



Nikolay Koldunov

# SOFTWARE USED IN TRR181

THINGS THAT WILL MAKE YOUR LIFE EASIER

AFTER INITIAL PERIOD OF SUFFERING AND FRUSTRATION ☺

Bremerhaven, 11.03.2021

- General productivity software.
- Specialized software (things that are used in research groups to do specialized tasks).
- Data handling/analysis/visualization software.



- Word processing (Google Doc/Word/OpenOffice Writer)
- Presentations (PowerPoint/OpenOffice Impress/Keynote)
- LaTeX - A document preparation system. Good place to start: <https://www.overleaf.com>



*"This CRC targets at two national climate models, **ICON-a/ICON-o** and **OpenIFS/FESOM**, which are with their modern unstructured grids at the forefront of the international model development. **All work will ultimately lead to improvements of these target models**, which therefore serve as a metric for the CRC success."*

At the end of the day our aim is to improve this two pieces of software ☺

- Written in FORTRAN programming language, so if you want your development to find it's way there it should probably be in FORTRAN.
- Created for massively parallel computers (tens of thousands of compute cores), so your solutions should work well in parallel mode and scale.
- Produce large amounts (Petabytes) of data, that some of you will work with. And you can't do this in Excel ☺



## General considerations:

Most of the time you probably will operate on arrays of numbers, plot graphs and maps. Only about 0.001% of this work will end up in your publications.

Better make good choice of the tools in the beginning and invest time to become proficient, it will pay off.

In general the recommended first choice is to use the same tools as your adviser/group use. However sometimes someone should start a revolution 😊

Invest in the skills that will be useful also in case you decide to leave academia.

In the field of climate modelling knowing basics of Unix/Linux is the must if you are doing any type of work related to models of model output. If you plan to work with large amounts of data it's probably a good investment for you too.



- Simple to start with and user friendly.
- Usually good for relatively simple tasks, like quick look at the results, or simple analysis.
- Often just black boxes for you.

Examples from climate modelling applications:

- **Cdo** <https://code.mpimet.mpg.de/projects/cdo/wiki/Cdo#Documentation>
- **Nco** <http://nco.sourceforge.net/>
- **Ncview** [http://meteora.ucsd.edu/~pierce/ncview\\_home\\_page.html](http://meteora.ucsd.edu/~pierce/ncview_home_page.html)
- **Panoply** <https://www.giss.nasa.gov/tools/panoply/>
- **Paraview** <https://www.dkrz.de/up/services/analysis/visualization/sw/paraview>



- There is relatively steep learning curve.
- You have control (however some things will still look like black magic).
- You can do things out of the box (this is how new and interesting science should work, right?).

## Main choices for data analysis in 2021:

- **Matlab** <https://www.mathworks.com/products/matlab.html>
- **Julia** <https://julialang.org/>
- **R** <https://www.r-project.org/>
- **Python** <https://www.python.org/>
- **Jupyter notebooks** (not programming language but environment for interactive data exploration) <https://jupyter.org/>



## DEMO TIME





- Version control
  - Code handling
  - Code testing
  - Code packaging and distribution
  - Testing software
  - Community management.
- 
- FAIR principle (Findability, Accessibility, Interoperability, and Reuse)
  - Software licenses
  - Data sharing
  - Data archiving
- 
- IDEs, code editors



**Management of research data and knowledge** strategy. ...we make use of different state-of-the-art platforms to share information, store data and collaboratively work on software.

The “**Data and Code Management**” **task group** coordinates and supervises the efforts associated with this data management plan.

Subproject **S1, WP3: “Code and data management”**. In this WP, we will implement a strategy that allows making the results of the S1 subproject easily accessible, reusable and reproducible by project partners and the wider scientific community. This includes code development (improvement of the core functionality, usability, maintainability) and code management (documentation, testing and distribution).

Subproject **S2, WP4: “Inter-model technology transfer and data management”**. ... the exchange is facilitated by a web-based exchange platform for code, documentation, and common evaluation software....



Best way to learn programming/software is to apply it to your problem.

Ask your colleagues.

Don't dive too much in hard core python, for data analysis you need only basics, put more emphasis on learning libraries.

Select Data Science oriented courses, they are more practical.

If you like videos, SciPy conferences have excellent tutorials on many libraries used in science. A lot of SciPy and other videos on <https://pyvideo.org/>





**U Bremen  
Research  
Alliance**

## Data Train

Training in Research Data Management  
and Data Science

Given the massive demand for qualified co-workers in **research data management and data science**, the [U Bremen Research Alliance](#) together with the Federal State of Bremen initiated this **cross-domain education program for doctoral candidates**. Data Train intends to strengthen competences in data science, research data management and data literacy.

### Kick-off event, 3rd March 21 at 3-5 pm

Doctoral candidates from the member institutions of the U Bremen Research Alliance are cordially invited to take part in our pilot-phase starting in March 2021. You will receive further information about Data Train, its concept, and insights into the preliminary curriculum for 2021 in our talks.

#### Data Train – Towards a cross-domain Education Program on Research Data Management and Data Science

Prof. Dr. Iris Pigeot (BIPS), Dr. Tanja Hörner

#### The Intersection of Data Science and Data Stewardship

Dr. Lena Steinmann, Prof. Dr. Rolf Drechsler (DSC)

#### Data Management of the MOSAiC Expedition

Prof. Dr. Frank Oliver Glöckner (AWI)

This event is open for everybody who is interested (all status groups). Please register via our [Data Train website](#).

**We are looking forward to welcoming you at the virtual Kick-off event!**

## STARTER TRACK

Start in March 2021

Wed	10.03. 10-12 am	<b>Data science &amp; big data</b> <i>Björn Tings (DLR)</i>
Wed	17.03. 10-12 am	<b>Philosophical reflections on data science</b> <i>Prof. Dr. Dr. Norman Sieroka (UB, F9)</i>
Wed	14.04. 9-12 am	<b>Data &amp; information management</b> <i>Prof. Dr. Frank Oliver Glöckner (AWI) &amp; Dr. Ivaylo Kostadinov (GFBio e.V.)</i>
Wed	21.04. 2-4 pm	<b>Data protection &amp; licenses</b> <i>Prof. Dr. Benedikt Buchner, Dr. Dennis-Kerji Kipker (UB, F6)</i>
Tues	27.04. 2-4 pm	<b>Managing confidential data</b> <i>Dr. Julia Gottschall, Dr. Martin Dörenkämper (Fraunhofer IWES)</i>
Thur	06.05. 9-12 am	<b>Managing qualitative data</b> <i>Prof. Dr. Betina Holstein, Dr. Jan-Ocko Heuer (UB, F8)</i>
Mon	10.05. 10-12 am	<b>Statistical thinking</b> <i>Prof. Dr. Iris Pigeot (BIPS)</i>
Tues	18.05. 9-12 am	<b>Asking the right research questions in data science</b> <i>Prof. Dr. Vanessa Didelez (BIPS)</i>
Wed	02.06. 10-12 am	<b>About the meaningfulness of data</b> <i>Prof. Dr. Hans-Christian Waldmann (UB, F11)</i>
Tues	08.06. 10-12 am	<b>Computer science basics</b> <i>Prof. Dr. Rolf Drechsler (DSC)</i>
Wed	16.06. 10-12 am	<b>Overview about programming languages</b> <i>Prof. Dr. Christoph Lüth (DFKI)</i>
Tues	22.06. 10-12 am	<b>Cryptography basics</b> <i>Prof. Dr. Dieter Hutter (DFKI)</i>
Tues	29.06. 10-12 am	<b>Security &amp; privacy</b> <i>Prof. Dr. Dieter Hutter (DFKI)</i>

Open online overview lectures

## OPERATOR TRACKS

hands-on workshops (2-3 days)

### DATA STEWARD TRACK

Start in July 2021

**Getting started with R**  
*Dr. Christian Fieberg (UB, F7)*

**Erste Schritte in MATLAB**  
*Dr. Christian Fieberg (UB, F7)*

**Python**  
*Dr. Nikolay Koldunov (AWI)*

**Tools for FAIR data handling**  
*Prof. Dr. Frank Oliver Glöckner (AWI),  
Dr. Ivaylo Kostadinov (GFBio e.V.)*

**Versioning**  
*Prof. Dr. Frank Oliver Glöckner (AWI),  
Dr. Ivaylo Kostadinov (GFBio e.V.)*

**Reproducibility in science – How and why?**  
*Dr. Arjun Chennu (ZMT)*

**Data preparation**  
*Prof. Dr. Betina Holstein,  
Dr. Jan-Ocko Heuer (UB, F8),  
Dr. Nikolay Koldunov (AWI),  
Dr. Bianca Kolthorst (BIPS),  
James Imber, Karl Kortum (DLR)*

**Data base skills**  
*Prof. Dr. Sebastian Maneth (UB, F3)*

**Data provisioning**  
*Prof. Dr. Kristina Klein (UB, F7),  
Dr. Brenner Silva (AWI)*

### DATA SCIENTIST TRACK

Start in October 2021

**Quantitative analyses**  
*Prof. Dr. Thorsten Dickhaus (UB, F3)*

**Causal learning**  
*Prof. Dr. Vanessa Didelez (BIPS)*

**Machine learning**  
*Prof. Dr. Marvin Wright (BIPS)*

**Deep learning**  
*Prof. Dr. Peter Maas (UB, F3)*

**Evaluation of ML/AI algorithms**  
*Prof. Dr. Werner Brannath (UB, F3),  
Dr. Max Westphal (Fraunhofer MEVIS)*

**Analyses of qualitative data & mixed methods**  
*Prof. Dr. Betina Holstein,  
Dr. Jan-Ocko Heuer (UB, F8)*

**Visualisation techniques**  
*Prof. Dr. Dr. Norman Sieroka (UB, F9),  
Dr. Antonie Haas (AWI)*

**Mapping**  
*Dr. Antonie Haas (AWI)*

Dates for the Operator Tracks are coming soon.  
Announcements and registration via our website.

<https://uni-bremen.de/data-train>

 Dr. Tanja Hörner  
Data Train Coordinator

 (+49) 421-218 56788

 data-train@vw.uni-bremen.de