6th Data Science Symposium
8th/9th November 2021
Haus der Wissenschaft, Bremen

Abstracts

with contributions from

The abstracts are given in aphabetical order.

Program Committee: Stephan Frickenhaus, Bernadette Fritzsch, Antonie Haas, Tilman Dinter
Co-Organization: Heike Lipka-Nixdorf

Date	Time	Speaker	Centre	Title
8.11.2021	**13:00**	**Welcome and Opening**		**Stephan Frickenhaus**
		Artificial Intelligence/ Machine Learning in ESS, Chair: Stephan Frickenhaus		
	13.00-13.30	Keynote *Markus Reichstein*	Max Planck Institut für Biogeochemie, Jena	Machine-learning-model-data integration for a better understanding of the Earth System
	13.30-13.45	Talk 1 Andreas Dietz	DLR	Status of the Artificial Intelligence for Cold Regions (AI-CORE) project - data integration and method,implementation on the infrastructures
	13.45-14.00	Talk 2 David Greenberg	hereon	Machine Learning Surrogates for Physical Earth Science Models: Conceptual and Infrastructure Challenges
	14.00-14.15	Talk 3 Felipe de Amorim	AWI	Evaluation of Machine Learning Predictions of a Highly Resolved Time Series of Chlorophyll-a Concentration
	14.15-14.30	Talk 4 Sebastian Primpke	AWI	A tool for the harmonized analysis of microplastics: Systematic Identification of MicroPLastics in the Environment (siMPle)
14:30 Coffee break				
		Remote Sensing and Big Data Science, Chair: Tilman Dinter		
	15.00-15.30	Keynote *Brice Mora*	Communications & Systèmes, Toulouse	WEkEO, the DIAS platform powered by EUMETSAT, ECMWF, MERCATOR OCEAN, and EEA
	15.30-15.45	Talk 1 Lars Kaleschke	AWI	Earth Observation (EO) - sea ice - present & future research and development requirements and needs
	15.45-16.00	Talk 2 Sebastian Mieruch	AWI	M-VRE: The MOSAiC Virtual Research Environment
	16.00-16.15	Talk 3 Karolin Thomisch	AWI	A progress report on OPUS - The Open Portal to Underwater Soundscapes to study sound in the global ocean
	16.15-16.30	Talk 4 Linda Baldewein	hereon	Coastal Pollution Toolbox – A PoF-IV Topic 4 product using data management techniques

16:30 Poster session & refreshing drinks
19:00 Dinner

Date	Time	Speaker	Centre	Title
9.11.2021	**9:00**	**Welcome and Opening**		**Stephan Frickenhaus**
		Software/Collaboration/Dataflows, MareHUB/DAM, Chair: Angela Schäfer		
	09.00-09.20	Keynote *Alexander Struck*	Humboldt University of Berlin	Research Software and its Engineers
	09.20-09.35	Talk 1 Philipp Sommer	hereon	Beyond FAIRness with the Model Data Explorer
	09.35-09.50	Talk 2 Julia Boike	AWI	Thawing Permafrost worldwide — A new app- and community-driven monitoring project
	09.50-10.15	Talk 3 Michael Schlund	GEOMAR	TSG data work flow from sensor to publication: status quo and future perspectives
		Robert Kopte	IFG Uni Kiel	Standardized treatment of ADCP underway data from ship to repository: Linked workflows for acquisition and quality control
	10.15-10.30	Talk 4 Angela Schaefer	AWI	MareHub & DAM-DM activities: research data work flows, SOPs, data portal and data viewers

10:30 Coffee break

		Sociocultural Dimensions of Digitalization in Earthsyrtem Sciences, Host: Nike Fuchs		
	11.00-11.15	Talk 1 Peter Braesicke	KIT	Social aspects of digitalisation pt 1: NFDI4 Earth
	11.15-11.30	Talk 2 Gauvain Wiemer	DAM	Social aspects of digitalisation pt 2: Community building and FAIR principles
	11.30-11.45	Talk 3 Jens Greinert	GEOMAR	Social aspects of digitalisation pt 3: Digital Earth
	11.45-12.30	Podiumsdiskussion / Nike Fuchs: Peter Braesicke, Gauvain Wiemer, Jens Greinert, Alexander Struck		**Socio-cultural dimensions of digitalization in Earth system sciences:** *panel discussion with participation from the audience*

12:30 Closing and Lunch

Coastal Pollution Toolbox – A PoF-IV Topic 4 product using data management techniques

Linda Baldewein[1], Dr. Marcus Lange[1], Ulrike Kleeberg[1]
[1] Helmholtz-Zentrum Hereon

The Coastal Pollution Toolbox (https://www.coastalpollutiontoolbox.org/index.php.en) is developed within PoF-IV Topic 4 as a digital working environment to assess the dynamics of carbon, nutrients, and contaminants within the coastal zone. The Toolbox allows to investigate the propagation, concentration, and mitigation strategies of emergent substances and prioritized pollutants via a digital atlas and various WebGIS products. The Coastal Pollution Toolbox (CPT) comprises of three compartments: science tools, synthesis tools and management tools. Data management techniques are used to take scientific results and refined tools of the science compartment, such as validated analytical data, scenario calculations or model results, and make them publicly available. The data will be curated according to the FAIR principles, stored in data bases, and made available in national and international portals such as the DataHub and the NDFI4Earth.
To reach different kinds of audiences for the scientific output, the results are edited for specific user groups and published on the CPT website. To transfer knowledge to the public, story maps describing pollution research results and analytical approaches are generated using ESRI Storymaps. Specific web applications, such as a decision support tool in case of accidental oil spills, are generated by software developers using the data prepared and published by the data management team. Various web maps, dashboards and WebGIS applications are used to showcase specific results for different user groups, e.g. for campaign planning, urban air quality forecasts or general information on microplastics and analytical measures for detecting them. Some of these tools are co-designed with potential users. A foresight and user-centered design strategy has been implemented to consider specific user interests from the offset of the tool development.
References:

- Proposal for the Helmholtz Research Program Changing Earth – Sustaining our Future Research Field EARTH & ENVIRONMENT 2021 – 2027 (2019). Helmholtz Research Centers Earth & Environment
- Coastal Pollution Toolbox (2021). https://www.coastalpollutiontoolbox.org/index.php.en

Social aspects of digitalization - PT 1: NFDI4Earth

Peter Braesicke (IMK-ASF)
KIT, Herrmann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen

All of us wear many different hats in our professional life. Thus, being part of a professional community and how we perceive the community is shaped by our different roles - the hats we wear. In my case the perspective I have on the Earth System Sciences (ESS) community is shaped by being professor for meteorology at KIT, having a history as an atmospheric chemistry-climate modeller, participating in the Helmholtz PoF IV Program "Changing Earth", and co-chairing the DataHub of the Research Field (RF) Earth and Environment – and, in addition, contributing to the digitalisation effort of the RF. Thinking about FAIR Research Data Management (RDM) in ESS in the context of all the above is obviously only one of many valid viewpoints. Thus, when we try to build a big community effort around the National Research Data Infrastructure (NFDI) it is important to gather as many different viewpoints as possible to generate a large community buy-in. In this respect the NFDI4Earth already brought together many ESS actors in Germany, knowing that the consortium had to stay open to allow more actors to join after the start of the effort. Now, NFDI4Earth is a funded consortium coordinated by TU Dresden. In this respect, the first of the four task areas (2Participate) in NFDI4Earth enables the growing of the community with pilots, incubators, an academy, and diverse educational measures. 2Facilitate and 2Interoperate (task areas two and three) form a "tree" in which the many different user perspectives form the leaves that come together in branches (e.g. disciplines in ESS) and combine to the trunk that supports the tree with the OneStop4All and the User Support Network (reacting to user needs). The roots can be thought of as a synthesized architecture including a knowledge hub. Of course, something as complex as the NFDI4Earth requires a continuous coordination effort (2Coordinate, task area four) that also deals with the governance and general community support.
On this basis an inclusive, welcoming environment is created and maintained that will continuously develop to foster FAIR RDM within the ESS community and beyond, and novel data science building onto it.

Status of the Artificial Intelligence for Cold Regions (AI-CORE) project - data integration and method implementation on the infrastructures

Andreas Dietz, DLR

Funded by the Helmholtz Foundation, the aim of the Artificial Intelligence for COld REgions (AI-CORE) project is to develop methods of Artificial Intelligence for solving some of the most challenging questions in cryosphere research. These use cases are challenging due to diverse, extensive, and inhomogeneous input data and their high relevance is given in the context of climate change. In a collaborative approach, the German Aerospace Center, the Alfred-Wegener-Institute, and the Technical University of Dresden work together to address not only the methodology of how to solve these questions, but also how to implement procedures for data integration on the infrastructures of the partners. The presentation will give a brief overview of the geoscientific use cases and then address the different challenges that emerged so far in this still on-going project in terms of data integration. Four use cases have been elaborated, which led to four AI-driven solutions that need to be implemented on the infrastructures of the three partners. We will give an overview of the status of this

implementation and demonstrate the already available functionalities. The ultimate goal it to develop a toolbox of available AI methods that can be used by the whole Helmholtz community on their infrastructures.

Machine Learning Surrogates for Physical Earth Science Models: Conceptual and Infrastructure Challenges

David Greenberg, Marcel Nonnenmacher, Hereon

From the perspective of machine learning, Earth science presents a range of exciting new problems to be addressed with both established emerging techniques.
However, the size, complexity and many idiosyncracies of simulation and observation datasets present a host of new challenges for ML, requiring innovative new solutions all the way from theory to hardware. I will describe several ongoing projects in the model-driven machine learning group at Hereon, seeking to develop ML proxies that are faster or more accurate than existing simulation codes for physical processes. I will also discuss the infrastructure challenges related to scaling up ML training on modern HPC systems, and how these challenges are being addressed by the Helmholtz AI support team.

Standardized treatment of ADCP underway data from ship to repository: Linked workflows for acquisition and quality control

Robert Kopte[1], Maximilian Betz[2], Carsten Schirnick[3], Marianne Rehage[4], Gauvain Wiemer[5], Marius Becker[1]
1 Institute of Geosciences, Kiel University and Kiel Marine Sciences KMS
2 Alfred-Wegener-Institute, Helmholtz-Centre for Polar and Marine Research
3 GEOMAR, Helmholtz Centre for Ocean Research Kiel
4 Marum, Centre for Marine Environmental Sciences, PANGAEA
5 German Marine Research Alliance

Upper-ocean current measurements by shipboard Acoustic Doppler Current Profilers (ADCPs) contribute substantially to the understanding of key oceanographic processes on global to local scales. They are of importance both for global climate studies and the investigation of regional ocean dynamics. Being a part of the underway-research data project of the German Marine Research Alliance (DAM), we aim at the establishment and homogenization of semi-automated underway ADCP data streams for the four large German research vessels Maria S. Merian, Meteor, Sonne, and Polarstern. On board, this includes the implementation of standard workflows for data acquisition using underway standard configurations, the integration of remote near-realtime monitoring of data quality and the organization of a reliable transfer of ADCP raw data from the ships to the DAM data space. There, ADCP raw data are processed and quality controlled following international standards. Final datasets are prepared as netCDF and ascii versions including automated processing report and abstract generation tailored for publication in the PANGAEA data repository. Future steps include the development of an appropriate quality flagging scheme and the visualization of ADCP data in the marine data portal.

Evaluation of Machine Learning Predictions of a Highly Resolved Time Series of Chlorophyll-a Concentration

Felipe de Luca Lopes de Amorim, AWI

Pelagic chlorophyll-a concentrations are key for evaluation of the environmental status and productivity of marine systems, and data can be provided by in situ measurements, remote sensing and modelling. However, modelling chlorophyll-a is not trivial due to its nonlinear dynamics and complexity. In this study, chlorophyll-a concentrations for the Helgoland Roads time series were modeled using a number of measured water and environmental parameters. We chose three common machine learning algorithms from the literature: the support vector machine regressor, neural networks multi-layer perceptron regressor and random forest regressor. Results showed that the support vector machine regressor slightly outperformed other models. The evaluation with a test dataset and verification with an independent validation dataset for chlorophyll-a concentrations showed a good generalization capacity, evaluated by the root mean squared errors of less than 1 µg L^{-1}. Feature selection and engineering are important and improved the models significantly, as measured in performance, improving the adjusted R^2 by a minimum of 48%. We tested SARIMA in comparison and found that the univariate nature of SARIMA does not allow for better results than the machine learning models. Additionally, the computer processing time needed was much higher (prohibitive) for SARIMA.

M-VRE The MOSAiC Virtual Research Environment

Sebastian Mieruch-Schnülle[1], Julia Freier[1], Mohamed Chouai[1], Felix Reimers[1], Merret Buurman[2], Irfan Khan[2], Marcus Paradies[3], Arne Osterthun[3]
[1] AWI, [2] DKRZ, [3]DLR

The M-VRE project aims at implementing and developing powerful methods and software tools for the MOSAiC consortium, the global climate community, and the general public for efficiently exploring, analyzing, and visualizing MOSAiC data in an online and user-friendly manner. Key feature is the easy access to interdisciplinary, large, complex and heterogeneous datasets. M-VRE will be part of AWI's O2A system and be built on AWI cloud infrastructure. Following web services will be implemented:
webODV provides interactive and interdisciplinary access to MOSAiC data to create publication ready visualizations and analyzes. Data Cubes will be set up to easily access large datasets using programmable interfaces, e.g. Jupyter Notebooks, Python or R.
DIVAnd (Data-Interpolating Variational Analysis in n dimensions) is a powerful interpolation tool and allows to create smooth and continuous fields in one, two, three or n dimensions from a set of observations. DIVAnd is operated via Jupyter Notebooks.
SalaciaML is an AI algorithm to support scientists in the

quality control procedures on arctic ocean temperature profiles.
The great advantages of the M-VRE are:

I. interdisciplinarity; by bringing data from different fields (ocean, ice, atmos, etc.) together, cross-disciplinary analyzes are facilitated
II. being online; users do not need to install any software on their local computers, and can use any operating system using a modern browser (e.g. Chrome, Firefox). Working in the M-VRE is possible from anywhere if an internet connection is available.
III. reproducibility; all analyzes and processing can be easily repeated by storing the processing protocols e.g. in Jupyter Notebooks or webODV views and logs.
IV. being collaborative and transparent; Jupyter Notebooks and webODV views and logs can be shared by users, strengthening the collaborations and can be transparently published together with research papers.

WEkEO, the DIAS platform powered by EUMETSAT, ECMWF, MERCATOR OCEAN, and EEA

Brice Mora, *Communications & Systèmes, 5 rue Brindejonc des Moulinais, 31506 Toulouse, France*

The WEkEO DIAS (Data & Infrastructure-Access-Service) platform is powered by EUMETSAT, ECMWF, MERCATOR OCEAN, and EEA. WEkEO facilitates access to Copernicus data and information from the Copernicus programme. By providing data and information access alongside processing resources, tools and other relevant data, WEkEO is able to boost user uptake, encourage innovation. Its offer is threefold: first, it offers in a harmonised and seamless way access to Copernicus data and information; the WEkEO's REST-based single protocol allows to scale and evolve your code, easily integrating all data sources from a virtual station, a Jupyter Notebook or a desktop application of your choice, and using homogeneous subsetting attributes. Second, it offers processing capabilities and tools based on standard cloud technologies, allowing you to render your own services and promote them. Infrastructure-, Platform- and Software- As A Service are available including ready to use virtual machine pre-configured for data access and standard tools. The third strong asset of WEkEO is the user support capitalizing on a recognized long-standing experience in training and education. This presentation aims to present also the latest developments of the WEkEO.

A tool for the harmonized analysis of microplastics: Systematic Identification of MicroPLastics in the Environment (siMPle)

Sebastian Primpke, AWI

The harmonization of microplastics analysis is one of the main research gaps for the analysis of microplastics due to the various steps involved ranging from sampling, work up/purification to analysis. Every of these steps has its own challenges and for the analysis part even different methods based on spectroscopy or thermoanalysis are available with different data quality and comparability. Among the spectroscopic methods, the state of the art analysis by spectroscopy via FTIR imaging allows the analysis of complete filter areas independent of human bias in a short time. Still, for the development of standardized operational protocols (SOPs), comparable data determination is hampered by the size of the generated datasets (ranging from 0.2 to >92 million spectra) and different manufacturer as well as commercial software solutions available. To overcome this challenge and allow the harmonization of data analysis we developed the tool siMPle. It allows the analysis of datasets measured on different instruments from various manufacturers. Here, every spectrum can either be selected individually or analyzed currently via two pipelines for the automated analysis, the original MPhunter- and the widely applied AWI developed automated analysis pipeline. It could be shown that even very large datasets can be handled by it with relative ease. The generated data was benchmarked and validated in accordance to the original automated analysis approach using Bruker OPUS to allow a harmonized comparison of results. In addition, it significantly reduces the calculation time from more >24 h down to 2 h for a reference data set containing 1 million spectra. This tool is made available as Freeware and allows the harmonization of MP data analysis for spectroscopic data for research. For the future, the implementation of further novel approaches using machine learning and artificial intelligence shall be investigated and suitable candidates added to the software.

MareHub & DAM-DM activities: research data workflows, SOPs, data portal, and data viewer

Angela Schäfer[1], Robin Heß[1] & MareHub/DAM-DM working groups, [1] AWI

The marine data consortia of the Earth and Environment Helmholtz initiative MareHub (AWI, GEOMAR, Hereon) and the data management project “Underway Data” of the Deutsche Allianz Meeresforschung (DAM) works with several cross-partner, thematic working groups on integration of institutional infrastructures, services and tools for basic data management workflows to enable the dissemination of FAIR research data and consequently to enhance data science. In the first pilot phase, these working groups design and visualize data products from oceanographic, seafloor, acoustic, bathymetric, video, imagery and sampling data up to underway and expedition data.
The whole range from establishing standard processing procedures, data system and storage solutions as well as data publishing and harvesting from assigned data repositories will lead to findable and accessible data in the marine data portal following the FAIR data principles. Furthermore, interactive data visualization services in the form of map-based viewers enable intuitive data exploration and access through time and space. For this purpose, in cooperation with the working groups, thematic viewers for ocean observations, seabed observations and ocean acoustic data are created. The use of established standards such as OGC services for these data products is intended to

ensure interoperability with other portals, viewers and software solutions. In addition, the viewer implementations themselves are based on existing open source solutions and are also published as a modular and reusable open source framework.

TSG data workflow from sensor to publication: status quo and future perspectives

Michael Schlundt[1], Claas Faber[1], Marianne Rehage[2], Gauvain Wiemer[3], Hela Mehrtens[1]
[1] GEOMAR, [2] University of Bremen (MARUM), [3] German Marine Research Alliance (DAM)

Thermosalinograph (TSG) observations of near-surface temperature and salinity obtained along vessel tracks are a large contribution to the near-surface hydrographic monitoring of the global oceans. TSGs are used for a long time and they are existent on almost every research vessel around the globe. However, for a scientific usage of TSG data a standardized data treatment with reproducible workflows is necessary, to eventually get publicly accessible data, which fulfill all conditions of FAIR data policies. As part of the German Marine Research Alliance (DAM) project “Underway research data” we aim to establish all required steps of TSG data workflows, initially on the four large German research vessels Maria S. Merian, Meteor, Polarstern, and Sonne. This includes also the daily NRT data transmission of processed TSG data to the European operational data center CORIOLIS. Full datasets are postprocessed after the individual cruise and have to pass through several quality control checks. If applicable, the data are calibrated against water samples or other available independent data like CTD observations. Final datasets, the so-called delayed-mode data, are afterwards published in a standard format on PANGAEA along with a detailed documentation about all applied steps. The whole workflow is as automatic as possible, and as manual as necessary. All published data should be easily accessible in the future via the dedicated marine data portal marine-data.de.

Beyond FAIRness with the Model Data Explorer

Philipp S. Sommer[1], Linda Baldewein[1], Housam Dibeh[1], Hatef Takyar[1], Max Böcke[1], Rehan Chaudhary[1], Ulrike Kleeberg[1], Tilman Dinter [2]
[1]Hereon, [2]AWI

Making Earth-System-Model (ESM) Data FAIR is challenging due to the large amount of data that we are facing in this realm. The upload is time-consuming, expensive, technical, and every institution has their own procedures. It can be difficult for a scientist to find the data of interest.
Non-ESM experts face even more problems. FAIR principles (https://www.go-fair.org/fair-principles) are restricted to metadata, and therefore limited to a small scientific community. Pure data portals are therefore hardly usable for inter- and trans-disciplinary communication of ESM data and findings, as this level of accessibility often requires specialized web or computing services.
With the Model Data Explorer, we want to go one step beyond the FAIR principles. We want to simplify the generation of web services from ESM data, and we want to automatically link and acknowledge the data to involved parties, individuals, or datasets, in a decentralized federative manner. Our framework makes ESM data better findable and makes important findings easier to integrate in trans-disciplinary communications.
We build our decentralized framework upon three major parts: A web portal to access the individual datasets, a configuration interface where scientists configure the appearance and linkage of datasets, and a framework for an efficient remote processing of distributed ESM data. Our approach is scientists first! We aim for a framework where web-based communication of model-driven data science can be maintained by the scientific community. With fair reward for the scientific work and adherence to the FAIR principles, and without too much overhead and loss in scientific accuracy.
The Model Data Explorer is in the progress of development at the Helmholtz-Zentrum Hereon, together with multiple scientific and data management partners in the dataHub and DAM communities. The full list of contributors is constantly updated and can be accessed at https://model-data-explorer.readthedocs.io.

Keynote - Research Software and its Engineers

Alexander Struck, Cluster of Excellence "Matters of Activity",
Humboldt-Universität zu Berlin

The talk sheds light on the issues software and the developers are facing. Current research often requires software tailored to solve specific problems. Without it, a major part of our research endeavor would not be possible. Despite its importance, research software and its engineers lack the recognition they deserve. As an example, software written during research is still considered as a part of research data although different FAIR principles, licenses and archiving strategies apply. Research software engineers are domain experts and create the necessary tools but may lack the skill set and incentives to make software more robust, reliable and sustainable. These problems are intensified for 'end-user developers' with short-term, often project-based contracts. Several organizations lobby for better recognition of these researchers and to improve their handling of research software. Among them are the Software Sustainability Institute in the UK and US, and hopefully one for Europe in the future. The FAIR4RS working group at the Research Data Alliance (RDA) works on a defintion of Research Software and adoption of the principles. The Research Software Alliance (ReSA) and several national organization, like de-RSE.org or the Research Software Working Group of the "Allianz-Initiative" engage with the topic. The community published policies and recommendations to be implemented in the future. Funding may play a role here but is currently too often restricted to newly written software which leads to duplication of effort and decreases reuse. If software is reused it is rarely cited as it is not yet considered a citable item. The Software Citation Principles and the two format specifications Citation-File-Format (CFF) and CodeMeta, both adopted by GitHub and Software Heritage respectively, shall make citing software easier and more widespread. If software is mentioned in publications it is often inaccessible

which can be caused by depublication. Here, the choice of an appropriate publishing platform is essential to achieve long-term availablity and to reach the relevant audience. Often, research software is not published at all due to intellectual property conflicts, licensing issues, quality concerns, competition, lack of resources or sensitive content. The abundance of possible software locations complicates discovery. In many cases, a general purpose web search engine, a social network or the domain literature are consulted to find reusable software. Assumptions about the absence of relevant solutions or the not-invented-here syndrome forestall efforts to discover existing code or applications. As it is still difficult to find software we experience duplication of work. A lack of resources impacts proper searches and the evaluation of found software. It's not yet reasonable to expect journal reviewers to evaluate software submissions although it seems necessary. Anecdotal evidence from few research software reviews raises questions about the overall quality, reliability and reproducibility of results computed by such software. Especially the disagreement in results of different implementations of the same use-case solution are worrysome. Trust in third-party (research) software is sometimes misplaced as the example of random number generators (RNG) shows. Sampling, clustering, simulations and other methods utilize RNGs and seed numbers are supposed to make results reproducible which has been documented to fail. Sufficient randomness is essential for trust in computational results, yet several popular research software applications, e.g. MATLAB, Mathematica, NumPy, SPSS or GNU R, failed randomness tests. Several aspects of our research software infrastructure need to be improved to foster FAIR software, Open Science and the community of research software engineers.
6thDataScienceSymposium_Struck_keynote_ResearchSoftwareEngineering.pdf

A progress report on OPUS - The **O**pen **P**ortal to **U**nderwater **S**oundscapes to study sound in the global ocean

Karolin Thomisch[1], Michael Flau[1], Robin Heß[1], Andy Traumüller[1], Olaf Boebel[1]
[1] AWI

Facing an era of rapid anthropogenically induced changes in the world oceans, ocean sound is considered an essential ocean variable (EOV) for understanding and monitoring long-term trends in anthropogenic sound and its effects on marine life, biodiversity and ecosystem health.
The International Quiet Ocean Experiment (IQOE) has identified two major research interests in the context of monitoring the distribution of ocean sound in space and time: i) estimating current levels and distribution of anthropogenic sound in the ocean, and ii) assessing trends in anthropogenic sound levels across the global ocean, with recommendations on ambient noise monitoring being also part of the EU Marine Strategy Framework Directive. To address these research foci by international collaborative research efforts, the OPUS (**O**pen **P**ortal to **U**nderwater **S**oundscapes) data portal is currently being developed by the Ocean Acoustics Group of the Alfred Wegener Institute (AWI), financially supported by the MAREHUB initiative.
Furthermore, an Ocean Sound Software for **M**aking **A**mbient **N**oise **T**rends **A**ccessible (MANTA) will be provided to data owners of underwater passive acoustic data worldwide to generate standardized ocean sound level data products from passive acoustic recordings according to IQOE Guidelines.
OPUS will accept MANTA-processed data together with related metadata and make them accessible under customized licensing policies via a map- and time-based selection tool and shopping basket functionality. Data products including the compiled MANTA data, parameter-naming conventions, instructions for citing the data, and other information necessary for data use according to FAIR standards will be regularly produced by OPUS. OPUS is envisioned to promote the use of acoustic data collected worldwide, thereby contributing to an improved understanding of the world's oceans soundscapes and anthropogenic impacts thereon over various temporal and spatial scales.

Research | data management: some observations made at the interface

Gauvain Wiemer, German Marine Research Alliance (DAM)

The German marine Research Alliance (DAM) is dedicated to the FAIR and open data principles. The core area "data management and digitalization" of the DAM is a coordination platform which aims at promoting FAIR and open data principles and supporting marine scientist in their data management activities. Additionally, the DAM coordinates the project "Underway"-Data which provides FAIR and open bathymetric and oceanographic "underway"-data in cooperation with chief expedition scientist on bord MARIA S. MERIAN, METEOR, POLARSTERN and SONNE.
In this talk, Gauvain Wiemer, responsible for the DAM core area "Data Management and Digitalization" shares his opinion formed by experiences made in three "fields of experience" at the interface of the research and data management communities: project management, research data management workshops and the promotion of data management activities within the DAM research missions. In his opinion data management is gaining in attention within the research community. However, some substantial improvement is needed in co-design of data infrastructures and co-work in data management activities of the two communities in order to provide open access to truly FAIR, distributed data also widely suitable for the application of AI-methods in the future. This talk is an impulse for the following panel discussion.

Panel Discussion and Resumé

Nike Fuchs, AWI
Peter Braesicke, KIT
Jens Greinert, GEOMAR
Angela Schäfer, AWI
Alexander Struck, Humboldt University of Berlin
Gauvain Wiemer, German Marine Research Alliance (DAM)