

# Data Train – The Cross-disciplinary Training in Research Data Management and Data Science

## U Bremen Research Alliance

Program speakers: Prof. Dr. Rolf Drechsler<sup>1</sup>, Prof. Dr. Frank Oliver Glöckner<sup>2</sup>, Prof. Dr. Iris Pigeot<sup>3</sup>

Coordinators: Dr. Tanja Hörner ⁴, Dr. Maya Dalby⁴

Dealing with data and innovative technologies are key competencies of our time. To meet the massive demand in science and economy, the cross-discipline training program 'Data Train - Training in Research Data Ma- nagement and Data Science' of the U Bremen Research Alliance (UBRA) teaches competencies in data literacy, i. e. in research data management (RDM) and data science, for doctoral researchers. The program is associated with the German National Research Data Infrastructure (NFDI) (Fig. 1).



Figure 1: Inside the rectangle: The Federal State of Bremen and the me Figure 1: Inside the rectangle: The recers state of premen and the members of the cooperation network of the Bremen Research Alliance (UBRA). Outside the rectangle: NFDI consortia in which member institutions of the UBRA are currently involved (September 2023). The consortia and the members of the cooperation network contribute to the development of the Data Train program and the training. ©UBRA

#### The Curriculum

A team of enthusiastic lecturers from member institutions of the UBRA (Fig. 1) voluntarily created a flexible and cross-disciplinary curriculum, with lectures tailored towards data analysis, data science and RDM. The training is targeted at doctoral researchers and open to all interested in strengthening their data skills, if capacities are available (Fig. 2).

### STARTER TRACK

In the 'Starter Track', thematic overview lectures provide a general knowledge of important topics in research data management, data science and various questions all about data. This track is held online or in hybrid format and is thus open to evervone

### **OPERATOR TRACK 'DATA STEWARD'**

In this track, essential competencies for the efficient handling of data in accordance with the FAIR data principles (i. e. data should be Findable, Accessible, Interoperable and Reusable) are taught. The track includes hands-on workshops on programming languages, data management plans, reproducibility, data preparation, data provisioning, and database skills

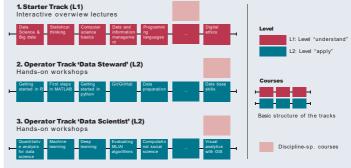


Figure 2: Schematic illustration of the curriculum's concept with exemplary courses from the past Figure 2: Schematic illustration of the curriculum's concept with exemplary courses from the past two years. Boxes symbolize individual courses within a track. Starter Track courses (red boxes) are overview lectures in the first half of the year (level "understand"). The operator track (blue boxes) is made of hands-on workshops taking place in the second half of the year. It is divided into "Data Steward" and "Data Scientist", which alternate every two years, i.e. one year "Data Scientist", the next year "Data Scientist" workshops. Light red boxes stand for discipline-specific courses that are offered in the scientific domains or the NFDI consortia building on the basics taught in the Data Train program. @Modified after Hörner et al. 2021s

#### **OPERATOR TRACK 'DATA SCIENTIST'**

Participants learn methods from mathematics, statistics, artificial intelligen- ce, and computer science basics for data analysis as well as data visualization skills. Data science is about extracting information and gaining knowledge from data using computational analysis methods such as statistical techniques, but also artificial intelligence (e.g. machine learning, deep learning).

Doctoral researchers can complete the entire tracks or attend only individual lectures/workshops. Participation is confirmed after completion of the respective track by a 'Certificate of Participation' listing all attended courses.

#### The Schedule

The Starter Track takes place in the first half of the year. The two operator tracks take place in the second half of the year alternating every two years, i. e. 'Data Steward' in one year, 'Data Scientist' in the other year (Figs. 2 and 3).



#### **The Data Stories**

Data Stories are inspiring talks about data management and data science in the private sector and in academia. The invited speakers shed light on the importance of data skills with regard to their respective fields of work and discuss current challenges. Whenever possible, the events end with an opportunity to network and socialize with colleagues from other disciplines and the 'storytellers'. The Data Stories are open to everyone.



data-train@vw.uni-bremen.de / Phone: (+49) 421 – 218 60044 / X @UBRA\_HB

























<sup>\*\*</sup>University of Bremen, Faculty of Mathematics and Computer Science, Data Science Center / German Research Center Artificial Intelligence \*\*
Affided Wegener Institute, Helmholtz-Centre for Polar and Marine Research / University of Bremen, MARUM – Center for Marine Environmental Science / PANGAEA – Data Publisher for Earth & Environmental Science / GFBio e. V. / NFDI4Biodive \*\*
\*\*Labbiz: Institute for Prevention Research und Epidemiology – BiPS / University of Bremen, Faculty of Mathematics and Computer Science / NFDI4Health

Bremen Research Alliance / Univers tps://doi.org/10.17192/bfdm.2021.3.8343