Leibniz Institute for Baltic Sea Research Warnemünde

MV, Germany
ORDS
Open Reproducible Data Science and Statistics

ORDS-MV MEETS THE SEA

JOIN US FOR A FULL-DAY WORKSHOP AND NETWORKING EVENT FOCUSED ON MARINE AND GEOSPATIAL RESEARCH DATA

14.10.2024
Leibniz Institute for Baltic Sea Research Warnemünde
9:30am to 5pm
Registration: stat-consult@fbn-dummerstorf.de

PROGRAM

9:30 WELCOME
Welcome at IOW

9:45 WORKSHOP: GEOREFERENCED OCEANOGRAPHIC DATA RESOURCES AT IOW
Dr. Susanne Feistel *(Leibniz Institute for Baltic Sea Research Warnemünde (IOW))*

11:00 HANDS-ON WORKSHOP: MAKING MAPS WITH OPEN DATA AND OPEN SOURCE SOFTWARE: R AND QGIS - PART 1
Dr. Anja Eggert *(Research Institute for Farm Animal Biology (FBN))*

12:00 KEYNOTE: NATIONAL DATA FOR INTERNATIONAL SCIENCE - THE EXAMPLE OF FISHERIES RESEARCH
Dr. Christian von Dorrien *(Institute of Baltic Sea Fisheries, Thünen Institute)*

13:00 LUNCH BREAK
Including a guided tour through the IOW exhibition for interested parties

14:00 HANDS-ON WORKSHOP - PART 2
Dr. Anja Eggert *(Research Institute for Farm Animal Biology (FBN))*

15:15 DEMO: FOUNDATION ODELS: REPOSITORIES, APIS AND FINE-TUNING
CANCELLED
JProf. Stefan Lüdtke *(University of Ro Science)*
Stefan Oehmcke: topic tba

15:45 FAREWELL

16:00 SOCIALIZING WITH PIZZA

14.10.2024
Leibniz Institute for Baltic Sea Research Warnemünde
9:30am to 5pm
Registration: stat-consult@fbn-dummerstorf.de

WORKSHOP

GEOREFERENCED OCEANOGRAPHIC DATA RESOURCES AT IOW

The Oceanographic Database of the IOW (IOWDB) was originally created for the internal needs of the Leibniz Institute for Baltic Sea Research. It includes historical and recent oceanographic data, collected since 1949, including over 88 million data points from 985 research campaigns. These data, ranging from 1877 to 2024, cover measurements such as CTD profiles, hydrochemical and biological samplings, and long-term monitoring. To provide public access, the ODIN2 tool was launched in 2018, offering a user-friendly online interface with advanced search options. All data is licensed under CC BY 4.0 and can be exported in various digital formats. In this workshop we will explore ODIN2 to download oceanographic data from IOWDB, which we will then import into R and QGIS for visualization.

Dr. Susanne Feistel works as a data steward and software developer in the IT and data management group at IOW.

14.10.2024
Leibniz Institute for Baltic Sea Research Warnemünde
9:30am to 5pm
Registration: stat-consult@fbn-dummerstorf.de

HANDS-ON WORKSHOP

MAKING MAPS WITH OPEN DATA AND OPEN SOURCE SOFTWARE: R AND QGIS

We will explore the quickest ways to access and visualize geospatial marine data from open repositories using R. With a focus on showing how geospatial data can be integrated in a familiar tidyverse data-analysis workflow, we will demonstrate how R can be used for making maps. However, while R is powerful for data processing, QGIS could offer more intuitive mapping capabilities. The choice of tools is yours! We will use real-world examples from Baltic Sea environmental and fisheries data, and you will have the option to participate in hands-on exercises. This workshop is designed for R users at a basic level or those curious about what R can do, particularly in incorporating spatial data into their everyday workflows. Advanced GIS topics will not be covered.

Dr. Anja Eggert works as a Statistical Consultant at the Research Institute for Farm Animal Biology (FBN).

14.10.2024
Leibniz Institute for Baltic Sea Research Warnemünde
9:30am to 5pm
Registration: stat-consult@fbn-dummerstorf.de

KEYNOTES

NATIONAL DATA FOR INTERNATIONAL SCIENCE - THE EXAMPLE OF FISHERIES RESEARCH

For over 100 years, the countries bordering the North Atlantic have recognized that the sustainable use of fish stocks can only work in an international cooperation. The Thünen Institute of Baltic Sea Fisheries contributes significantly to this by collecting "fishery-independent" data on economically important fish stocks such as herring, sprat and cod on German standard research cruises. This data is quality-checked and then uploaded to databases at the International Council for the Exploration of the Sea (ICES) so that it is available to the international research community.

In addition, the Thünen Institute of Baltic Sea Fisheries has many data sets that are collected as part of other scientific studies, but which are currently not easily found and available. The Thünen Institute is therefore currently developing its own research data management system to make this data available to the research community.

Dr. Christian von Dorrien

14.10.2024

Leibniz Institute for Baltic Sea Research Warnemünde

9:30am to 5pm

Registration: stat-consult@fbn-dummerstorf.de

KEYNOTES

FOUNDATION MODELS: REPOSITORIES, APIS AND FINE-TUNING

The availability of large pre-trained models (e.g., large language models, but also large pre-trained models for images or environmental data) leads to a paradigm change in machine learning: Instead of training new models from scratch for each task, researchers can now reuse and fine-tune large pre-trained models, often resulting in the need for training only providing impressive results. In this talk I will show how the Hugging Face ecosystem and the transformers library can be easily used to run and fine-tune models. I will also briefly introduce for geospatial modeling foundation models.

CANCELLED

Instead you will listen to an interesting presentation by Stefan Oehmcke

Stay tuned!

Stefan Oehmcke:
Juniorprofessur Visual and Analytic Computing in Ocean Technologies.
University Rostock

14.10.2024
Leibniz Institute for Baltic Sea Research Warnemünde
9:30am to 5pm
Registration: stat-consult@fbn-dummerstorf.de

ORDS-MV NETWORK

Open Reproducible Data Science and Statistics (ORDS) is a scientific network in Mecklenburg-Vorpommern. Its goal is to bundle regional expertise in the fields of data analysis and statistics with open and reproducible science. The focus is on the exchange of expertise between doctoral candidates and postdocs, but all other interested scientists are also welcome.

Due to the interdisciplinary character of modern data-driven science, the network explicitly addresses all disciplines. Besides general questions of data analysis, statistics and reproducibility, the network also focuses on programming environments such as R and Python. While R is mainly used for statistical data analysis and data visualization, Python is especially common in the application of machine learning methods. Besides the actual data analysis, another central point is the management and versioning of data and source code. Modern tools such as the Git version control system, Jupyter notebooks and Docker containers can be used for this purpose.

We want to meet at regular intervals and inform each other about current projects and methods in the field of data analysis and statistics. Furthermore, we will organize seminars, workshops and other events.

More details:
https://ords-mv.github.io
Mailing list:
https://www.listserv.dfn.de/sympa/subscribe/ords-mv