

## Workshop Summary

On September 16th -17th the first workshop of the Alumni Circle on Data Cooperation was conducted as a digital meeting. We aimed at pursuing two goals at this workshop:

- **Identifying major challenges** for researchers **to manage, archive and retrieve data** from repositories.
- Formulating ideas for **training and information materials** to overcome those challenges and to **promote FAIR research data handling**.

The newly formed ZMT Alumni Circle consists of 4 international Alumni, and two ZMT experts. The workshop was also attended by the DatAlumni project coordinator Dr. Janna Just as well as the head of the ZMT Academy and Alumni Relations Dr. Janine Reinhard. Four invited guest speakers shared their perspectives on different aspects of data management.

Participants of Workshop ZMT Alumni Circle on Data Cooperation		
João Marcelo Brazão Protázio	Laboratory of Mathematical-Statistical Modeling, ICEN - UFPA, Brazil	Alumni Experts
Emmanuel Acheampong	Department of Fisheries and Aquatic Sciences, University of Cape Coast, Ghana	
Lata Gawade	National Institute of Oceanography, Panaji, India	
Tridewi Pribadi	Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Inodensia	
Theresa Schwenke	Center for Marine and Tropical Research, Germany	ZMT Experts
Julian Lilkendey	Center for Marine and Tropical Research, Germany	
Janna Just	Center for Marine and Tropical Research, Germany	ZMT Alumni Network
Janine Reinhard	Center for Marine and Tropical Research, Germany	
Nicole Bloy	Center for Marine and Tropical Research, Germany	
Hauke Reuter	Center for Marine and Tropical Research, Germany	Guest Speakers
Arjun Chennu	Center for Marine and Tropical Research, Germany	
Marianne Kunkel	Alfred-Wegener Institut Bremerhaven, PANGAEA, Germany	
Tanja Hörner	University of Bremen Research Alliance, Germany	

## Challenges of Data Management - Day 1

After getting to know each other, the focus of the initial workshop day was to identify the challenges of Data Management through group discussions and guest speaker presentations. Dr. Hauke Reuter, who is responsible for the ZMT data base spoke on the “Importance of Metadata”. After the presentation we discussed about challenges to make scientists aware of required formats and metadata that need to be collected in the field. We further talked about the difficulties for interdisciplinary research project to utilize a common data base. Dr. Arjun Chenu introduced his perspective and of data management with regard to the FAIR principles. We discussed afterwards how to cope, if after the time of data collection, necessary metadata have not been recorded. The importance to clearly defined data management plans was emphasized by the participants. Dr. Julian Hilkeney introduced the [AANChOR](#) project, and the database AANoCID (working title All Atlantic Ocean Initiatives Database) he is developing on marine initiatives. By now, the later platform is available internally but should be integrated into the AANChOR data portal. Dr. Janna Just showed the results of the online DatAlumni Survey “Challenges of Data Management”.

In two break-out groups the Alumni Circle identified the major challenges of data management and data archiving and retrieval from repositories. These challenges were afterwards discussed in the big group and categorized.

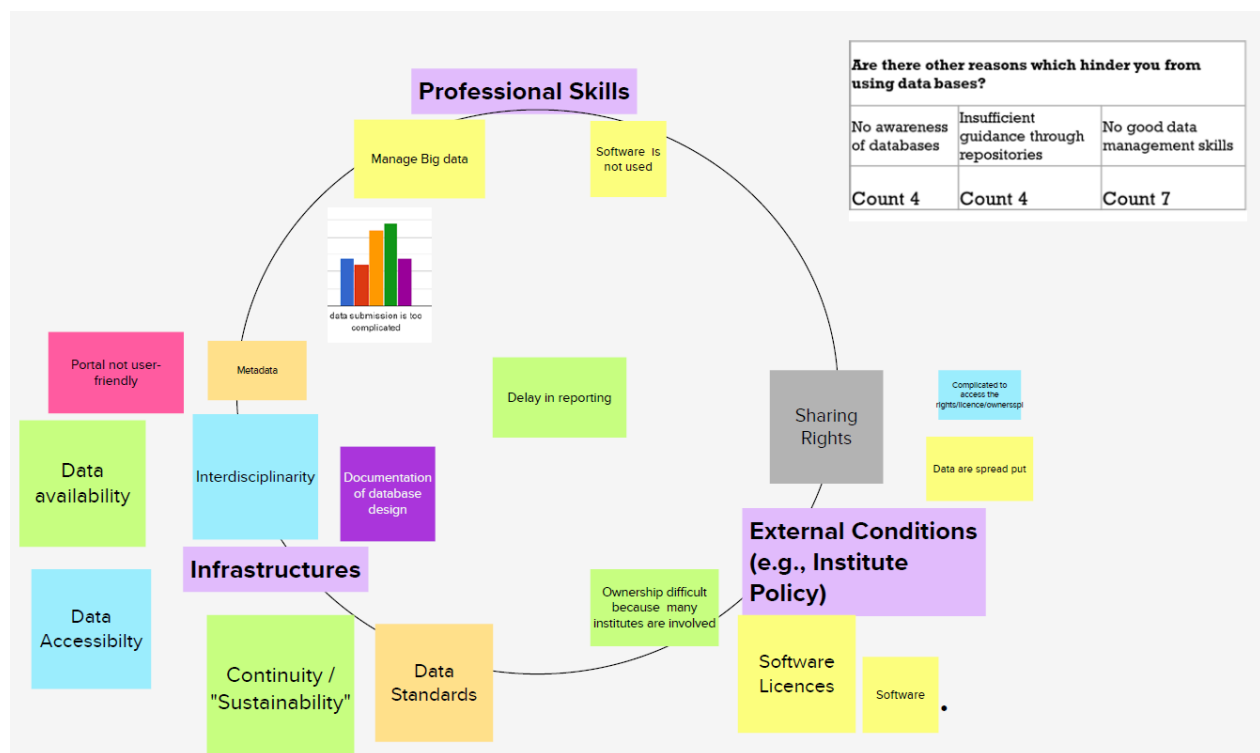


Figure 1: Categorized challenges of data management identified by the workshop participants. Inset in upper right corner displays reasons hindering the usage of data bases outlined by participants of the online DatAlumni survey on data management.

### Infrastructures:

Among challenges with regard to infrastructures, **undefined metadata** and the unavailability of data standards was remarked by many participants. In particular, pre-defined data standards in “structured

data repositories” might not be applicable for all scientific disciplines, so that interdisciplinary is not ensured. To the user end, **the documentation** and tutorials are sometimes insufficient. A third important point was the **access to the stored data** (open-access, or just visibility) as well as the sustainability of the data bases and **continuity of maintenance**.

### External Conditions

In large research projects, with multiple participating scientists and institutions, data are often stored at **different instances** and **ownership** of the data is not clearly defined. Insufficient **access to software** licenses makes data management more complicated for the user.

### Professional Skills

Many (young) reserachers are not **trained to manage** big amounts of data. We discussed at the workshop that this is on one hand related above mentioned insufficient access to software, and not enough training in basic computer applications. The later point was also among the highest challenges from the DatAlumni online survey.

## Training Concepts for Data Management – Day 2

The second day of the workshop was dedicated to developing ideas for training materials for data management. Dr. Tridewi Pribadi gave us an interesting overview on the data management system in the human resources and student administration at Universitas Padjadjaran, Inodensia. For the management of research data, she pointed out, that many universities in Indonesia have their own data bases, and centralized structures need yet to be developed. In particular, trained and dedicated data managers, that secure a long-term maintenance is indispensable. Dr. Emmanuel Acheampong continued to give insight into his work at Africa Centre of Excellence in Coastal Resilience ([ACECoR](#)) and the Database [FishCoM Ghana](#) for which he integrated fisheries data from various institutions. Afterwards Dr. João Marcelo Brazão Protázio introduced the newly initiated Data Management Strategy of the Postgraduate Program in Mathematics and Statistics at Universidade Federal do Pará. A protocol for meta data implementation, data submission, quality control and data archiving is set into function.

External guest speaker Marianne Kunkel from the Alfred-Wegener Institute provided insights into the workflow at the [PANGAEA repository](#) and gave insights into the challenges of data curators. She provided practical tips, what to consider when/before submitting data to a database. We exchanged ideas, how the engagement of researchers in data archiving can be made visible. Marianne introduced that data archived at PANGAEA can be linked to researcher’s [ORCID](#) profiles. Dr. Tanja Hörner, the coordinator Graduate Program of Data Science of the University of Bremen Research Alliance ([UBRA](#)), presented the concept of this freshly initiated program. An education in the *Data Steward track* and *Data Scientist track* are offered to all Graduate Students of the UBRA, through participating in interdisciplinary courses in research data management and data science. The discussion afterwards emphasized a strong interest to consider such a program in the international context and stimulated the following group work on possible training concepts.

In two groups the participants and expert guest speakers discussed how potential training material could be designed, in order to overcome the challenges (Day 1) of data management. A discussion with all workshop participants followed, during which we categorized the training activities.

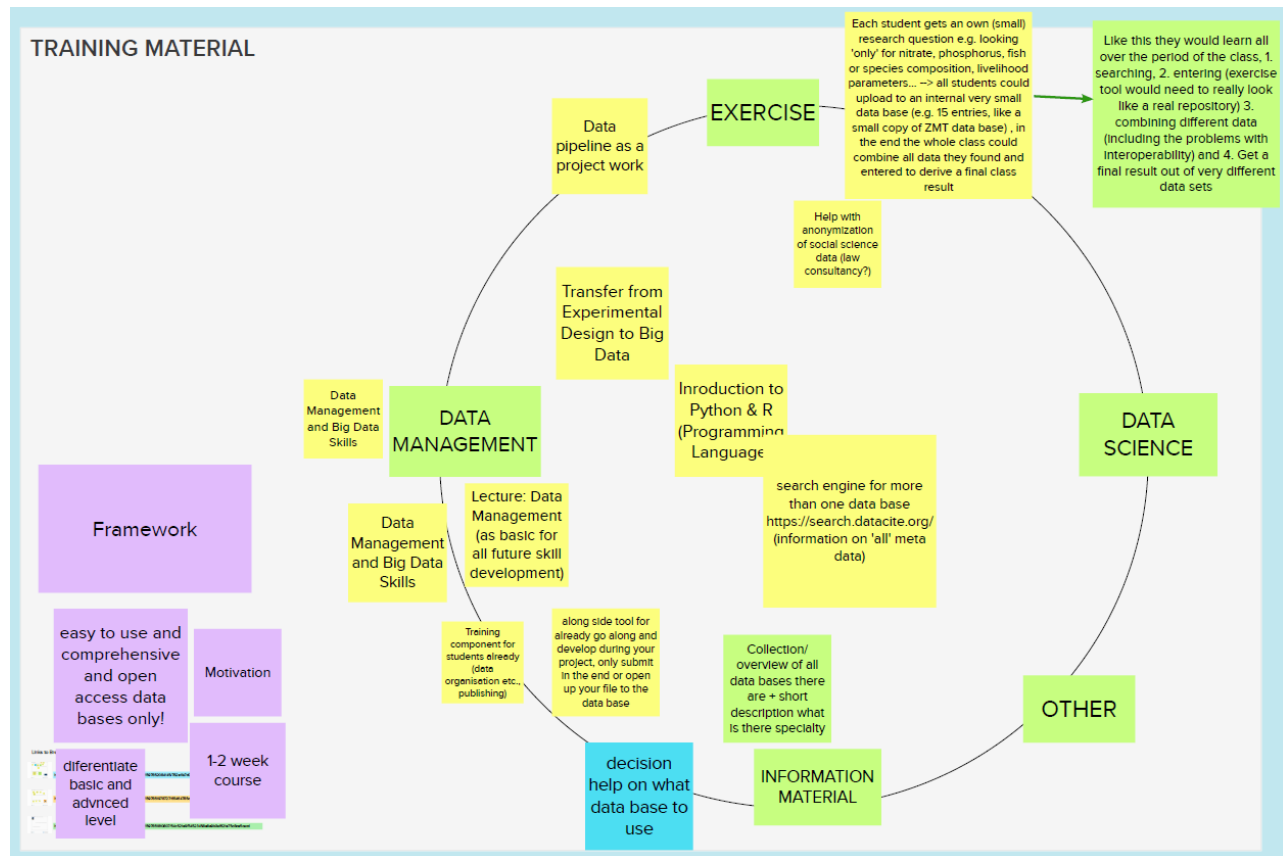


Figure 2: Ideas for potential training materials helpful for supporting data management expertise were collected by the participants. Besides the categories Exercises, Data Management, Data Science and Information Material, we identified the framework, that should be considered when designing training materials.

### Exercises, Lectures and Courses on Data Management

Exercise ideas mainly focused on **hands-on training**, where course participants complete the whole **data cycle** of data retrieval from a data base, data organization and data upload into a data base. Other trainings, as helping with the **anonymization of social science** data were identified as immensely helpful.

Improving data literacy by offering courses on **the basics of computer application**, which the participants can afterwards build upon for statistical analysis was strongly suggested by all participants (e.g., R, Python). Further offers could be lecture on **Data Management Plans**, potentially in combination with a practical exercise.

### Information Material

We discussed the usefulness of a list of **existing research data bases** with a short description, helping for decision making which data base would be most applicable for the ZMT Alumni's data. Moreover, **the information on necessary metadata** and how to structure the data on the personal file, should be made clear from the beginning. Such information would feed into a potential course on data management plans (see above).

## Following Activities

For the training activities and the DatAlumni project, the following steps were identified at the workshop:

- Create a **list of available data bases** with regard to the different disciplines, advantages/disadvantages
- Develop concept **for course on data management plans** (Lecturer/Content)
- Establish a regular meeting open to ZMT Alumni with regard to data management at **ZMT Data Round Table**
- Plan **R, and Data Management courses, Anonymization of Social Science data for Alumni Conference** in spring 2021 (parallel courses)

The team of the ZMT Academy with support from the Alumni Circle will continue their activities in those directions.

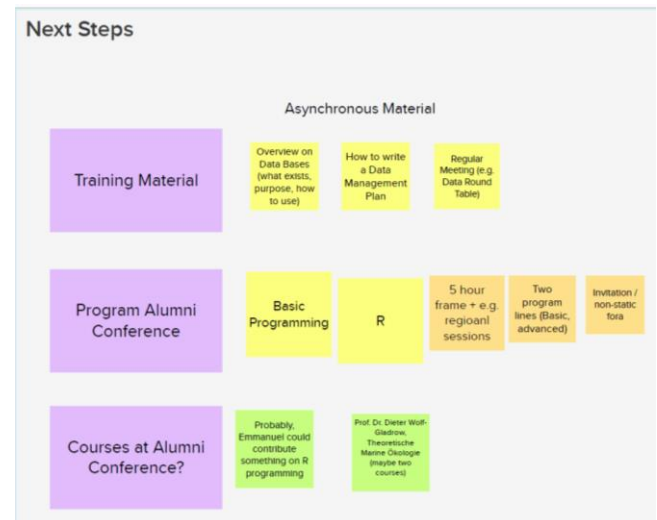
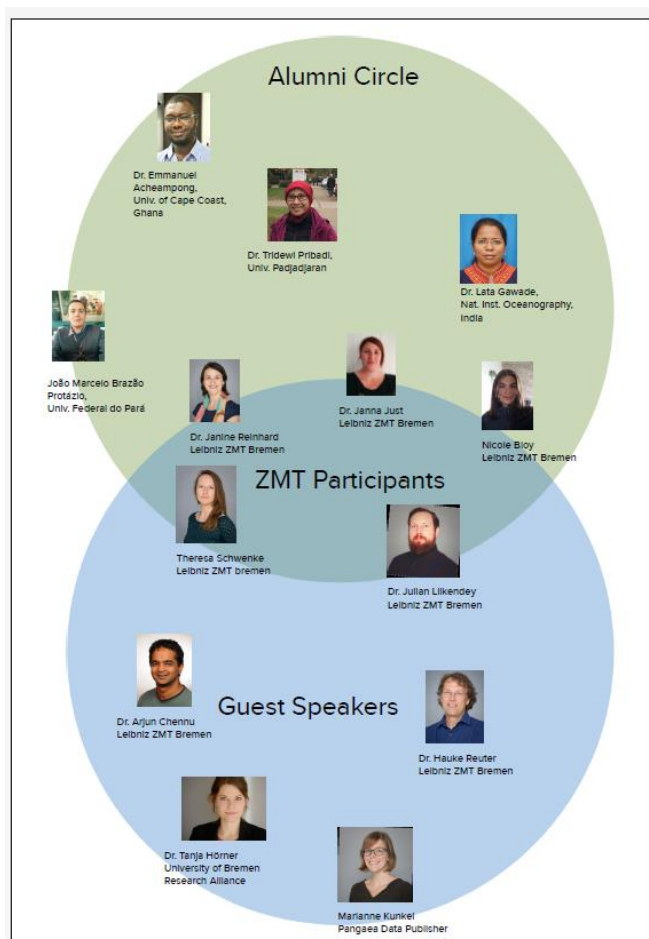


Figure 3: The participants collected ideas for the next steps in the DatAlumni project, to support data management strategies of ZMT Alumni.



We thank all participants and speakers of the Workshop of the Alumni Circle

