GEOMAR
Helmholtz-Zentrum für Ozeanforschung Kiel

GEOMAR Guide
for
Handling Supplementary Data

Contact

Data Management

Adress: GEOMAR Helmholtz Centre for Ocean Research Kiel
Wischhofstr. 1-3
24148 Kiel | Germany

Phone: 0431 / 600 2294

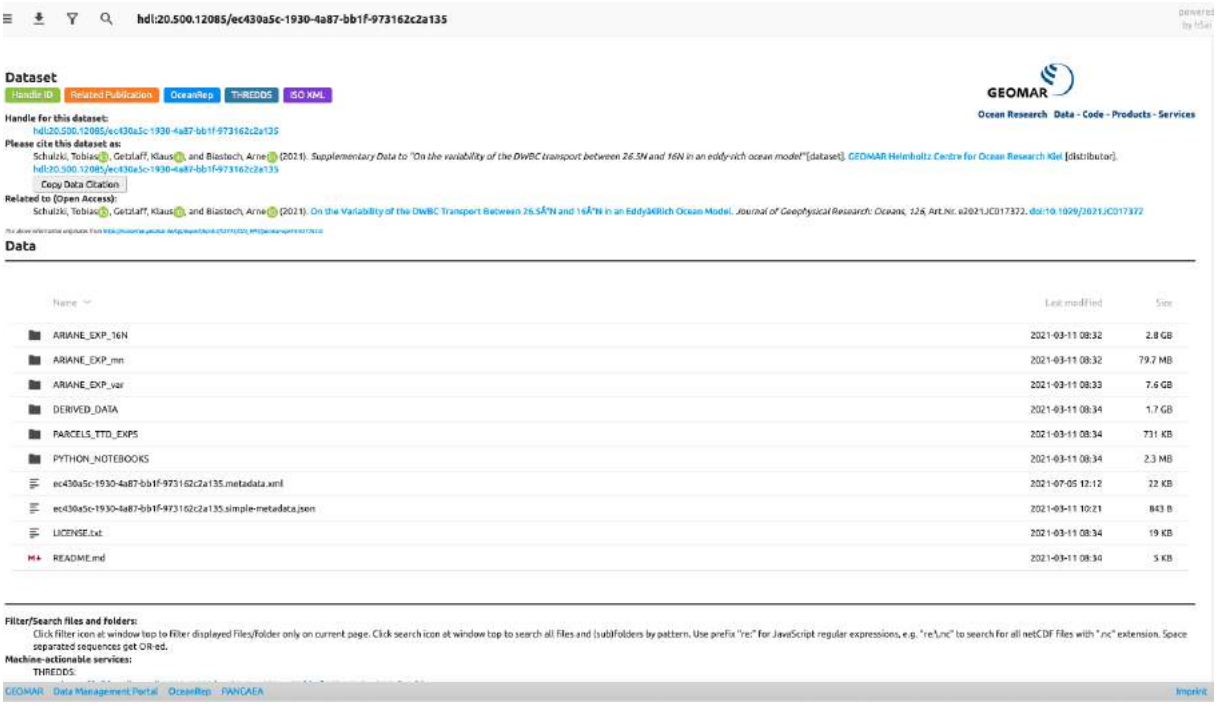
E-Mail: datamanagement@geomar.de

CONTENT

1. Which data access services are available

The DRS team support the sharing and publication of research data from disciplines that do not fit into a specific repository (like e.g. PANGAEA). We will provide a **unique handle as a Persistent IDentifier** (PID) for your data which will always redirect users to the proper site at GEOMAR services providing access to data and metadata! GEOMAR researchers have the opportunity to foster the FAIR process and comply with its principles by taking advantage of the GEOMAR Data Management support.
The official printable citation of such a handle is for e.g.:
hdl:20.500.12085/ec430a5c-1930-4a87-bb1f-973162c2a135
which results then in a clickable Link/URL:
https://hdl.handle.net/20.500.12085/ec430a5c-1930-4a87-bb1f-973162c2a135

Following the handle link leads first to a **landing page where your data is accessible for download.** Data-centric services may be available where applicable and are then shown on this page as well (e.g. WMS, THREDDS,…).



Hereby you will have the following choices for sharing your data:

a. **Open Access** - listed in public overview pages - full visibility e.g. data.geomar.de
b. **Access only for collaboration partners** in ongoing research projects - no visibility in any catalog or web page - you need to know the URL
c. **Access only for reviewers until publication** of peer reviewed articles / data - no visibility in any catalog or web page - they need to know the URL

The datasets related to journal articles will be interlinked in OceanRep GEOMAR for better findability.

NOTICE: Please **use always the CITATION variant of your handle** when citing your data!

The syntax for the displayed/printed **CITATION variant** is for example:
hdl:20.500.12085/4e104bfe-aa9e-4d1e-99fa-5a7d1b9f9313

The URL syntax for the clickable **HTML variant** is then the following:
https://hdl.handle.net/20.500.12085/ec430a5c-1930-4a87-bb1f-973162c2a135

2. What you need to do

2.1 Prepare your data for sharing or publication

Before you request a handle for your data you have to prepare your datasets by taking in account the following points.

2.1.1 Prepare metadata and documentation

Complete metadata are crucial for long-term archiving and subsequent data reusage.

1. We recommend you to set global attributes and variable standard names in NetCDF files following the **NetCDF Climate and Forecast (CF) Metadata Convention** (http://cfconventions.org/) (recommended standard by Unidata). Information about the CF are well summarized by CEDA: https://help.ceda.ac.uk/article/4507-the-cf-metadata-convention.

   (There is also a CF-Checker to test your files for compliance with the CF Convention: Visit the CF-checker website to use this service or alternatively download and install the CF-checker software.)

2. Create a **README file** for your dataset.

2.1.2 Structure your dataset and prepare your files

1. **Label your files, directories and data** for making reuse and proper citation possible.
   a. Special characters and spaces in file and directory names should be replaced by underscores and/or hyphens or appropriate substitutions (e.g. ü -> ue).
   b. Label your data with the controlled vocabulary (your community or project standard).
2. Check that the **file formats** are sustainable: Only non-proprietary data formats
   - Containers: TAR, GZIP, ZIP

- Databases: XML, CSV
- Geospatial: SHP, DBF, GeoTIFF, NetCDF
- Moving images: MOV, MPEG, AVI, MXF
- Sounds: WAVE, AIFF, MP3, MXF
- Statistics: ASCII, DTA, POR, SAS, SAV
- Still images: TIFF, JPEG 2000, PDF, PNG, GIF, BMP
- Tabular data: CSV
- Text: XML, PDF/A, HTML, ASCII, UTF-8

and for model data **Network Common Data Format (NetCDF)** – „NetCDF is a set of software libraries and self-describing, machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data." (https://www.unidata.ucar.edu/software/netcdf/)– **GRIB** or **HDF** are accepted.

3. Check that the file names have a proper **file name extension.**

4. **Calculate the overall size** of the data to be shared and provide this information to the data management team.

5. **Create checksums (**see blue box) for each file before your data submission (please provide sha256 checksums). That will help us to guard data integrity during the data submission workflow and in the long run when published.

Create checksums on command line

Command to create SHA256 checksum of a file:

Linux: `sha256sum FILE_NAME`

MAC: `shasum -a 256 FILE_NAME`

WINDOWS: find appropriate tool

Start at the top level of your folder containing all data and other files and calculate checksums for each file of your dataset (including all subdirectories) and write them to a single text file with extension "sha256":

Linux:
```
cd YOUR_DATA_DIRECTORY
find ./ -type f -exec sha256sum {} \; >> Checksums.sha256
```

MAC:
```
cd YOUR_DATA_DIRECTORY
find ./ -type f -exec shasum -a 256 {} \; >> Checksums.sha256
```

WINDOWS: find appropriate tool

2.2 Submit your dataset for sharing or publication

After the preparation of your datasets you need to submit your data for publication to the data management team.

For data exchange there are the following options to choose:

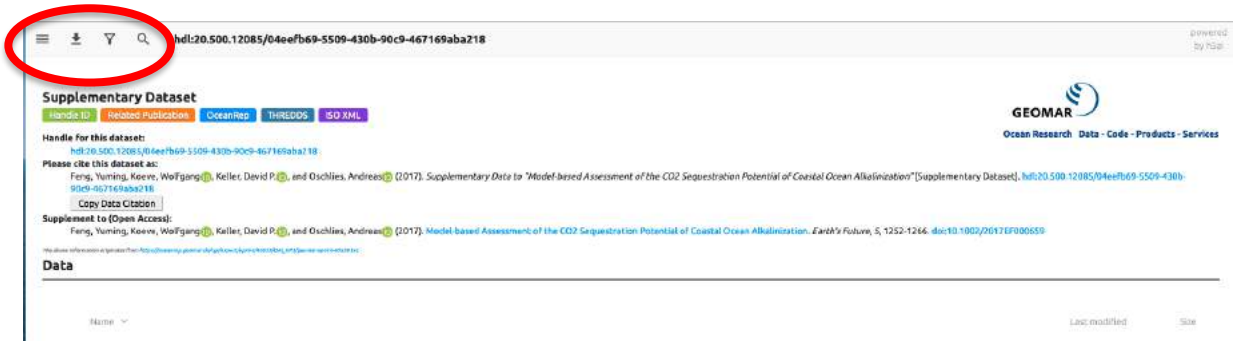
1. Upload the dataset you want to publish to the **GEOMAR FTP-Server**: contact helpdesk@geomar.de
2. Or alternatively upload the dataset to a **Git-LFS repository** for your project, which will be provided by the data management team (contact the data management team: datamanagement@geomar.de). For more information about Git see our Git Tutorial. By this, the data management team will have access to your data to process the publication request.
3. Or alternatively via a **SSH accessible storage** (GEOMAR Ceph, HPC Uni Kiel…)
4. Or alternatively via the **GEOMAR Cloud** (only if the other sharing options are not possible - limited space)

Fill in the data submission form https://portal.geomar.de/thredds-data-submission **to request a handle for your data!**

3. Overview of available services using handles and THREDDS for (model) data sharing

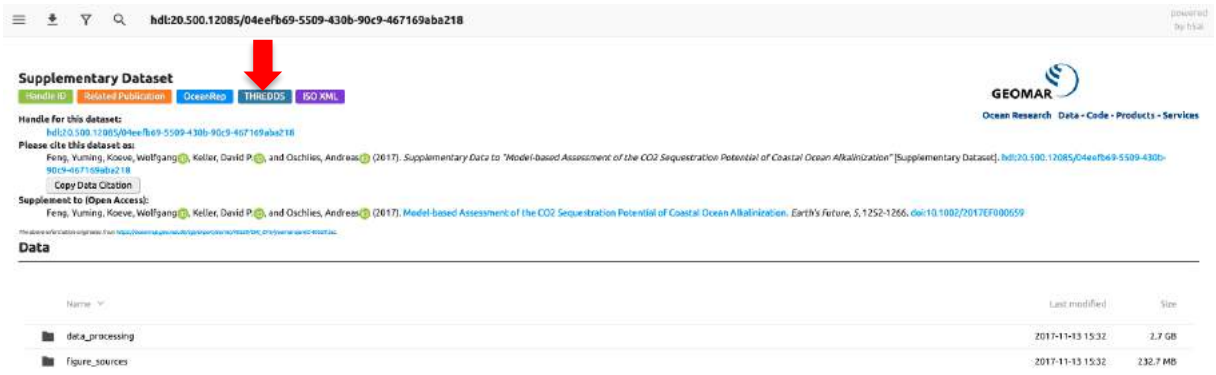
3.1 Direct Download

You can download single files or select multiple files in the "H5AI" browser view (see also filter button in the top line).



3.2 Options for accessing single NetCDF files:

Clicking on the THREDDS button will take you to the THREDDS browser based overview..



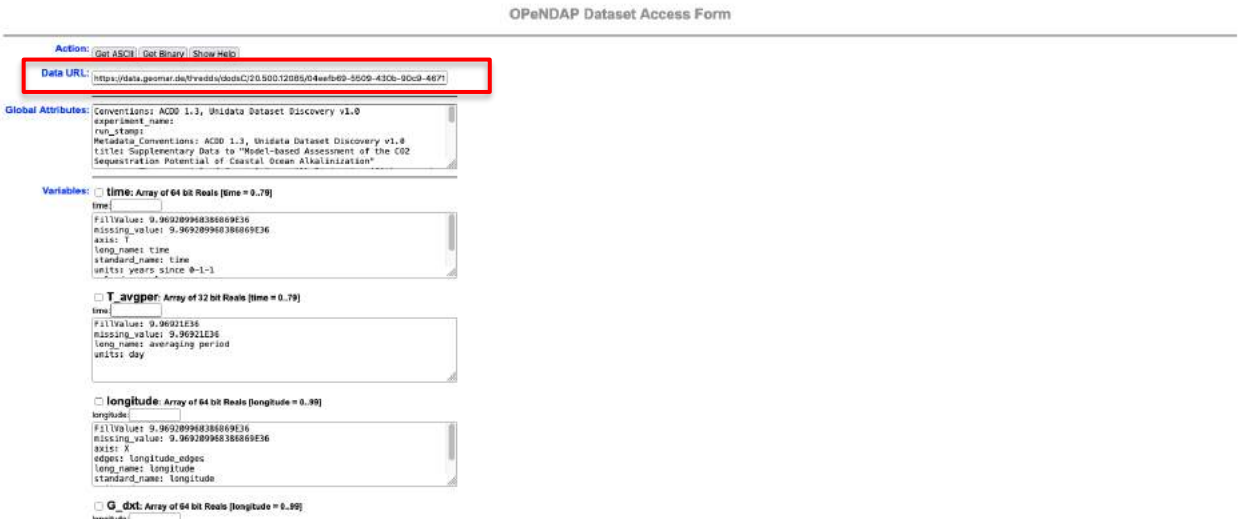
Thredds browser view:

Clicking on a *.nc entry gives you further access options:

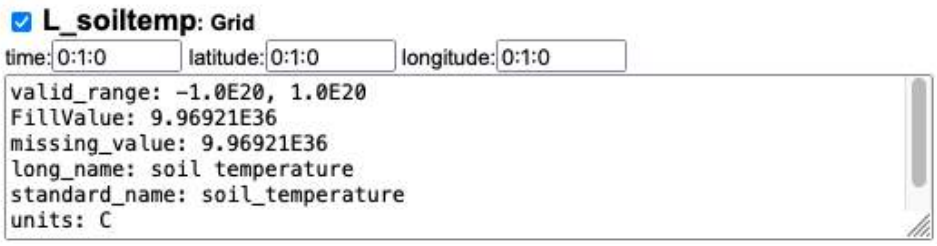
Overview of options for accessing single NetCDF files:

3.2.1 OPeNDAP Data Link

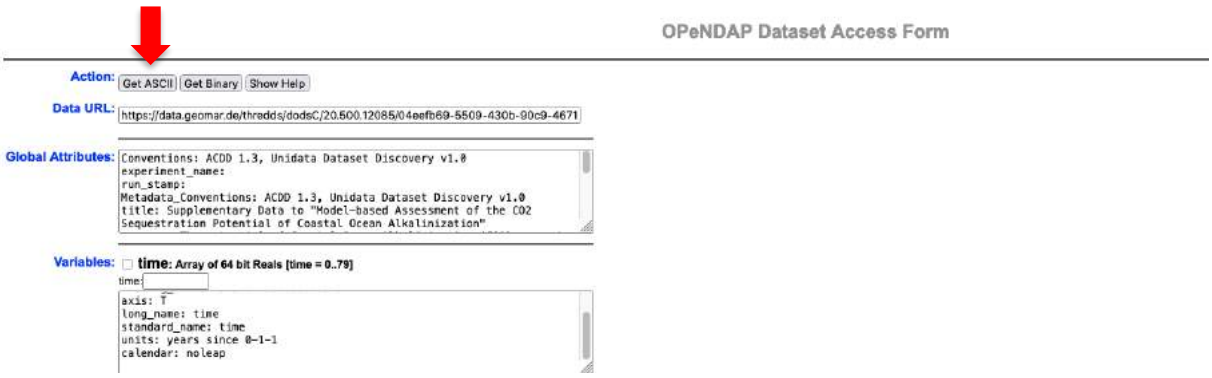
By clicking on the OPeNDAP-Link, you will get the OPeNDAP Dataset Access Form. Via OPeNDAP you can work on a subset of the dataset file without the need to download it to your computer. You can copy the ‚**Data URL**' and use it with your favorite software like Matlab etc. . In the other form fields, all variables of the dataset are listet.



You can get one subset of the data by checking the box of the interested variable (here e.g. L_soiltemp).



Then click on the ‚**Get ASCII**' button at the top of the form.



You will get the following result:

```
Dataset {
    Grid {
     ARRAY:
        Float32 L_soiltemp[time = 1][latitude = 1][longitude = 1];
     MAPS:
        Float64 time[time = 1];
        Float64 latitude[latitude = 1];
        Float64 longitude[longitude = 1];
    } L_soiltemp;
} 20.500.12085/04eefb69-5509-430b-90c9-467169aba218/model_runs/oliv1000_Omega3.4/results/tavg.02021.01.01.nc;
---------------------------------------------
L_soiltemp.L_soiltemp[1][1][1]
[0][0], 9.96921E36

L_soiltemp.time[1]
2020.5

L_soiltemp.latitude[1]
-89.0999984741211

L_soiltemp.longitude[1]
1.7999999523162842
```

If you have a look at the Data URL again, you will see that your subset of interest was added to the URL: e.g. https://data.geomar.de/thredds/dodsC/20.500.12085/04eefb69-5509-430b-90c9-467169aba218/model_runs/oliv1000_Omega3.4/results/tavg.02021.01.01.nc?L_soiltemp[0:1:0][0:1:0][0:1:0]
You can also use this URL with your favorite software and you will get your subset of interest.

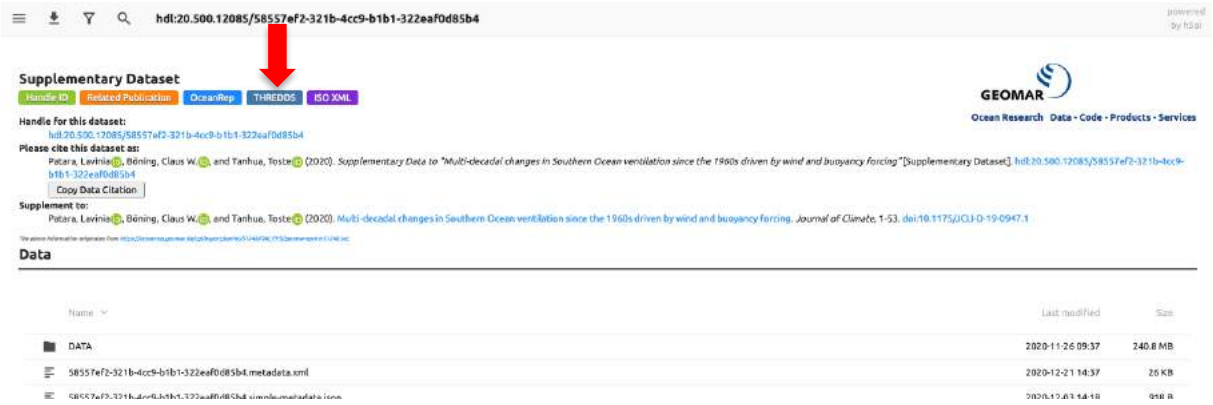
3.2.2 NetcdfSubset (NCSS)

If you follow the link "4. NetcdfSubset" on the Thredds overview of a Netcdf file, you will get to the NCSS interface. Here you can select which variables you would like to have and in which region. It may be worth taking a look at the NCSS documentation (link at the bottom of the page), where you can compile a URL for individual, customized downloads (parameters minx,miny,maxx,maxy) and build a *wget* script, for example.

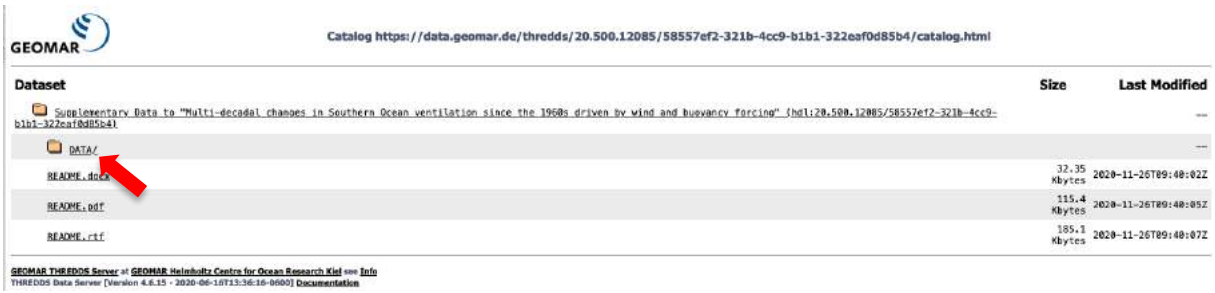
3.2.3 GODIVA2

Go to the Thredds overview of a Netcdf file and click on the first link "Godiva2" under "Viewers".

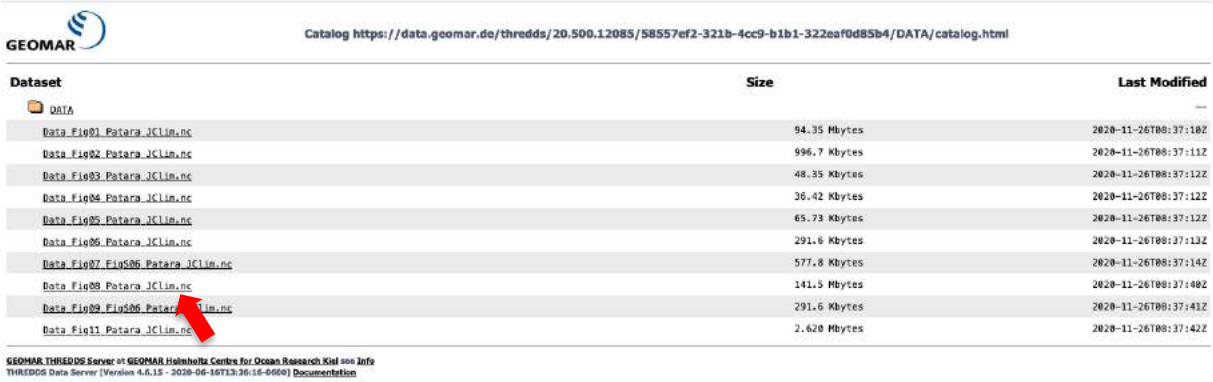
e.g. https://hdl.handle.net/20.500.12085/58557ef2-321b-4cc9-b1b1-322eaf0d85b4



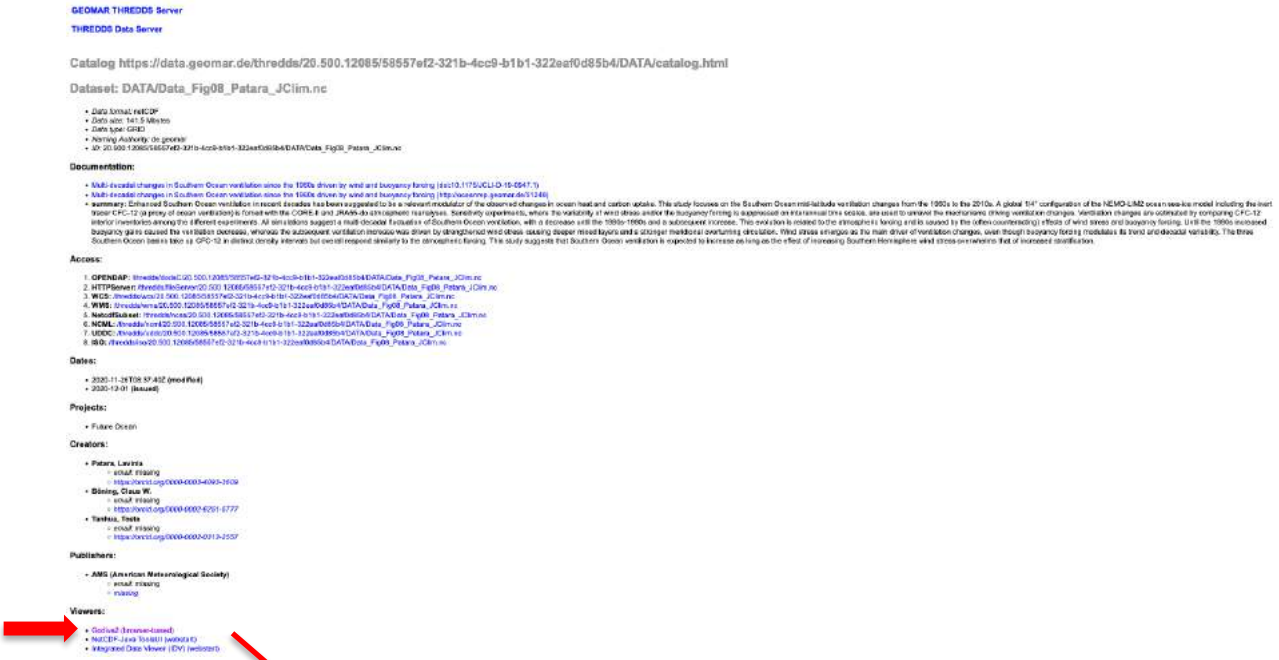
THREDDS browser view:



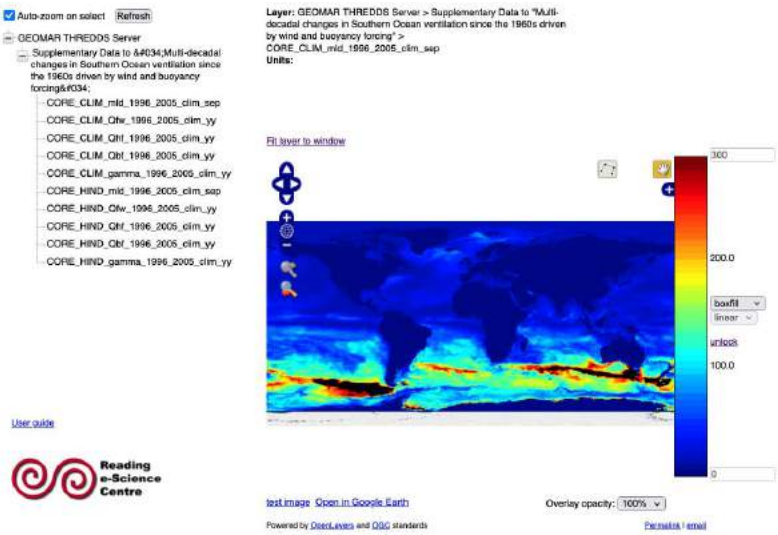
Clicking on a *.nc entry gives you the different viewers access options:



Click on ‚**Godiva2**' viewer at the bottom of the page.



Godiva2' viewer:



Select the variable and the time and modify the graphic. There is also a Google Earth link at the bottom of the graphics window and the animation function if you define the first and last frame via the drop-down menus and the links next to it. Additionally you can create an animation and a kmz file via the Google-Earth link and open it in Google-Earth.

If you have other questions or comments please contact the data management team:

Phone: 0431 / 600 2294

E-Mail: datamanagement@geomar.de

Location: Eastshore / Building 1/ Entrance 2/ Room 110 - 112

Adress: GEOMAR Helmholtz Centre for Ocean Research Kiel
Wischhofstr. 1-3
24148 Kiel | Germany