FROM THE DEEP SEA TO THE ATMOSPHERE

Establishing “Best practice” data workflows in marine research

Hela Mehrtens, P. Springer, C. Faber, L. Paglialonga, C. Schirnick
GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany

HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

GEOMAR

GEOMAR

Practising data management: who, when and how

- The components of data management @GEOMAR

- Examples:

    - Expeditions
    - Mesocosm experiments
    - Ocean modeling

GEOMAR

Central data management workflow @ GEOMAR

- Agreement on one **data management plan** (expected data, responsible persons, timeframes)
- One **information system** with metadata and linked data for exchange and monitoring deliverables
- **Versioning systems** (git/subversion)
- Accessibility and reuseability of research data based on **persistent identifiers** (DOI + handle)

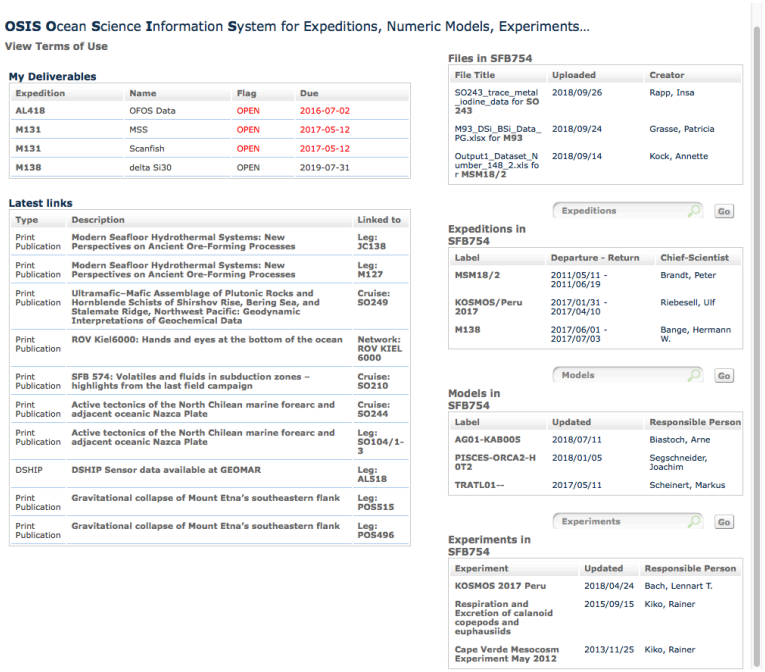
GEOMAR

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

GEOMAR

Ocean Science Information System - OSIS



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

View Terms of Use

My Deliverables

Expedition	Name	Flag	Due
AL418	OFOS Data	OPEN	2016-07-02
M131	MSS	OPEN	2017-05-12
M131	Scanfish	OPEN	2017-05-12
M138	delta Si30	OPEN	2019-07-31

Latest links

Type	Description	Linked to
Print Publication	Modern Seafloor Hydrothermal Systems: New Perspectives on Ancient Ore-Forming Processes	Leg: JC138
Print Publication	Modern Seafloor Hydrothermal Systems: New Perspectives on Ancient Ore-Forming Processes	Leg: M127
Print Publication	Ultramafic–Mafic Assemblage of Plutonic Rocks and Hornblende Schists of Shirshov Rise, Bering Sea, and Stalemate Ridge, Northwest Pacific: Geodynamic Interpretations of Geochemical Data	Cruise: SO249
Print Publication	ROV Kiel6000: Hands and eyes at the bottom of the ocean	Network: ROV KIEL 6000
Print Publication	SFB 574: Volatiles and fluids in subduction zones – highlights from the last field campaign	Cruise: SO210
Print Publication	Active tectonics of the North Chilean marine forearc and adjacent oceanic Nazca Plate	Cruise: SO244
Print Publication	Active tectonics of the North Chilean marine forearc and adjacent oceanic Nazca Plate	Leg: SO104/1-3
DSHIP	DSHIP Sensor data available at GEOMAR	Leg: AL518
Print Publication	Gravitational collapse of Mount Etna's southeastern flank	Leg: POS515
Print Publication	Gravitational collapse of Mount Etna's southeastern flank	Leg: POS496

Files in SFB754

File Title	Uploaded	Creator
SO243_trace_metal_iodine_data for SO243	2018/09/26	Rapp, Insa
M93_DSi_BSi_Data_PG.xlsx for M93	2018/09/24	Grasse, Patricia
Output1_Dataset_Number_148_2.xls for MSM18/2	2018/09/14	Kock, Annette

Expeditions Go

Expeditions in SFB754

Label	Departure - Return	Chief-Scientist
MSM18/2	2011/05/11 - 2011/06/19	Brandt, Peter
KOSMOS/Peru 2017	2017/01/31 - 2017/04/10	Riebesell, Ulf
M138	2017/06/01 - 2017/07/03	Bange, Hermann W.

Models Go

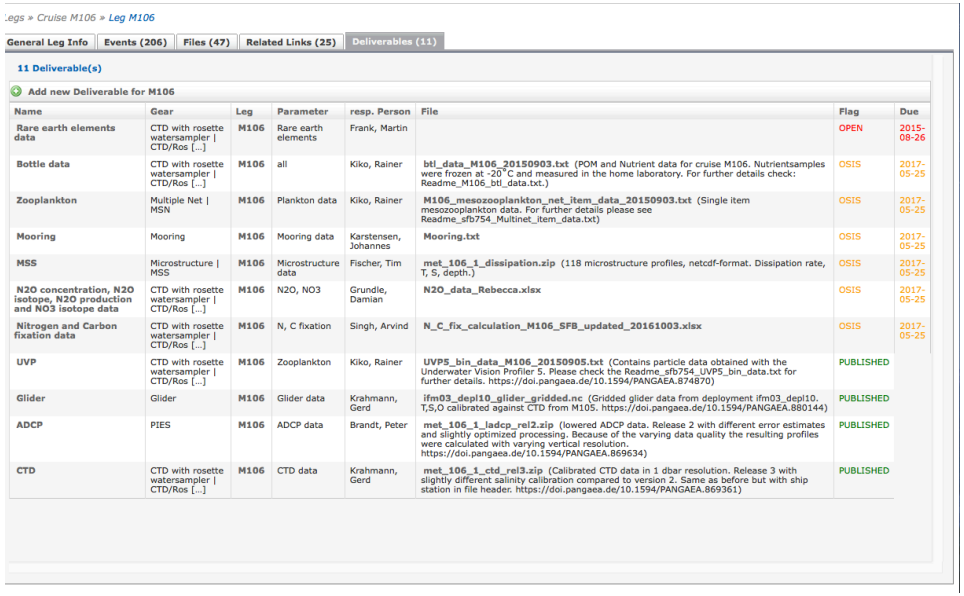
Models in SFB754

Label	Updated	Responsible Person
AG01-KAB005	2018/07/11	Biastoch, Arne
PISCES-ORCA2-H0T2	2018/01/05	Segschneider, Joachim
TRATL01--	2017/05/11	Scheinert, Markus

Experiments Go

Experiments in SFB754

Experiment	Updated	Responsible Person
KOSMOS 2017 Peru	2018/04/24	Bach, Lennart T.
Respiration and Excretion of calanoid copepods and euphausiids	2015/09/15	Kiko, Rainer
Cape Verde Mesocosm Experiment May 2012	2013/11/25	Kiko, Rainer



...egs » Cruise M106 » Leg M106

General Leg Info | Events (206) | Files (47) | Related Links (25) | Deliverables (11)

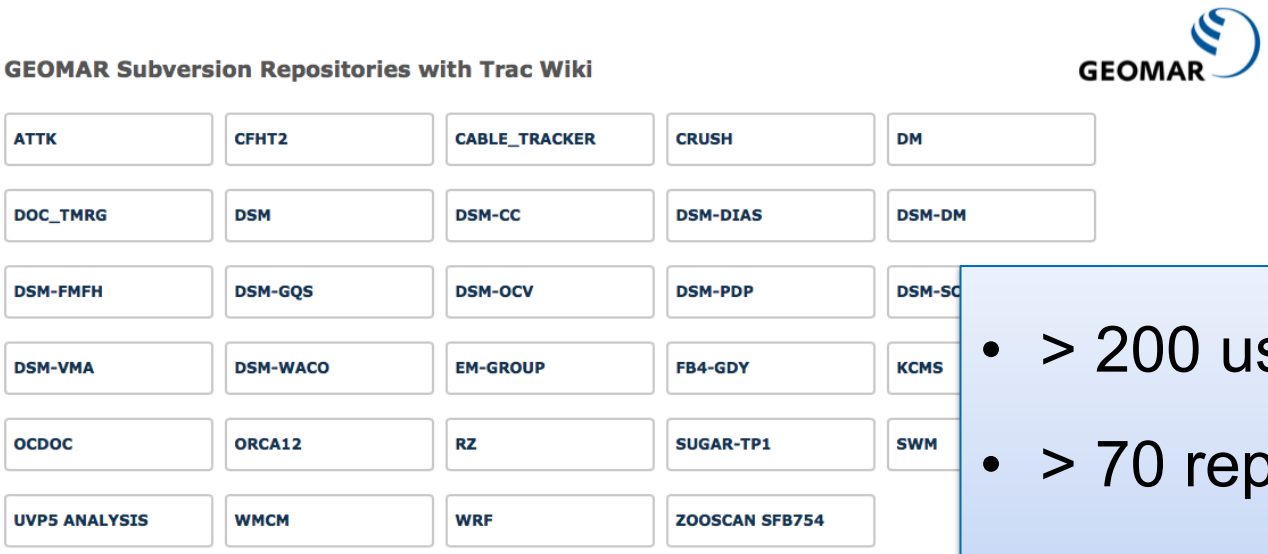
11 Deliverable(s)

Add new Deliverable for M106

Name	Gear	Leg	Parameter	resp. Person	File	Flag	Due
Rare earth elements data	CTD with rosette watersampler | CTD/Ros [...]	M106	Rare earth elements	Frank, Martin		OPEN	2015-08-26
Bottle data	CTD with rosette watersampler | CTD/Ros [...]	M106	all	Kiko, Rainer	btl_data_M106_20150903.txt (POM and Nutrient data for cruise M106. Nutrientsamples were frozen at -20°C and measured in the home laboratory. For further details check: Readme_M106_btl_data.txt.)	OSIS	2017-05-25
Zooplankton	Multiple Net | MSN	M106	Plankton data	Kiko, Rainer	M106_mesozooplankton_net_item_data_20150903.txt (Single item mesozooplankton data. For further details please see Readme_sfb754_Multinet_item_data.txt)	OSIS	2017-05-25
Mooring	Mooring	M106	Mooring data	Karstensen, Johannes	Mooring.txt	OSIS	2017-05-25
MSS	Microstructure | MSS	M106	Microstructure data	Fischer, Tim	met_106_1_dissipation.zip (118 microstructure profiles, netcdf-format. Dissipation rate, T, S, depth.)	OSIS	2017-05-25
N2O concentration, N2O isotope, N2O production and NO3 isotope data	CTD with rosette watersampler | CTD/Ros [...]	M106	N2O, NO3	Grundle, Damian	N2O_data_Rebecca.xlsx	OSIS	2017-05-25
Nitrogen and Carbon fixation data	CTD with rosette watersampler | CTD/Ros [...]	M106	N, C fixation	Singh, Arvind	N_C_fix_calculation_M106_SFB_updated_20161003.xlsx	OSIS	2017-05-25
UVP	CTD with rosette watersampler | CTD/Ros [...]	M106	Zooplankton	Kiko, Rainer	UVP5_bin_data_M106_20150905.txt (Contains particle data obtained with the Underwater Vision Profiler 5. Please check the Readme_sfb754_UVP5_bin_data.txt for further details. https://doi.pangaea.de/10.1594/PANGAEA.874870)	PUBLISHED	
Glider	Glider	M106	Glider data	Krahmann, Gerd	ifm03_depl10_glider_gridded.nc (Gridded glider data from deployment ifm03_depl10. T,S,O calibrated against CTD from M105. https://doi.pangaea.de/10.1594/PANGAEA.880144)	PUBLISHED	
ADCP	PIES	M106	ADCP data	Brandt, Peter	met_106_1_ladcp_rel2.zip (lowered ADCP data. Release 2 with different error estimates and slightly optimized processing. Because of the varying data quality the resulting profiles were calculated with varying vertical resolution. https://doi.pangaea.de/10.1594/PANGAEA.869634)	PUBLISHED	
CTD	CTD with rosette watersampler | CTD/Ros [...]	M106	CTD data	Krahmann, Gerd	met_106_1_ctd_rel3.zip (Calibrated CTD data in 1 dbar resolution. Release 3 with slightly different salinity calibration compared to version 2. Same as before but with ship station in file header. https://doi.pangaea.de/10.1594/PANGAEA.869361)	PUBLISHED	

GEOMAR

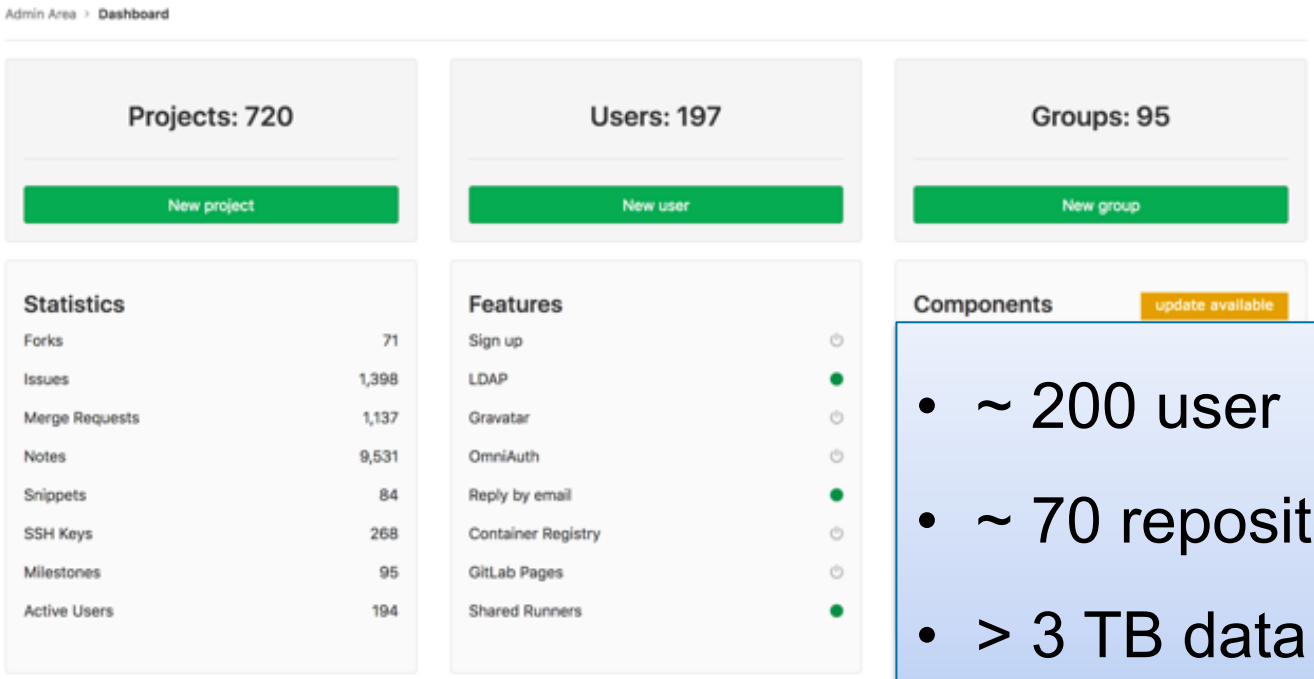
Versioning system: GEOMAR Subversion/Trac



- > 200 user
- > 70 repositories
- > 1 TB data + code

GEOMAR

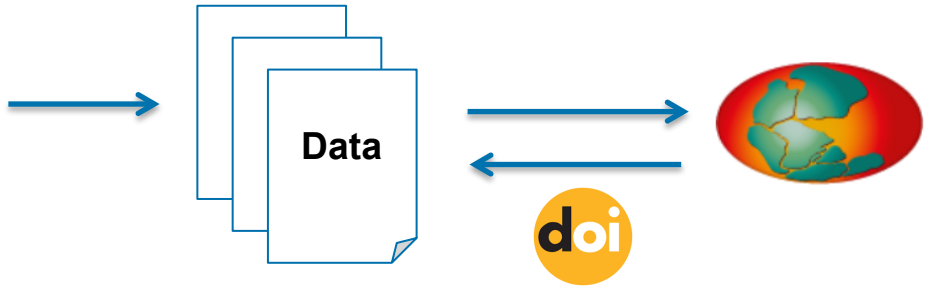
Versioning system: GEOMAR GitLab Server



- ~ 200 user
- ~ 70 repositories
- > 3 TB data + code

Persistent identifier

DOI-Registry at GEOMAR: Workflows



PANGAEA.

Data Publisher for Earth & Environmental Science

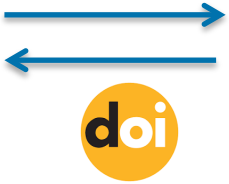
Scientists

METADATA, URL & zip-file

Software Code in (public) GitLab repository

GEOMAR Library/ OceanRep

METADATA & URL



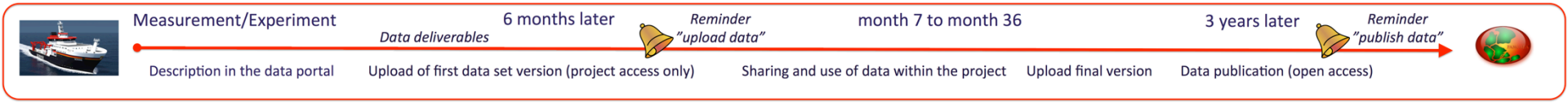
DataCite

FIND, ACCESS, AND REUSE DATA

GEOMAR

Example 1: Expedition workflow

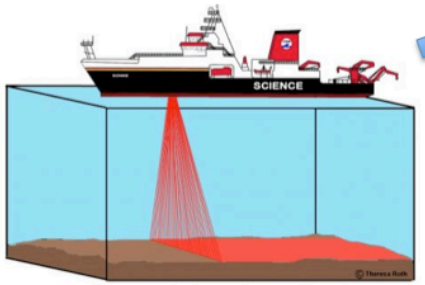
Data curation process



- Cruise planning involves DMP components
- Data reminder in OSIS used for > 80 expeditions
- 220 deliverables done: data are openly available as data publications with DOI at PANGAEA

GEOMAR

Workflow Bathymetry



http://www.oceanblogs.org ©Theresa Roth

DAVIS SHIP

Raw Data

+ Metadata

DATA

PANGAEA data sets

+ Quality Control

DATA

DATA

Data Products

Data availability:
World Data Centre
Long term archive
Public
Citable with DOI

AtlantOS

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

GEOMAR

Example 2: Mesocosm experiments

- Data policy finalised in June 2018
- Information System used for 10 experiments
- Data reminder used since June 2018
- 94 data publications with DOI at PANGAEA

Dataset Publication Year

- 2018 (17)
- 2017 (32)
- 2016 (34)
- 2015 (4)
- 2014 (2)
- 2013 (1)
- 2012 (1)
- 2010 (2)

more...

EXAMPLES

GEOMAR

Numerical Modeling Data - OPeNDAP Service

GEOMAR

Numerical Modeling Data Related to Peer Reviewed Journal Articles

2018 | 2017 | 2016 | 2015 | 2014 | 2013 | 2012 | before 2012

Biastoch et al. (2018), OM :: DOI 10.1016/j.ocemod.2017.12.002 :: Details in OceanRep

Greatbatch et al. (2018), JGR :: DOI 10.1002/2017JC013652 :: Details in OceanRep

Keller et al. (2018), GMD :: DOI 10.5194/gmd-11-1133-2018 :: Details in OceanRep

Kloewer et al. (2018), OM :: DOI 10.1016/j.ocemod.2018.09.006 :: Details in OceanRep

Rieck et al. (2018), JPO :: DOI 10.1175/JPO-D-17-0173.1 :: Details in OceanRep

Data Sets by Research Projects

Data Sets by Research Units

Software / Code

Bathymetric Data Sets

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.

GEOMAR

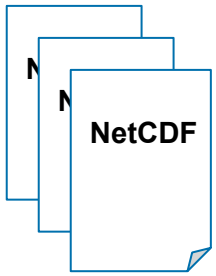
Catalog http://data.geomar.de/thredds/catalog/open_access/feng_et_al_2017_ef/catalog.html

Dataset	Size	Last Modified
Feng et al. (EF 2017)		--
README.txt	5.532 Kbytes	2017-11-13T15:45:52Z
data_processing/		--
feng_et_al_2017_ef_listing.txt	146.6 Kbytes	2017-11-15T15:01:05Z
figure_sources/		--
model_runs/		--

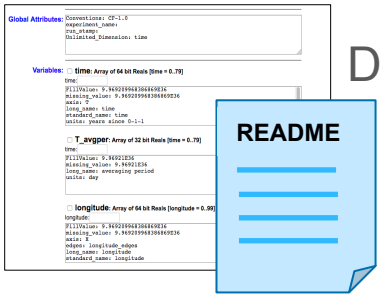
Catalog of Gridded Data at GEOMAR OPeNDAP Service see Info
THREDDS Data Server [Version 4.6.0 - 20150326.1318] Documentation

EXAMPLES

1. Preparation of model data for publication by data owner



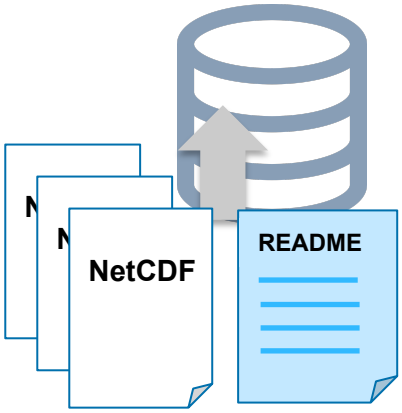
dataset organisation,
file formatting,
checksums



Metadata &
Documentation

GitLab for versioning,
project- & data management

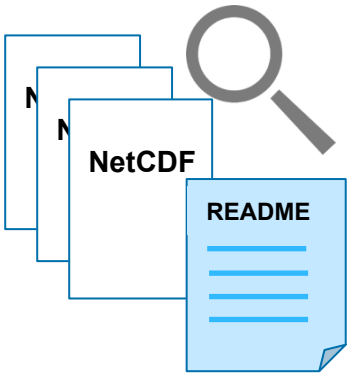
2. Dataset submission to the data management team



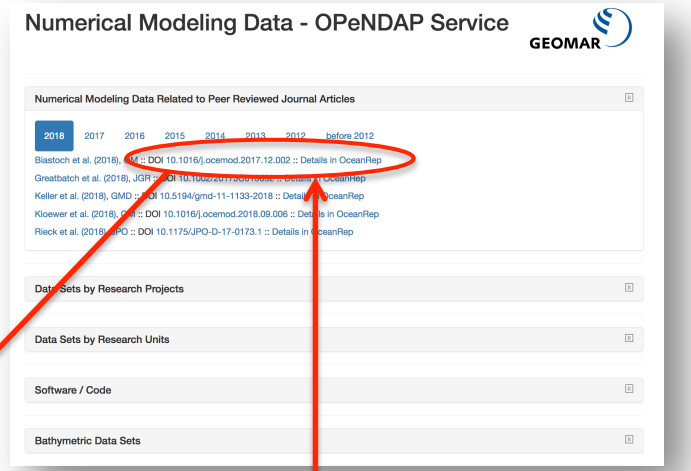
data transfer via
GEOMAR cloud or
(S)FTP server
&
data submission
information to DM
via web form

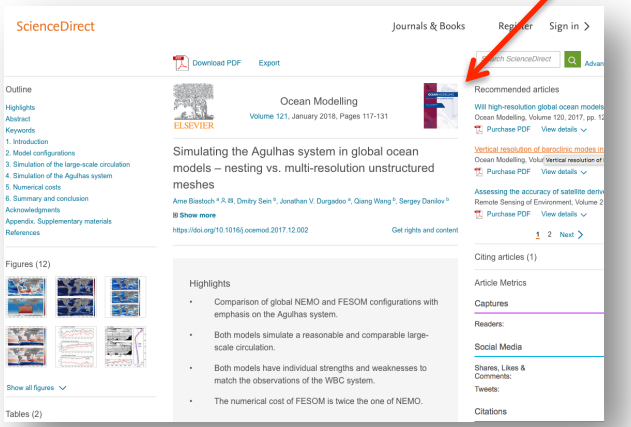
EXAMPLES

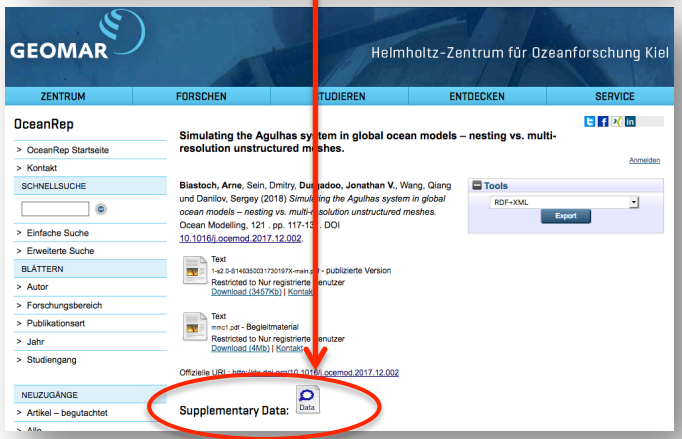
3. Data publication on data.geomar.de (OPeNDAP Service)



data integrity check
&
data dissemination
via THREDDS-Server
by data manager







Distribution of results

- OpenAire:

GEOMAR

Questions and comments?

Data Management Team @GEOMAR

datamanagement@geomar.de