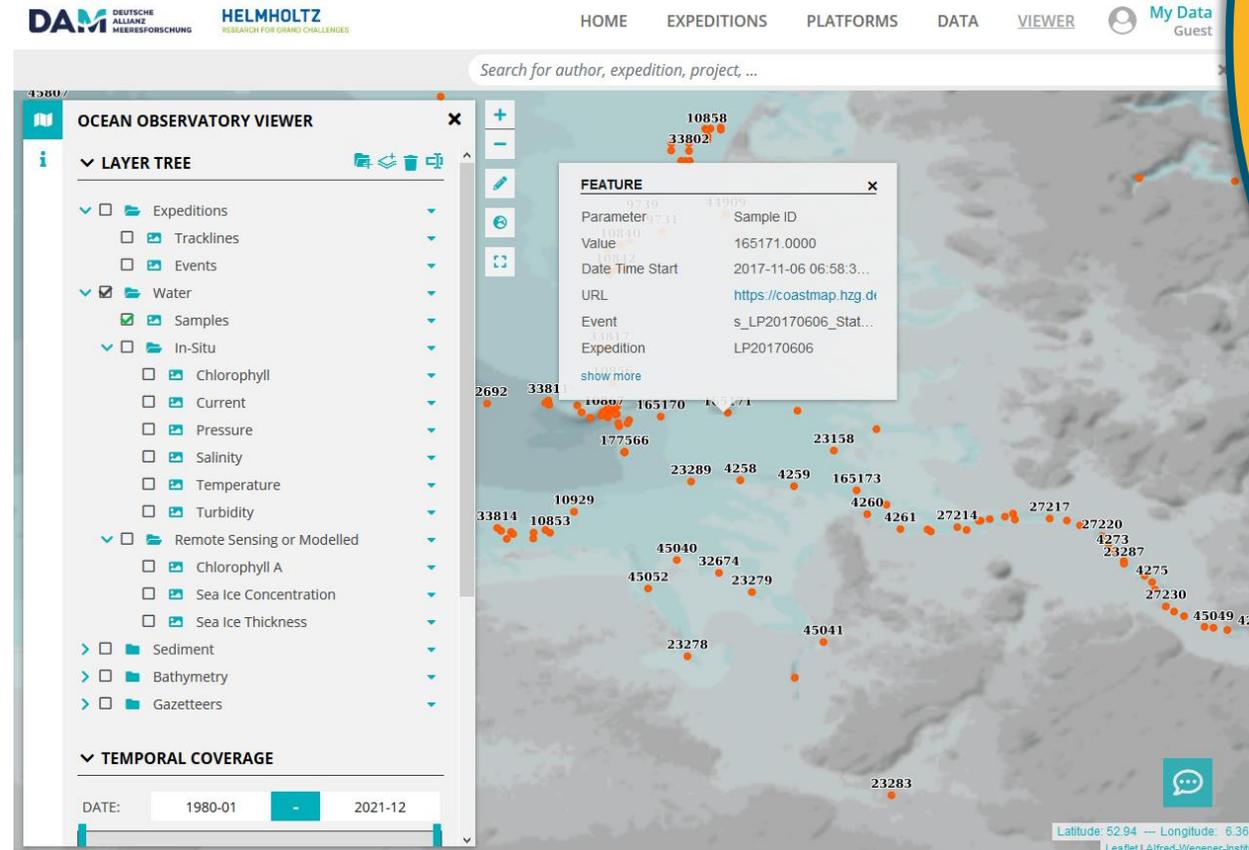


**Abstract:** UV absorbing compounds are of emerging concern due to their large production volumes, their persistence or pseudo-persistence, and their potential adverse effects. This is the first study investigating the environmental occurrence and potential hazard of organic UV stabilizers and UV filters in the North and Baltic Sea surface sediments, including the connecting Skagerrak and Kattegat straits. In total, nineteen substances were identified over the entire study area, including the rarely studied compounds ethylhexyl triazone (EHT) and bisoctrizole (UV-360). Octocrylene (OC) was the predominant compound in this study with regard to detection frequency (79%) and concentrations (up to 9.7 ng/g dwt). OC occurred frequently (86% of UV stabilizer concentrations) in the German Bight. The relative detection of EHT was identified in the



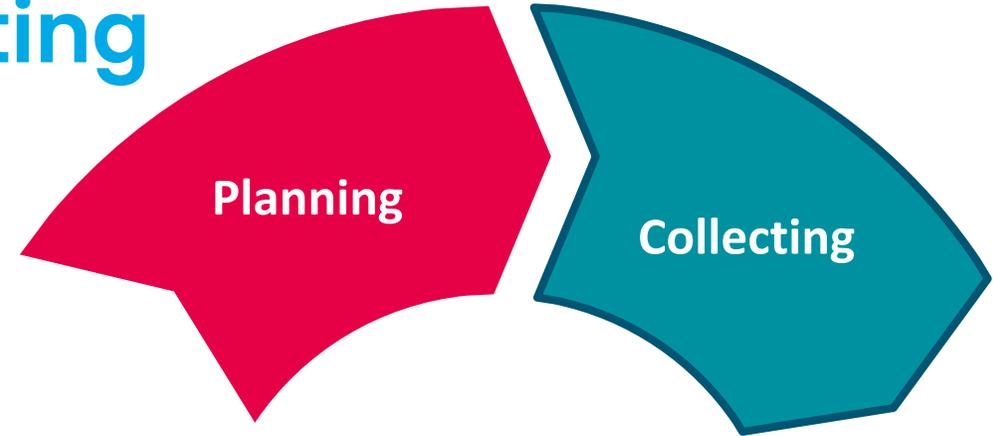
# Campaign data: Reusing

- Data available in data portals
- Other scientists can find, download and cite your data



# Model data: Planning and Collecting

- Which model will you use / develop?
- Gather data
  - Reuse data published by other scientists



PANGAEA.  
Data Publisher for Earth & Environmental Science

SEARCH SUBMIT HELP ABOUT CONTACT

Submit Data

Welcome to PANGAEA® Data Publisher

Our services are generally open for archiving, publishing, and re-usage of data. The World Data Center PANGAEA is member of the World Data System.

ALL TOPICS

Search for measurement type, author name, project, taxa,...

LITHOSPHERE (132338)

BIOLOGICAL CLASSIFICATION (33408)

ATMOSPHERE (29860)

PALEONTOLOGY (23637)

ECOLOGY (18051)

LAND SURFACE (7811)

BIOSPHERE (4290)

GEOPHYSICS (3831)

WDC CLIMATE

Home FAIRness Documentation Utilities Register Login

DKRZ

DKRZ long term archive

The German Climate Computing Center (DKRZ: Deutsches Klimarechenzentrum GmbH) provides a Long Term Archiving Service for large research data sets which are relevant for climate or Earth system research. This service includes archiving and retrieval capability of data for time periods of 10 years or longer. The long-term archive (LTA) of DKRZ is certified according to the criteria of the Core Trust Seal (CTS) and is, as World Data Centre for Climate (WDCC), accredited as regular member of the World Data System.

Find your data

Search ...

Services

Data access

Regarding the FAIR principles all metadata is openly accessible. The data download is free for

Data archiving

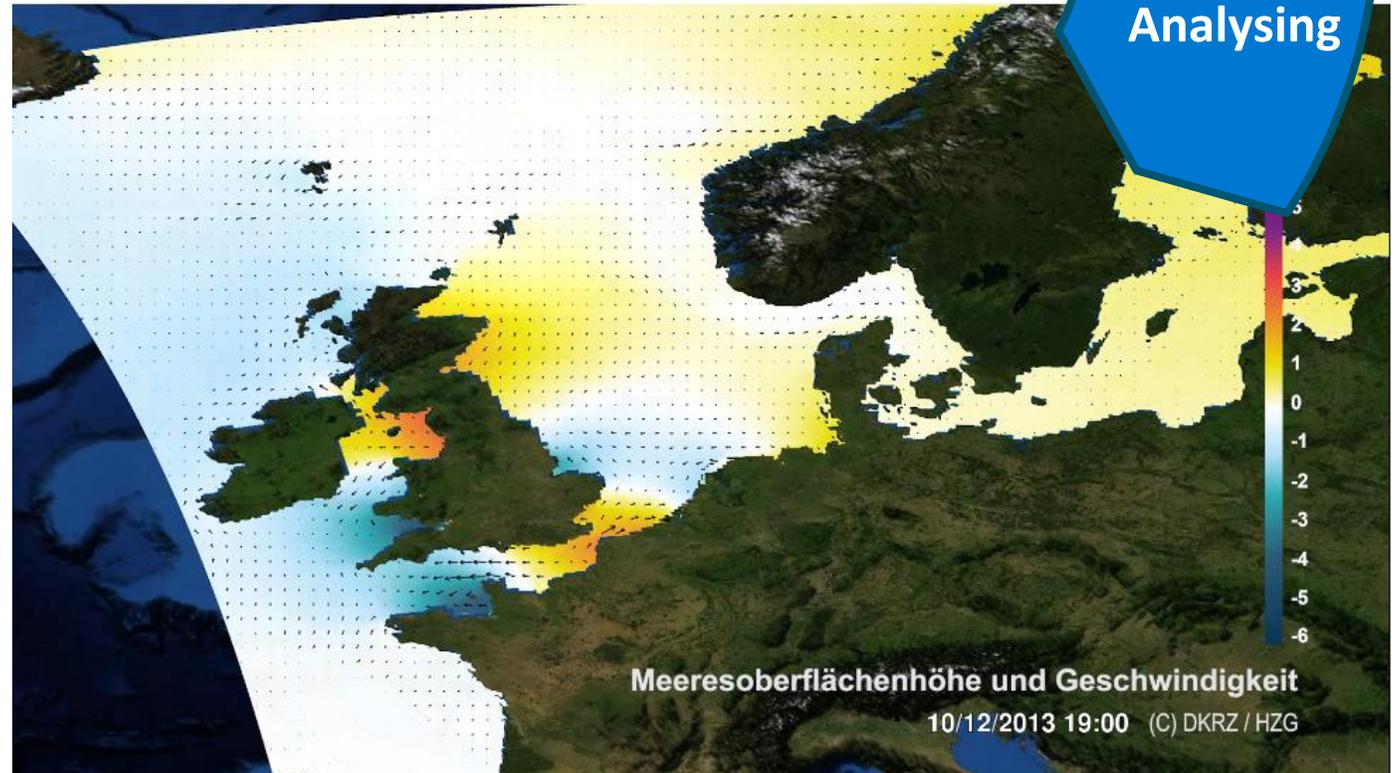
The DKRZ provides two different tiers of long-term archiving:

Data publication

The DKRZ offers a DataCite data publication service for WDCC long-term archived data.

# Model data: Processing and Analysing

- Metadata of your model
  - Conventions
  - Institution, originator, contact
  - Title
  - Source
  - Creation date
  - Coordinate reference system
  - Dimensions
  - Variables
  - Units
  - Standard names



# Model data: Preserving



- Store results locally
  - Meaningful file names
  - Shared storage with your working group
  - Include metadata
- Prepare for submission
  - Format based on requirements of repository
  - Use template, if available
  - Double check for typos, etc.
- General rules for data storage in netCDF files compiled at Hereon

|   |                           |   |
|---|---------------------------|---|
| <br>Helmholtz-Zentrum<br><b>hereon</b> | Hereon<br>Data Management | Name: <b>Binding Regulations for Storing Data as netCDF Files</b><br>Date: 18.05.2021 |
|---|---------------------------|---|

## 2. General Specifications

Provided here is general information that has nothing directly to do with the netCDF format itself but should, however, be taken into consideration.

### 2.1. File Names

- NetCDF files are designated with the extension .nc
- The file name begins with a letter or a letter sequence. The letter sequence should allow users to draw conclusions about the type of data found in the file.
- Customary sequences used thus far (e.g., "ctd" for CTD data, "sf" for ScanFish data) will be retained. Specification should still be designated for other devices.
- When dealing with model data, either the model name with the version designation or the ExperimentID should be used in the first segment of the filename.

Spaces, umlauts, "?", special characters (except underscore "\_", dash "-", and periods) are not

# Model data: Publishing and Sharing



- Publish your data
  - DOI registration with WDC CERA at DKRZ
- Your local data curators may help you. Ask them!

The screenshot shows the WDC CLIMATE website. The header is dark blue with the WDC CLIMATE logo on the left and navigation links: Home, FAIRness, Documentation, Utilities, Register, and Login. The main content area features a satellite image of Earth. On the left, there is a section titled 'DKRZ long term archive' with a 'CORE TRUST SEAL' logo and a paragraph of text. On the right, there is a search bar with the text 'Find your data' and a search icon.

## Services

**Data access**

Regarding the **FAIR** principles all metadata is openly accessible. The data download is free for

**Data archiving**

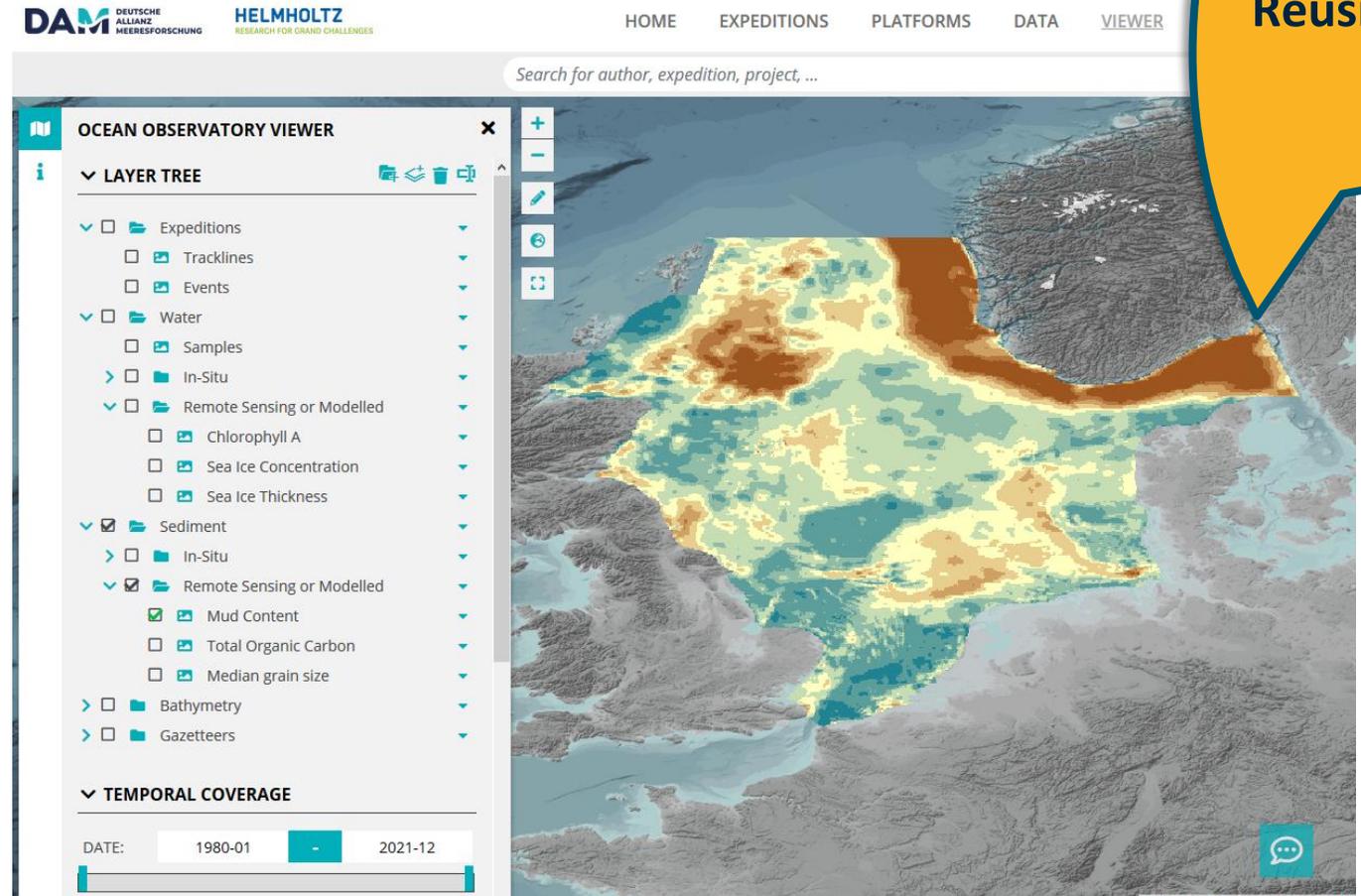
The DKRZ provides two different tiers of long-term archiving:

**Data publication**

The DKRZ offers a **DataCite** data publication service for WDCC long-term archived data.

# Model data: Reusing

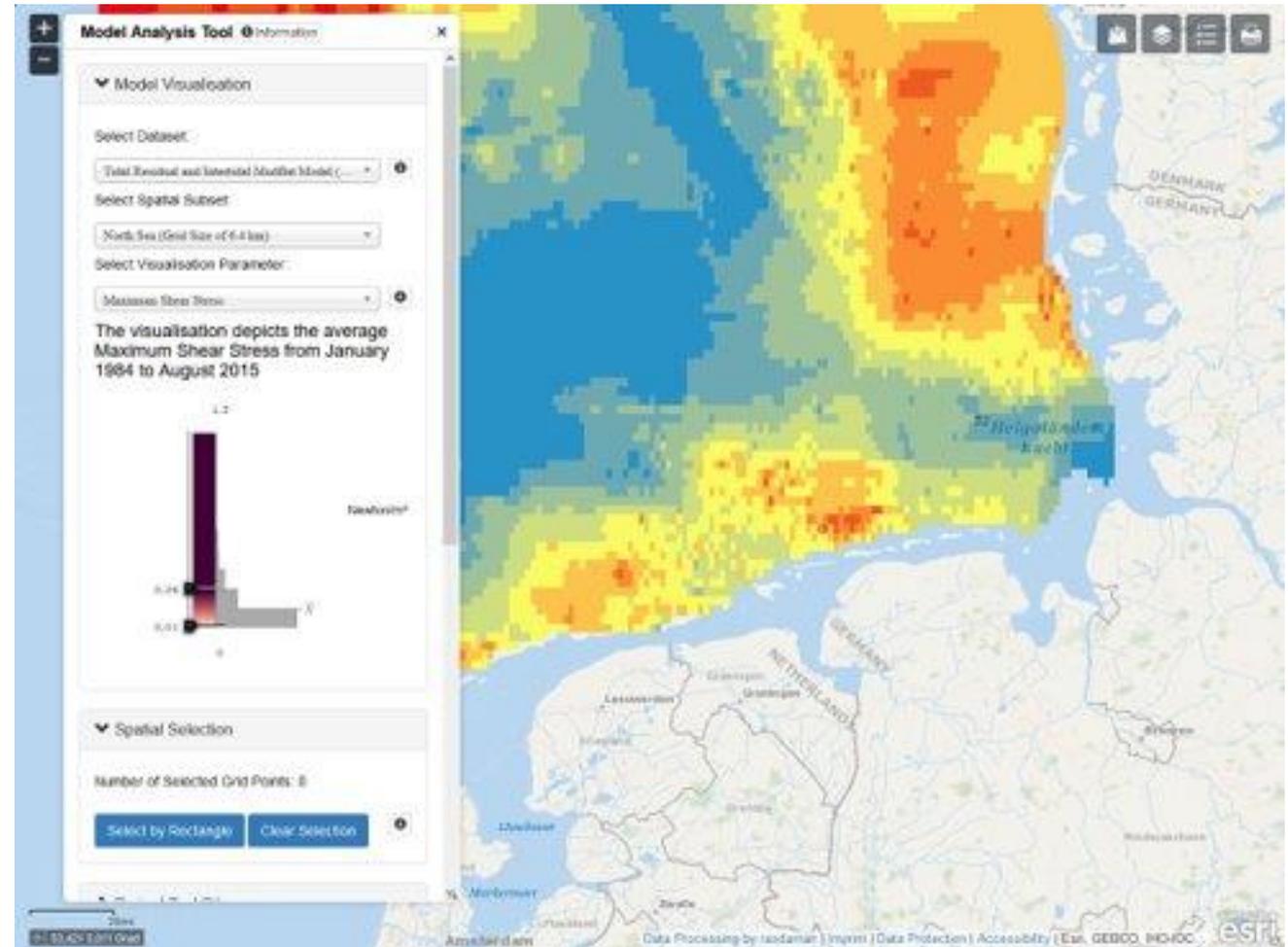
- Other scientists can find, download and cite your data
- Data available in data portals
  - Currently a lot of manual effort by curators



Reusing

# Model Data Explorer

- Central platform to access Model Data
  - 4D Model Data Map
  - Search and filter datasets
  - Compute and compare statistics on the data
  - Download raw data
- Currently in development
  - We need support from future users
  - Join our mailing list [mde-dev@listserv.dfn.de](mailto:mde-dev@listserv.dfn.de)
  - Check out the [development](#)



# Vielen Dank.

[www.hereon.de](http://www.hereon.de)

