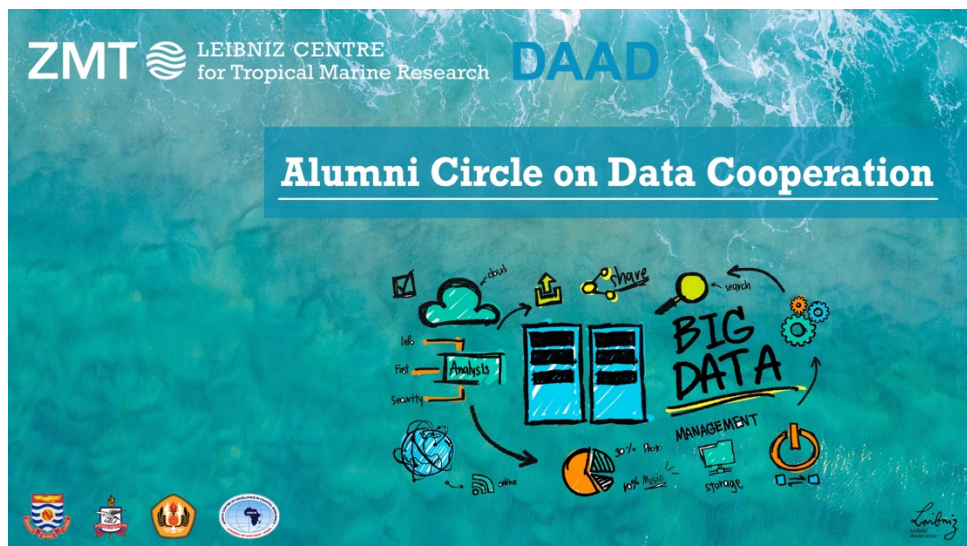


ZMT Alumni Circle on Data Cooperation

Workshop protocol

On June 28th, 2021 the second workshop of the Alumni Circle on Data Cooperation was conducted as a digital meeting via Zoom. The overall objective of the Alumni Circle is to foster interdisciplinary research addressing global environmental problems through data collaboration. Together with ZMT alumni and researchers, we pursued the following goals in this workshop:

- recap the **Alumni Conference on “Data management and collaboration for tackling marine challenges”** and discuss the results
- learn from ZMT and external colleagues about approaches to **research data management throughout the data life cycle**
- exchange of experiences in **working with databases and repositories**



CET	Program item
12:00 – 12:05	Welcome
12:05 – 12:15	Introduction of participants
12:15 – 12:30	Julian Lilkendey - Recap of the Alumni Conference (Numbers, Talks, Discussions, Outcomes)
12:30 – 13:10	Discussion
13:10 – 13:20	Coffee Break
13:20 – 13:45	Jan Härter (ZMT, Work Group Leader Complexity and Climate) - Big data and its handling in atmospheric science
13:45 – 14:15	Alexandra Nozik (ZMT, Geodata Specialist at the Submarine Groundwater Discharge Workgroup) - Supporting scientists in their work with geodata
14:15 – 14:50	Holger Kuhlmann (MARUM, Superintendent of the IODP Bremen Core Repository) – The Drilling Information System (DIS)
14:50 – 15:00	Closing

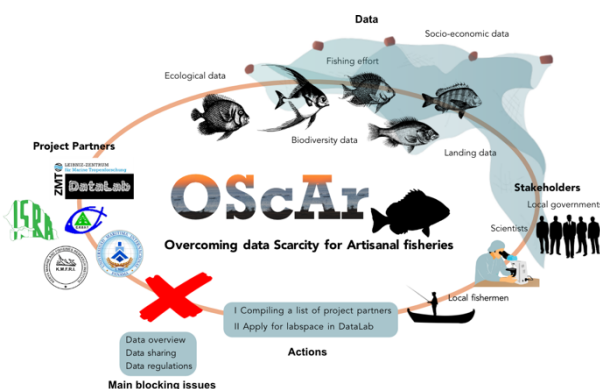
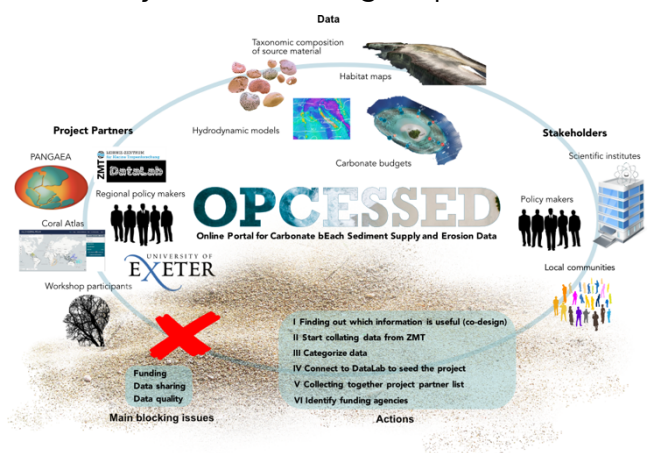
Participants of Workshop ZMT Alumni Circle on Data Cooperation		
João Marcelo Brazão Protázio	Laboratory of Mathematical-Statistical Modeling, ICEN - UFPA, Brazil	Alumni Experts
Arturo Dominici-Arosemena	International Maritime University of Panama (UMIP), Panama	
Bocar Baldé	Institut Sénégalais de Recherches Agricoles (ISRA) / Centre de Recherche Océanographique Dakar-Thiaroye (CRODT), Senegal	
Charles Faseyi	University of Cape Coast, Ghana	
Ima Kusumanti	College of Vocational Studies IPB University, Indonesia	
Ines Lange	University of Exeter, United Kingdom	
Sucharit Neogi	Coastal Development Partnership (CDP), Bangladesh	
Lucy Obare	Kisii University, Kenya	
Nehru Prabakaran	Wildlife Institute of India, India	
Manuel Jänig	Leibniz Centre for Tropical Marine Research (ZMT)	
Joscha Schmiedt	Leibniz Centre for Tropical Marine Research (ZMT)	
Fridolin Haag	Leibniz Centre for Tropical Marine Research (ZMT)	
Julian Lilkendey	Leibniz Centre for Tropical Marine Research (ZMT)	ZMT Alumni Network
Alexandra Nozik	Leibniz Centre for Tropical Marine Research (ZMT)	Guest Speakers
Jan Härter	Leibniz Centre for Tropical Marine Research (ZMT)	
Holger Kuhlmann	Center for Marine Environmental Sciences (MARUM) / IODP Bremen Core Repository	

Conference Recap

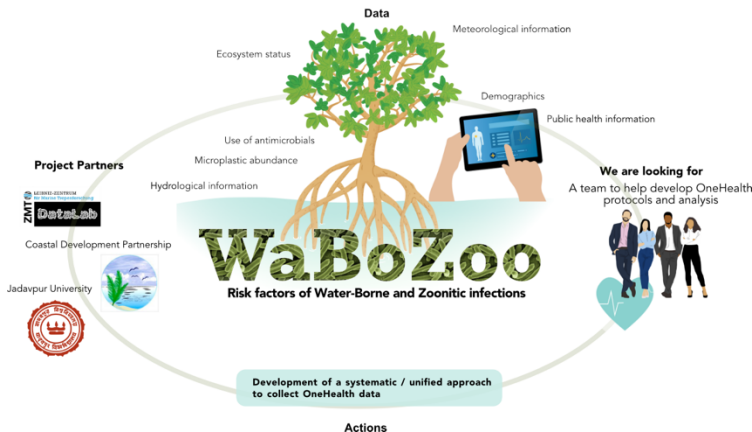
During the alumni conference entitled “Data management and collaboration for tackling marine challenges” we saw a total of 21 talks by ZMT researchers and alumni. Alumni joined from the most ZMT partner countries within the tropical belt including Panama, Brazil, Senegal, Ghana, South Sudan, Kenya, Philippines, Bangladesh, India, Indonesia and European countries such as Germany, Norway, Finland and the United Kingdom. The combined event consisting of conference and panel discussion was attended by 117 unique users, and the total streaming time amounted to 10 days and 11 hours, with 344 one-one-connections. Although we are sorry that we could not meet in person, one of the greatest successes of this event was to enable networking and exchange of ideas between ZMT researchers and alumni across borders and time zones, saving 33,000 kg of CO₂ and 270,500 travel kilometres. One of the conference highlights was the talk by Regina Therese Bacalso on **The Open Data Kit (ODK) for a quick, simple, and participatory fisheries data collection and accessibility in the tropics**. It's an innovative smartphone app that is used to collect fisheries data in the field. She pointed out perks of her methodology, such as cutting out the step of time-consuming data encoding but also highlighted problems such as internet connectivity and localization.

One of the main goals of our virtual alumni conference was to develop new project ideas and to find partners for data-centric collaborations. Together with ZMT researchers, our alumni have developed four auspicious project ideas for which they are now seeking cooperation in the grant writing process and data exchange:

During the conference ZMT alumnus Ines Lange, Gary Murphy, and Yannis Kappelmann (both ZMT) sat down together and formulated their project idea for an **Online Portal for Carbonate bEach Sediment Supply and Erosion Data (OPCESSED)**. Their main goal is to first find out what information is actually useful (co-design) and to create a network of institutes/stakeholders to collect data and inform them about the underlying processes of beach accretion/erosion. Through their efforts, they aim to engage local communities about processes that affect “their” beaches.



ZMT alumnus Bocar Sabaly Baldé and Manuel Jänig (ZMT) developed the project **Overcoming data Scarcity for Artisanal fisheries (OScAr)**. The two want to improve the quality of small-scale fisheries data and compare the means of data collection between regions and scientific fields. They also want to encourage colleagues to share their experiences in incentivising fishermen to provide data. Ultimately, their goal is to fill local data gaps using citizen science-based tools.

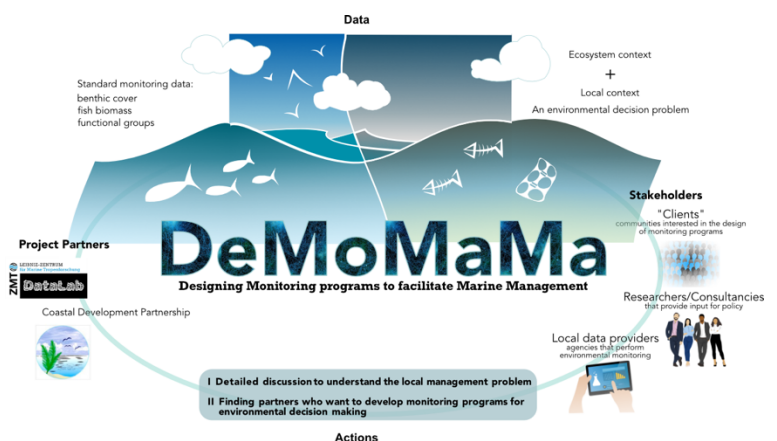


In our project development Session 2, ZMT alumnus Sucharit Neogi proposed the project **Risk factors of Water-Borne and Zoonotic infections (WaBoZoo)**.

One Health is an approach to designing and implementing programmes, policies, legislation and research in which multiple sectors communicate and work together to

achieve better public health outcomes. The areas of work in which a One Health approach is particularly relevant include food safety, the control of zoonoses (diseases that can spread between animals and humans, such as flu, rabies and Rift Valley Fever), and combatting antibiotic resistance (when bacteria change after being exposed to antibiotics and become more difficult to treat). Sucharit's aim is to form a multidisciplinary team to support his partners in data management beyond One Health criteria.

Last but certainly not least, ZMT's Fridolin Haag proposed the project **Designing monitoring programs to facilitate marine management (DeMoMaMa)**. In this project, Fridolin wants to conduct virtual experiments and collect monitoring data in order to optimise the cost-benefit ratio of environmental monitoring. His ultimate goal is to support the design of decision-making and monitoring in environmental management.



Working with remote sensing data and managing the earth's palaeoarchive

Jan Härter gave the audience an introduction to climate data, its collection and its storage volume. While some data are collected on the ground, remotely sensed satellite data make up an important part of all data in the field. However, most climate scientists actually work with simulation data, which allows, for example, the comparison of coupled ocean-atmosphere general circulation models. Jan then introduced the concept of "reanalysis data" - a hybrid between pure observational data and pure model data - and the netCDF file format.

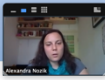
What is climate data?

e.g., Atmosphere: physical variables, usually 3D+time
(e.g., pressure p , temperature T , humidity q_v , condensed water q_c , rainfall r ; velocities u, v, w)
consider: 1000 km x 1000 km x 20 km domain, 1 km data resolution, 1 day at 10 min temporal resolution, double precision (8 bytes)
each variable: $2 \cdot 10^7 \times 144 \sim 10^{10}$ bytes = 10 gigabytes
all variables: ~100 gigabytes



Neither scientist, nor IT... what is a geodata specialist?

- focus of my position: data process design
(data management design for ZMT is in development by Research Data Management WG)



Spatial thinking

The goal of the project

base data search on your research question, know what you are looking for

Finding spatial data

Data sources

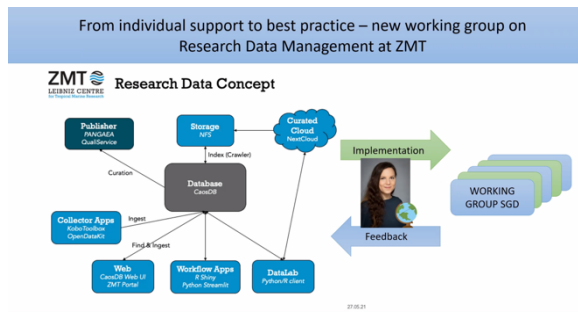
Where
When
Who
What

Handling spatial data

Data processing

Working with GIS software (import, processing, visualization)
Handling challenges: wrong projections, finding right tools

Alexandra Nozik also has a lot of experience in dealing with two- and three-dimensional remote sensing data, as she is mainly working with aquatic systems. Alexandra will soon join the DigiZ working group at ZMT to support researchers working with remote sensing data.



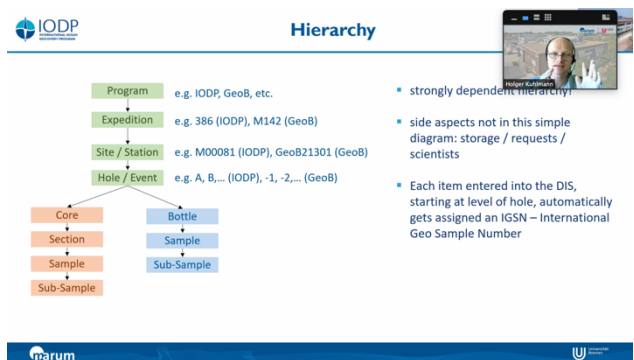
During her time as a research assistant in the Submarine Groundwater Discharge working group here at ZMT, she has already been involved in several projects spanning multiple disciplines. In her new position, she will focus on data process design, so she advises the audience to think about data and resources as early as possible in a project.

Alexandra acknowledges that there is very heterogeneous data in myriad databases and repositories, so there should also be a strong focus on where to find spatially resolved data. She is also able to help researchers develop workflows, especially when using different geographic information system (GIS) software.

Switching from present to past climates, as sediment cores are essentially geological paleoclimate archives. **Holger Kuhlmann** described past and present approaches of managing a sediment core collection which is more than 50 years old – the Integrated Ocean Drilling Program (IODP) Bremen core repository. The collection is one of three international core repositories. It contains approx. 158 km of cores in 220,000 individual pieces, which are stored professionally at +4 °C in a 1,100 m² cold storage hall.



Holger briefly introduced how the cores are sampled, only to then dive into describing the labelling system and metadata of the samples shortly afterwards. The inventory is managed by the Drilling Information System (DIS), which not only allows scientists to access the collection and order subsamples of sediment cores, but also to keep track of how much sample material is left per core. With his experience in database development and collection management, Holger emphasised that "consistency" is one of the main considerations in successfully managing a database.



Databases and repositories

In a 2020 survey, the majority of ZMT alumni responded that they have little overview about what data repositories exist and where they can store or access data. One aim of this workshop consequently was to identify relevant databases and repositories. The topic of databases was introduced by playing a video of the discussion after Daniel Pauly's (creator of **Sea Around Us**, and **FishBase**) talk at our alumni conference. During the discussion, Daniel Pauly explained that the crux of a specialized database such as FishBase is that you cannot analyse the relevant data when you enter it as a free text because they are not relatable. You need to break the information down into "bits", which can be given either as a number or as multiple choices, to be related and to be counted. In Daniel Pauly's view the achievement of FishBase was to reduce or express the whole discipline of ichthyology into a form of fields into which numbers were then entered (length-weight ratios, growth, length at maturity, food) to be manageable.

In the panel discussion following our alumni conference, participant Frank Oliver Glöckner is himself a testimonial for good data practice. Thus, he extensively promoted the publication of data - be it in the repository he and his colleagues developed (**PANGAEA**) or in international databases. Nina Wambiji made the iconic statement that "data is the answer to everything", because once you have basic ecological information available in databases or repositories, it can be analysed and you can make management recommendations depending on it. During the discussion, Luiz Gadelha pointed to a project he has been involved in that operates at the intersection of a data hub and citizen science: the **Brazilian National Animal Health Information System (SIZ)**. The application is particularly successful because it is also available for handheld devices. Nina Wambiji mentioned a similar application, **iNaturalist**, which publishes all volunteer-generated data in the **Global Biodiversity Information Facility (GBIF)**. Estradivari made reference to the **Spatial Monitoring and Reporting Tool (SMART)** as a valuable measure for enforcing management guidelines in marine protected areas. She also pointed out that citizen science projects are not only about collecting data, but also about raising awareness because people then directly take notice of what is happening around them. With this statement the discussion completed the detour via databases and citizen science projects to show a concrete way in which data management initiatives enable environmental sustainability in the tropics.

In the discussion after her talk, Alexandra Nozik mentioned the **European Space Agency's (ESA) Copernicus Open Access Hub** (<https://scihub.copernicus.eu/>) as one of her go-to online databases. In his talk, Holger Kuhlmann pointed the audience towards the **International Ocean Discovery Program's (IODP) portal for Bremen Core Repository Curation Data and Mission Specific Platform Expedition Data** (<https://iodp.pangaea.de/>). During the #DatAlumni Course "Introduction to Research Data Management: Why? When? How?" participants compiled a list of data bases and repositories (see stickies). To additionally inform ZMT researchers and alumni about data repositories to publish and retrieve data, we compiled a list of data repositories and portals which can be found online: <https://www.leibniz-zmt.de/en/research/research-projects/dat-alumni.html>

