Data management workflows for campaigns and model data

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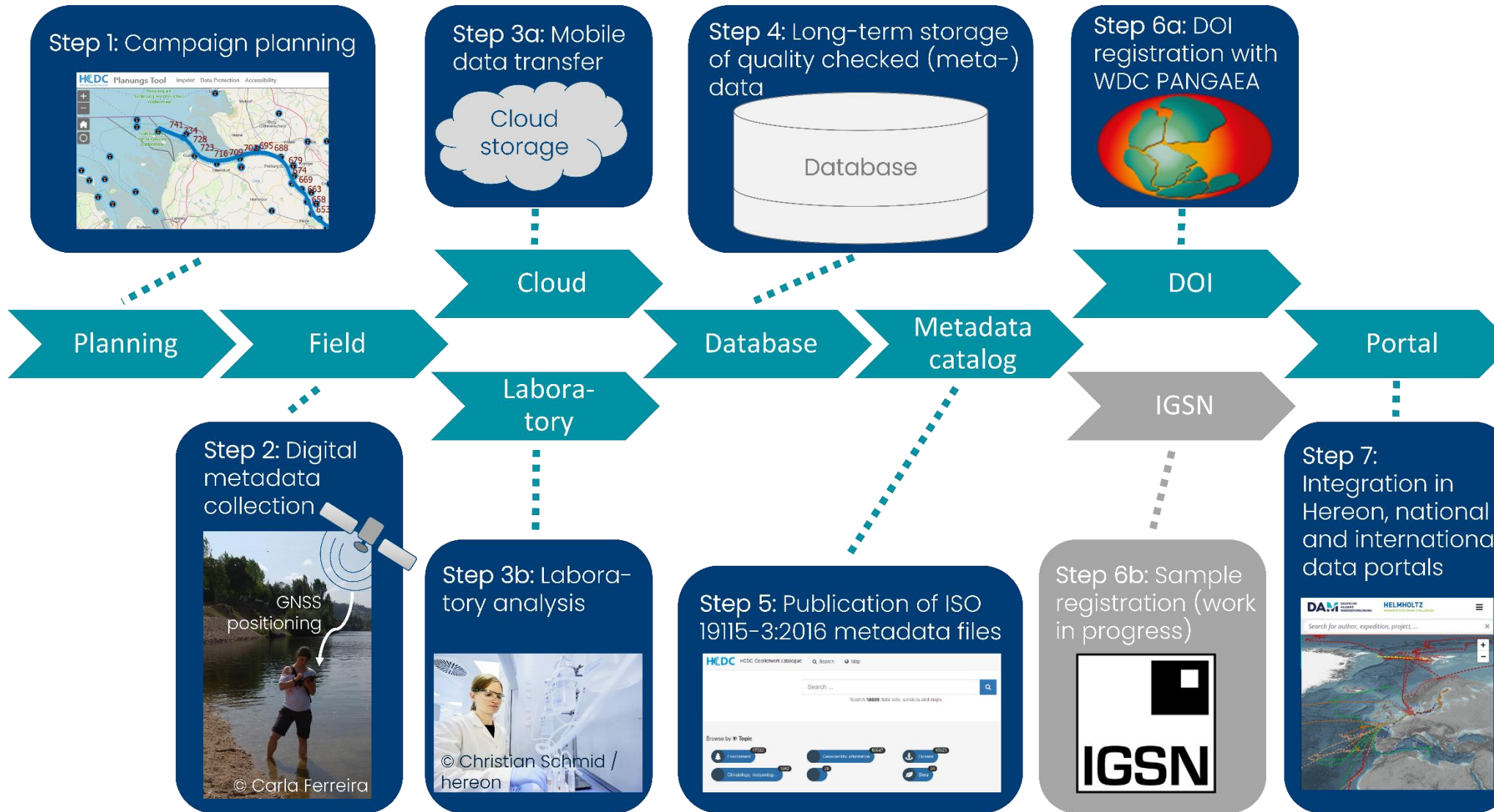Geesthacht, 29.09.2021

DAM

Helmholtz-Zentrum
hereon

Agenda

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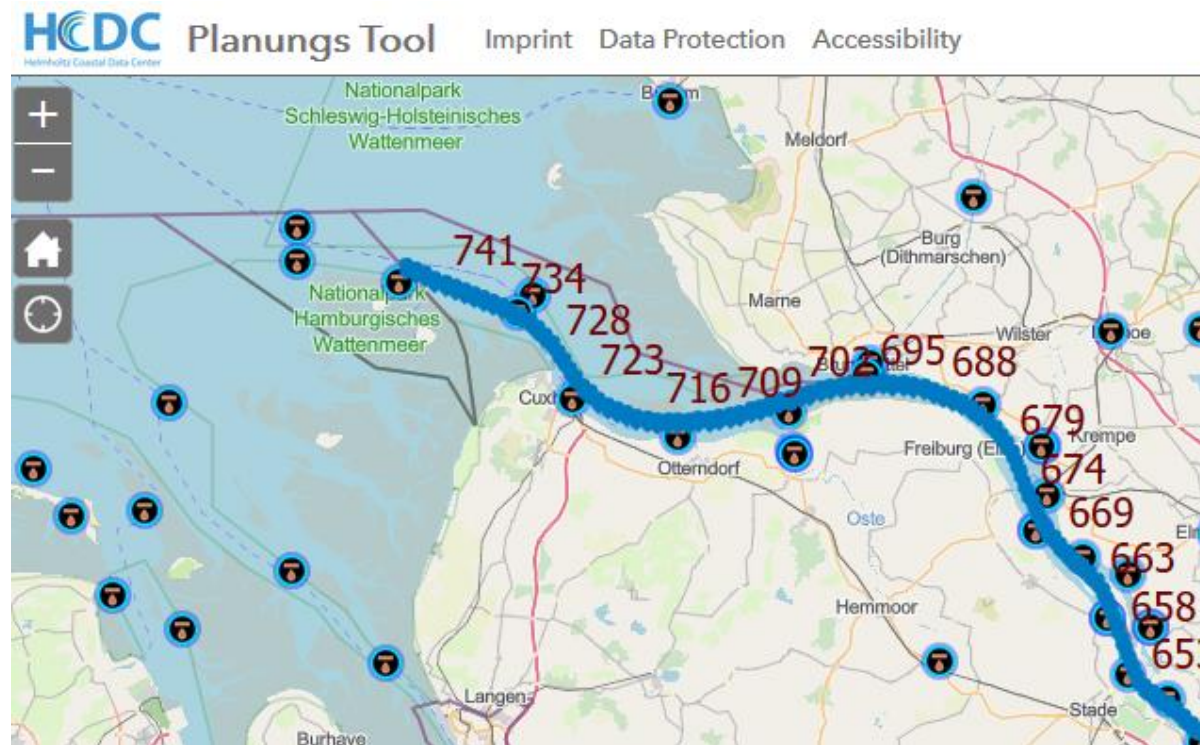 - Introduction to the Model Data Explorer

29.09.2021   Data workflows for campaigns and model data

Campaign data workflow

Step 1: Campaign planning

Step 3a: Mobile data transfer

Cloud storage

Step 4: Long-term storage of quality checked (meta-) data

Database

Step 6a: DOI registration with WDC PANGAEA

Cloud

DOI

Planning → Field → Database → Metadata catalog → Portal

Labora-tory

IGSN

Step 2: Digital metadata collection

GNSS positioning

© Carla Ferreira

Step 3b: Labora-tory analysis

© Christian Schmid / hereon

Step 5: Publication of ISO 19115-3:2016 metadata files

Step 6b: Sample registration (work in progress)

IGSN

Step 7: Integration in Hereon, national and international data portals

DAM

Helmholtz-Zentrum
hereon

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
Comment    Device Operation Number Device Operation    Device Operation Comment    Device Operation
Closed Device Operation Label  Device Shortname   Device Comment  Underway    Event Time Event
Comment    Action  Latitude    Longitude   Depth (m)   Speed (kn)  Course  Latitude (deg)  Longitude
(deg) Wind Dir    Wind Velocity
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

Collecting

© Carla Ferreira

© Ebeling / Hereon

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)

Preserving
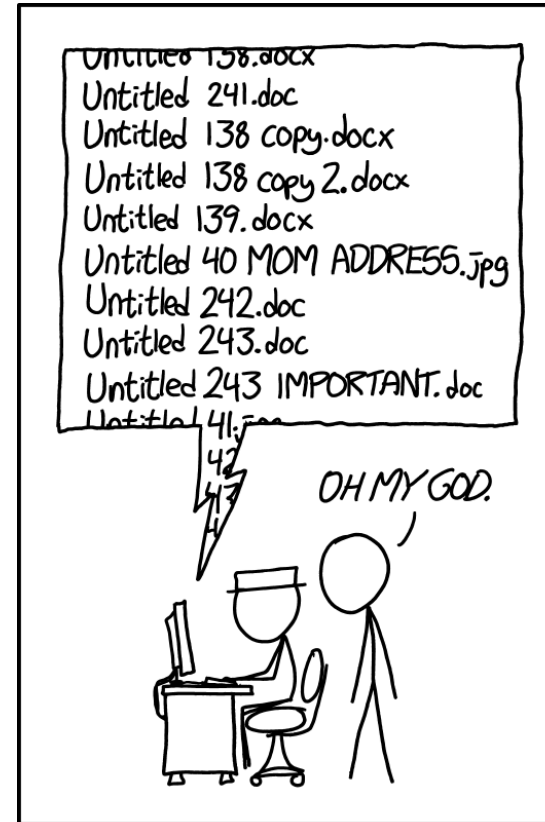
Randall Munroe, 2014, CC BY-NC 2.5

Campaign data: Publishing and Sharing

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!

Campaign data: Reusing

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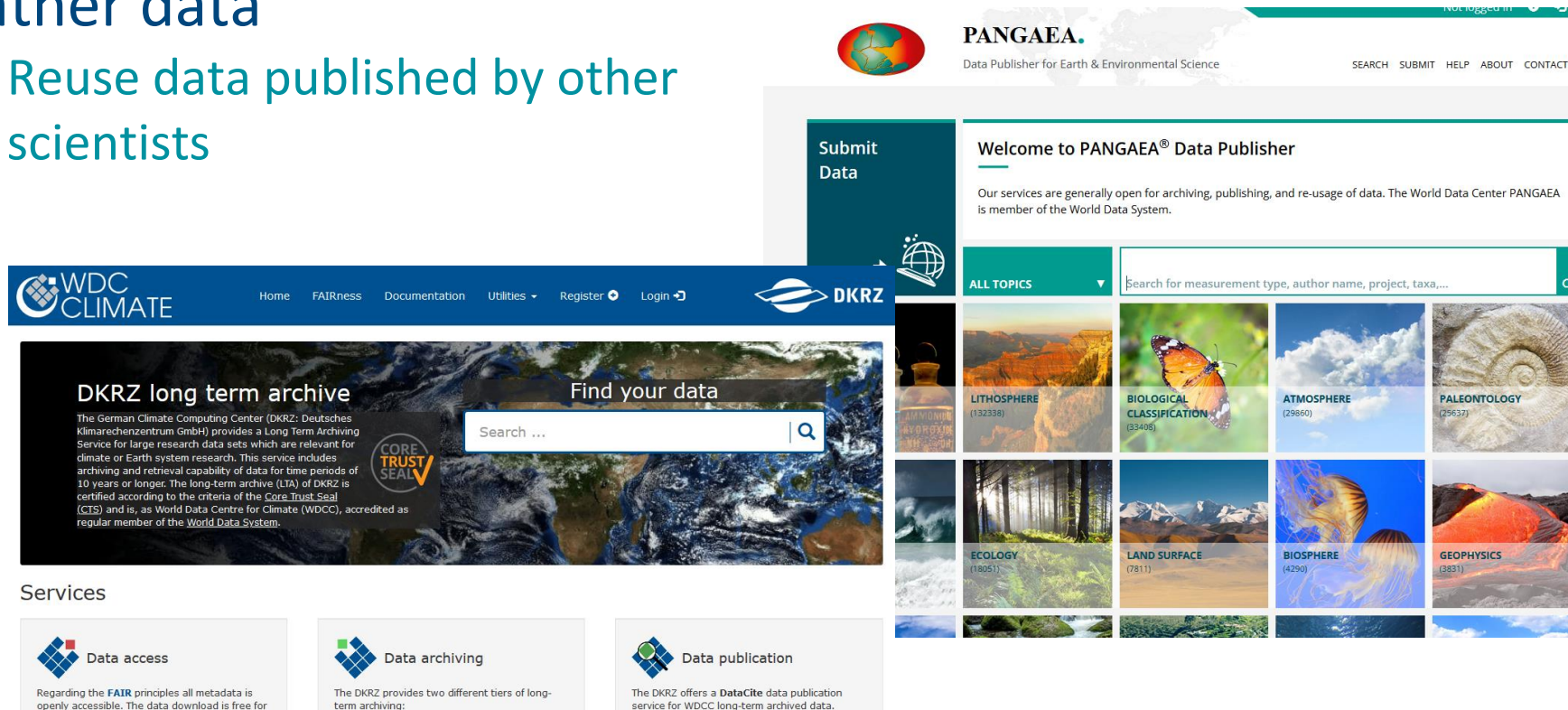- Other scientists can find, download and cite your data
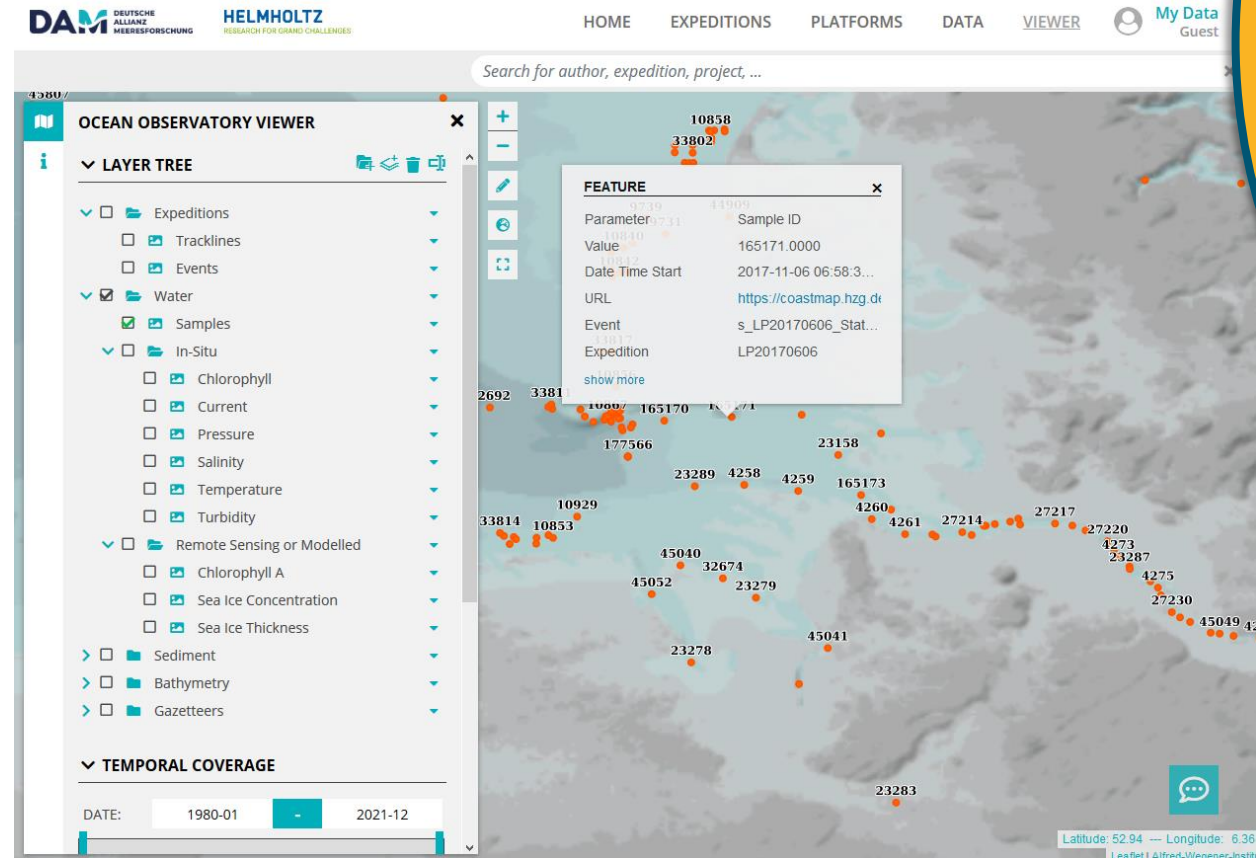
Reusing

Model data: Planning and Collecting

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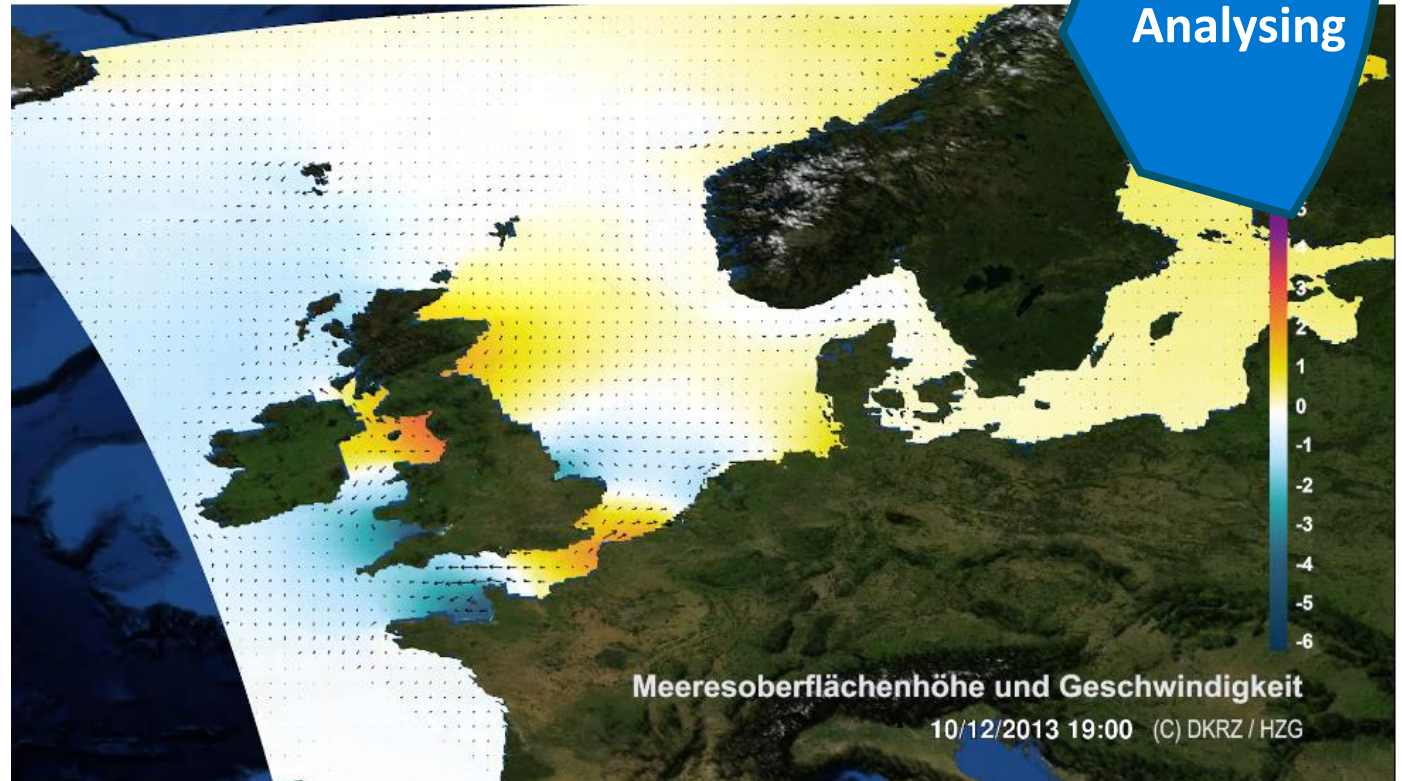

Data workflows for campaigns and model data

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

Processing and Analysing

Meeresoberflächenhöhe und Geschwindigkeit
10/12/2013 19:00   (C) DKRZ / HZG

Model data: Preserving

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.
- <u>General rules</u> for data storage in netCDF files compiled at Hereon

| Helmholtz-Zentrum hereon	Hereon Data Management	Name: **Binding Regulations for Storing Data as netCDF Files** 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, "ß", special characters (except underscore "_", dash "-" and periods) are not

Model data: Publishing and Sharing

Publishing

and

Sharing

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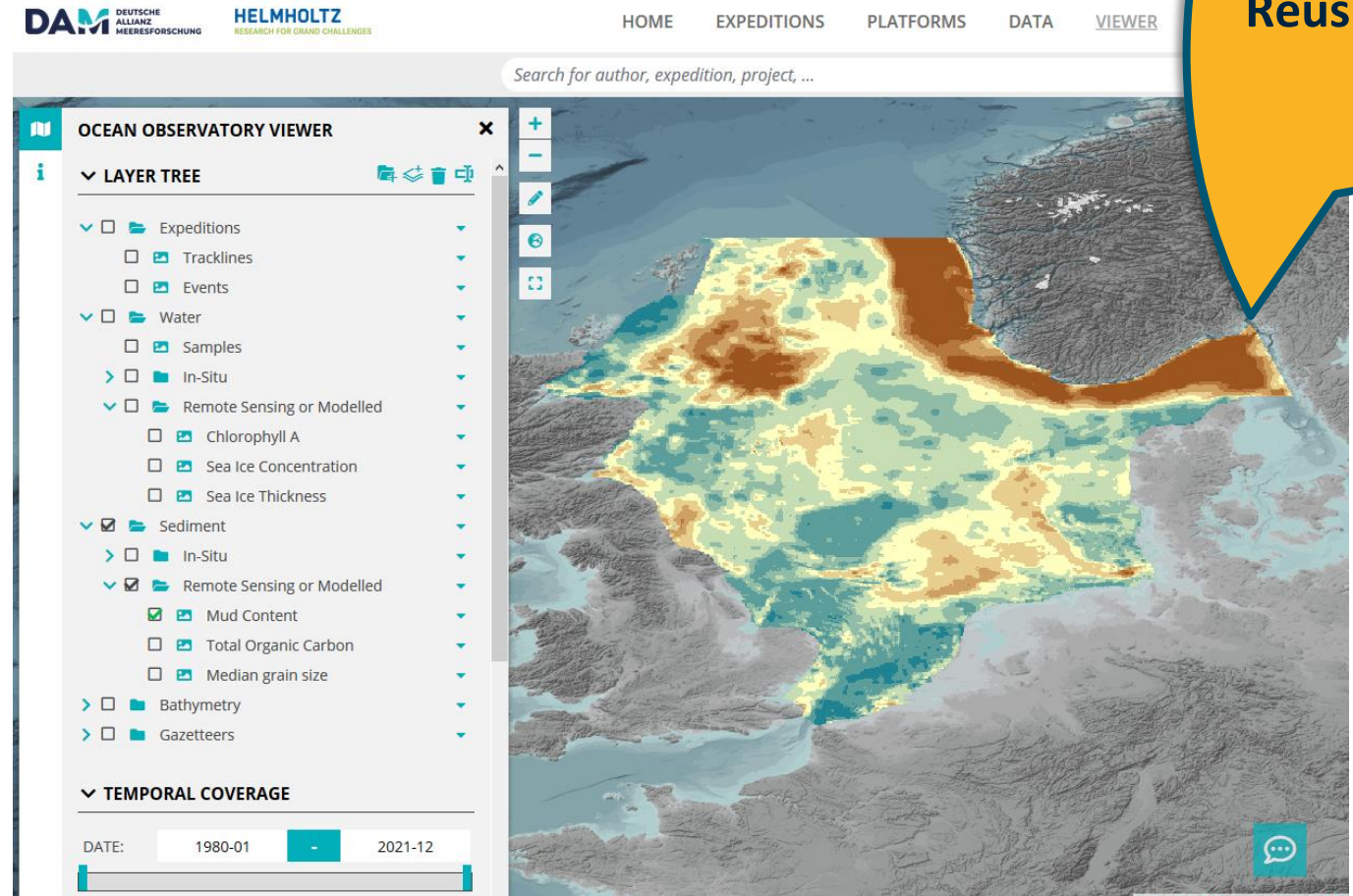- Your local data curators may help you. Ask them!
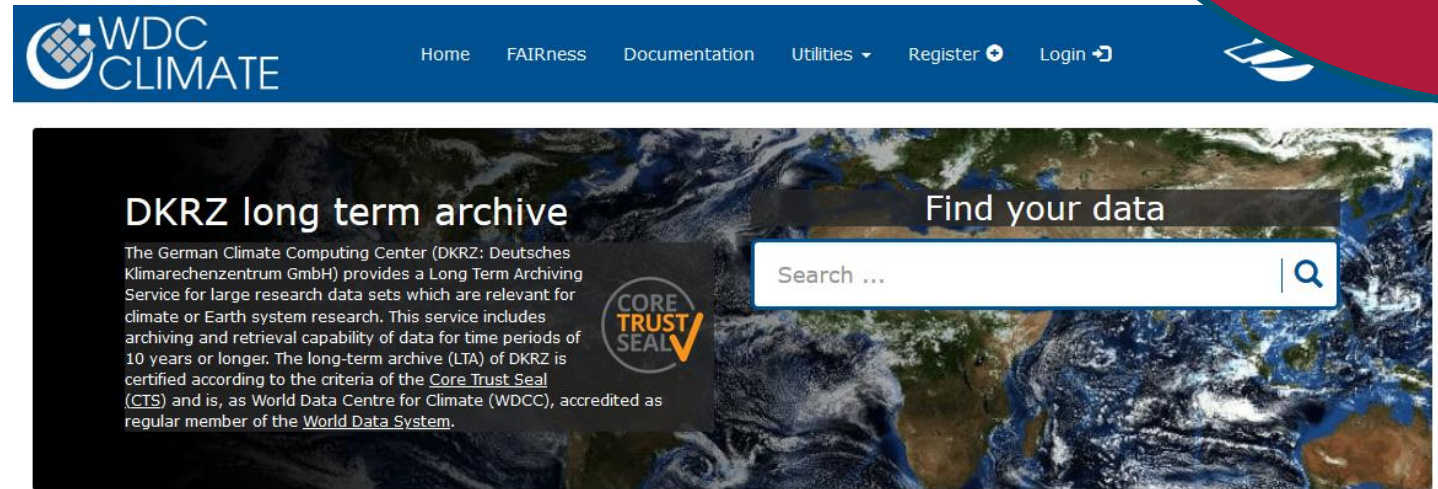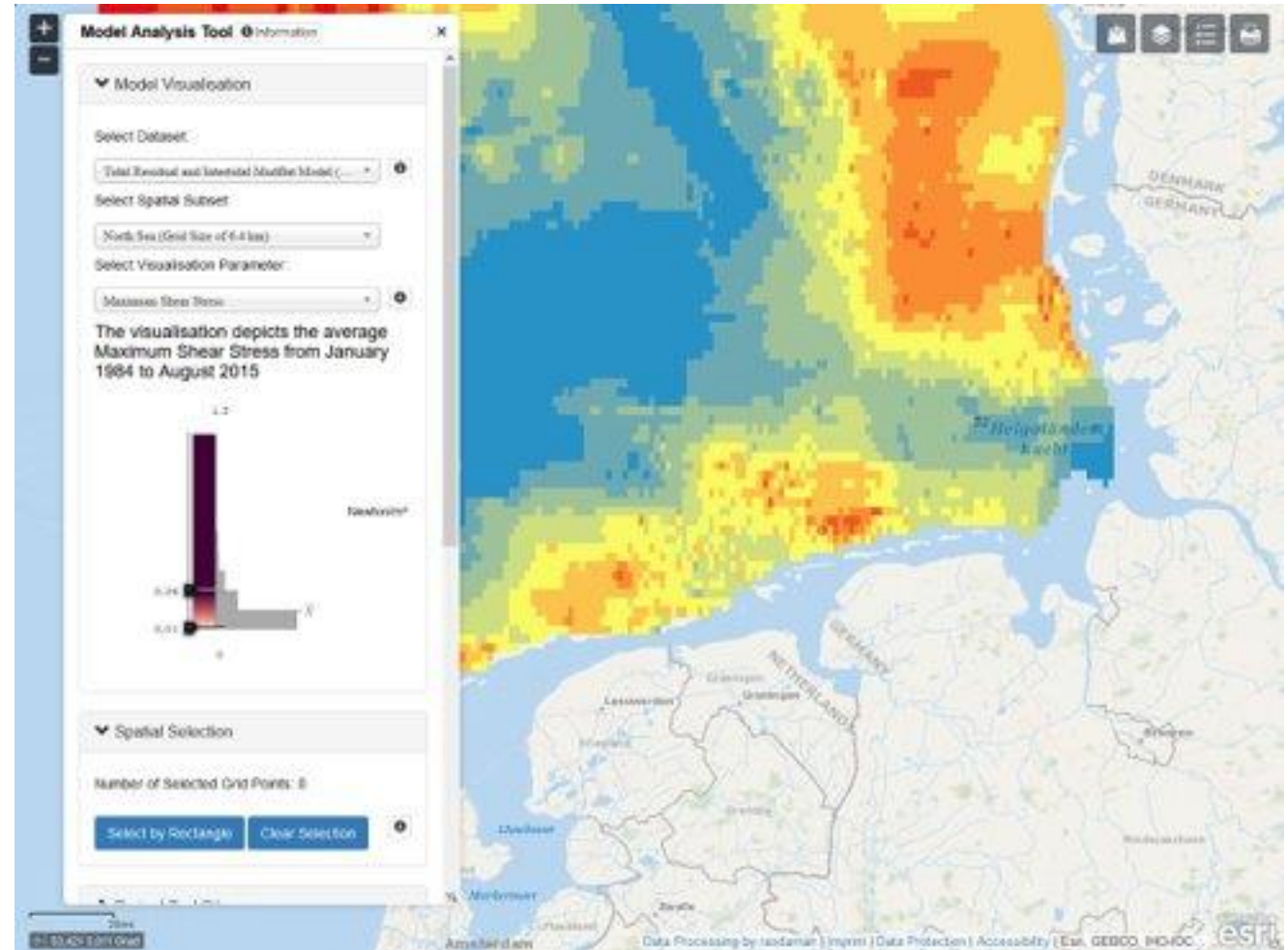
Model data: Reusing

- 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
  - Check out the development

Vielen Dank.

Helmholtz-Zentrum
hereon