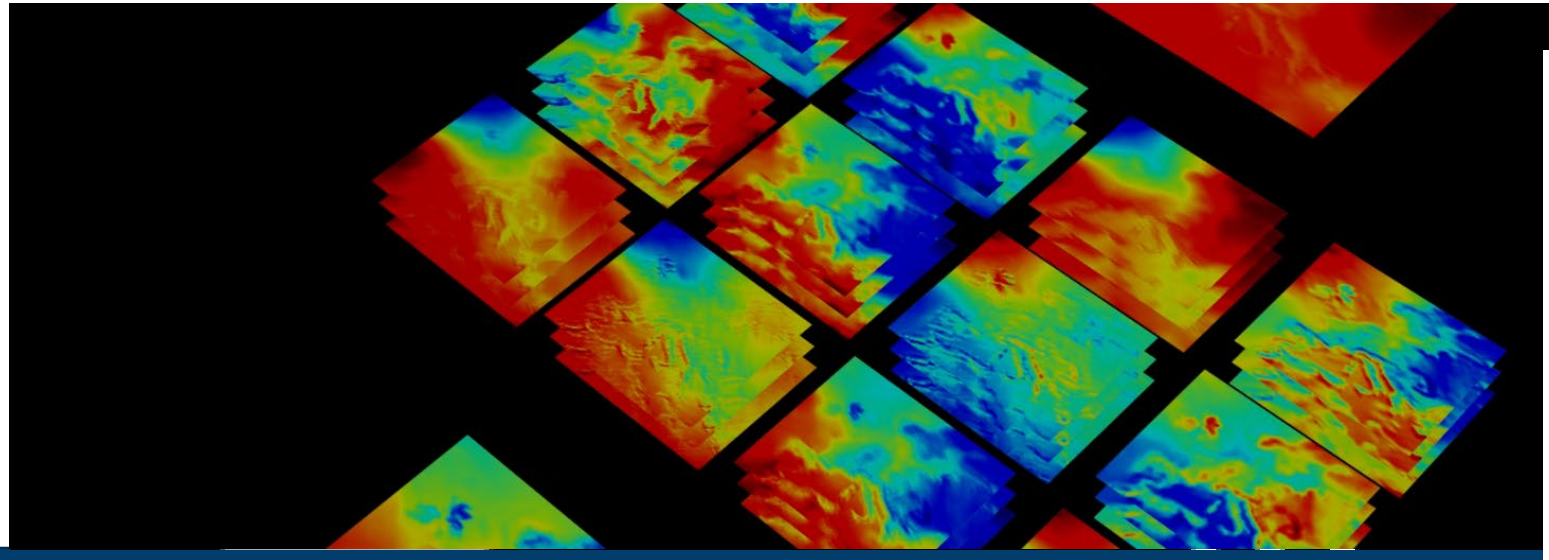


Martin Schultz

TEAM LEADER

0000-0003-3455-774X

GEODATA AT THE JÜLICH SUPERCOMPUTING CENTRE

Geo-IT Round Table
16 February 2021

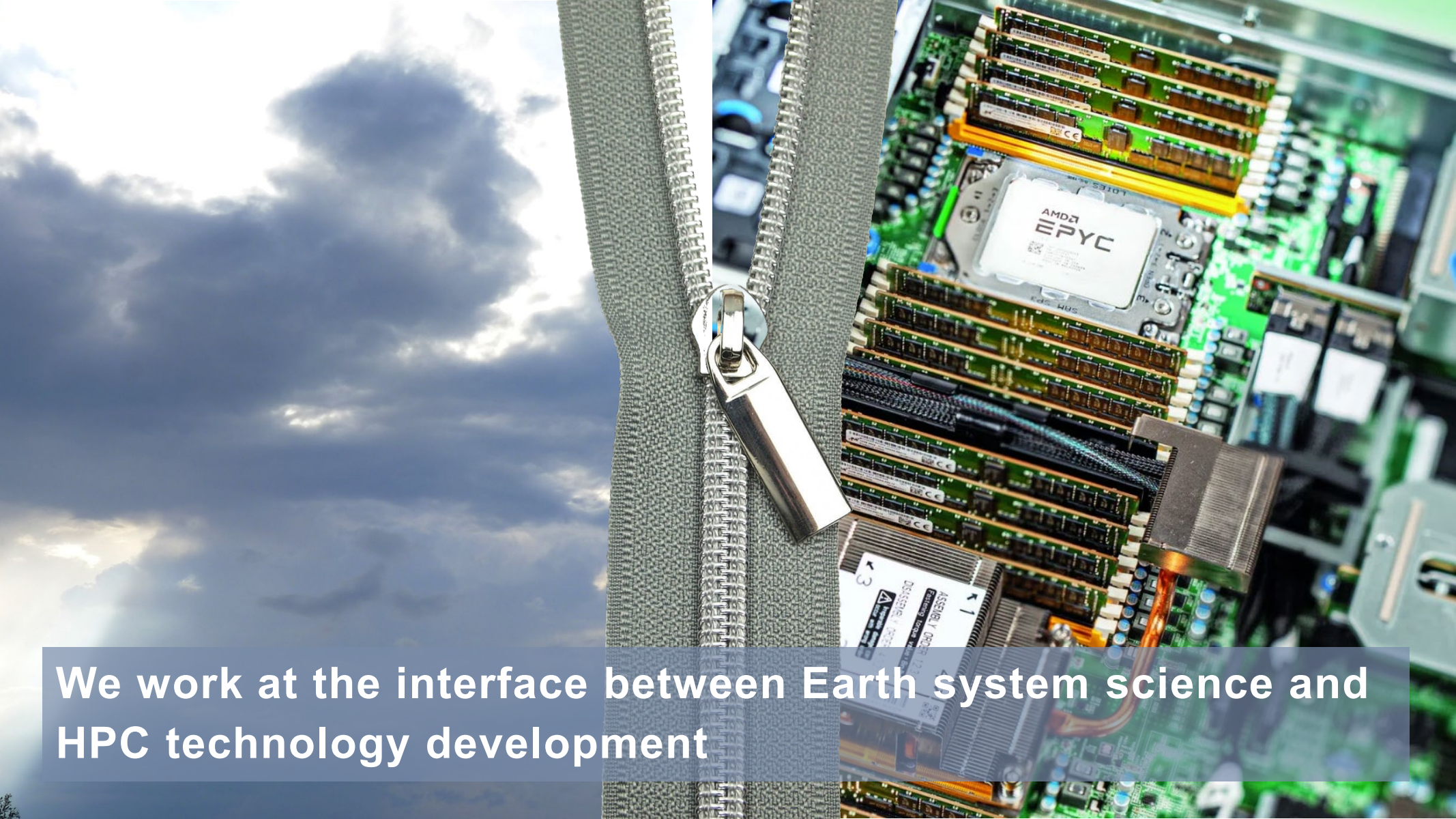
AKTUELLES

Supercomputer made in Jülich setzt neue Maßstäbe

Der Jülicher Supercomputer JUWELS erzielt dank eines neuen Booster-Moduls nun 85 Petaflops, was der Rechenleistung von mehr als 300 000 modernen PCs entspricht. JUWELS kann damit die Grenzen von

JUWELS Booster Porting Workshop 2021 (online)

Dieser Workshop hat die Portierung von bereits GPU-fähigen HPC-Anwendungscodes auf den neuen JUWELS-Booster mit mehreren Quad-GPU-Rechenknoten sowie die Inbetriebnahme von Code, der noch nicht für die Nutzung von (mehreren)



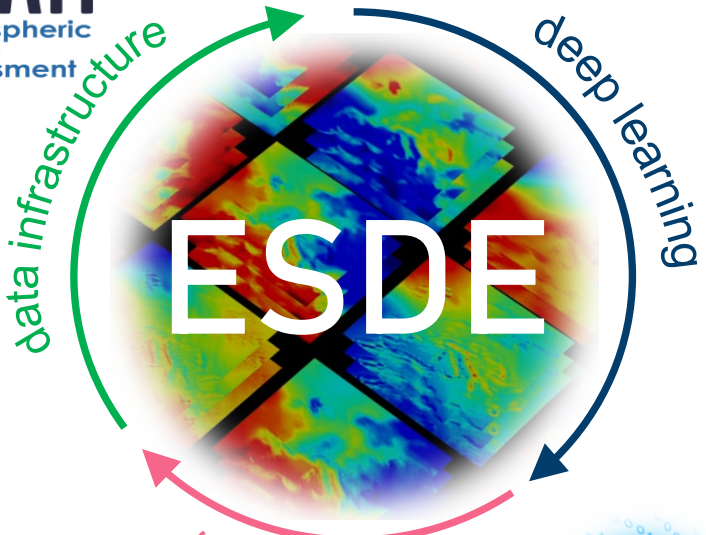
We work at the interface between Earth system science and HPC technology development

The Earth System Data Exploration Group

EUDAT Collaborative Data Infrastructure
Data shared and preserved across borders and disciplines

TOAR tropospheric ozone assessment report

IntelliAQ



KI:STE

HI-CAM (with IBG-3)

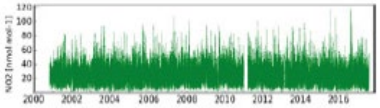
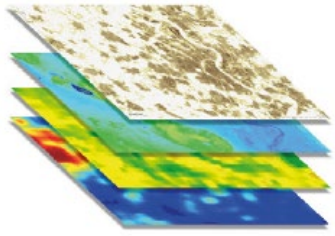
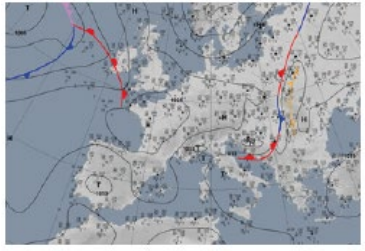
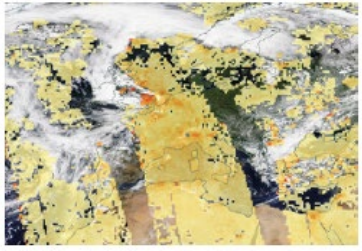
Bioökonomierevier
DAIS-RR (with IBG-3)

DeepRain

ExaESM
Pilot Lab Exascale
Earth System Modelling

MAELSTROM

Our data

Data type				
Category	Time series, (mostly) observational data	Geospatial (2D), (mostly) EO data	Gridded (3D), Model data	EO data (2D)
Content	Air pollutant conc‘s, Weather variables	Topography, Population density, Landcover, etc.	Weather variables	Column integrals of air pollutant con‘s
Source	Environmental agencies, OpenAQ, Universities, WMO	EO institutes, NASA, ESA, OSM	ECMWF, DWD	NASA, ESA, Jaxa
Format*	Database and REST API	Database and REST API	Grib and netCDF files	HDF5 files
Size	Gbytes - TBytes	GBytes	Tbytes - PBytes	Tbytes - PBytes
Storage location	HDF Cloud VM	HDF Cloud VM	MeteoCloud	MeteoCloud

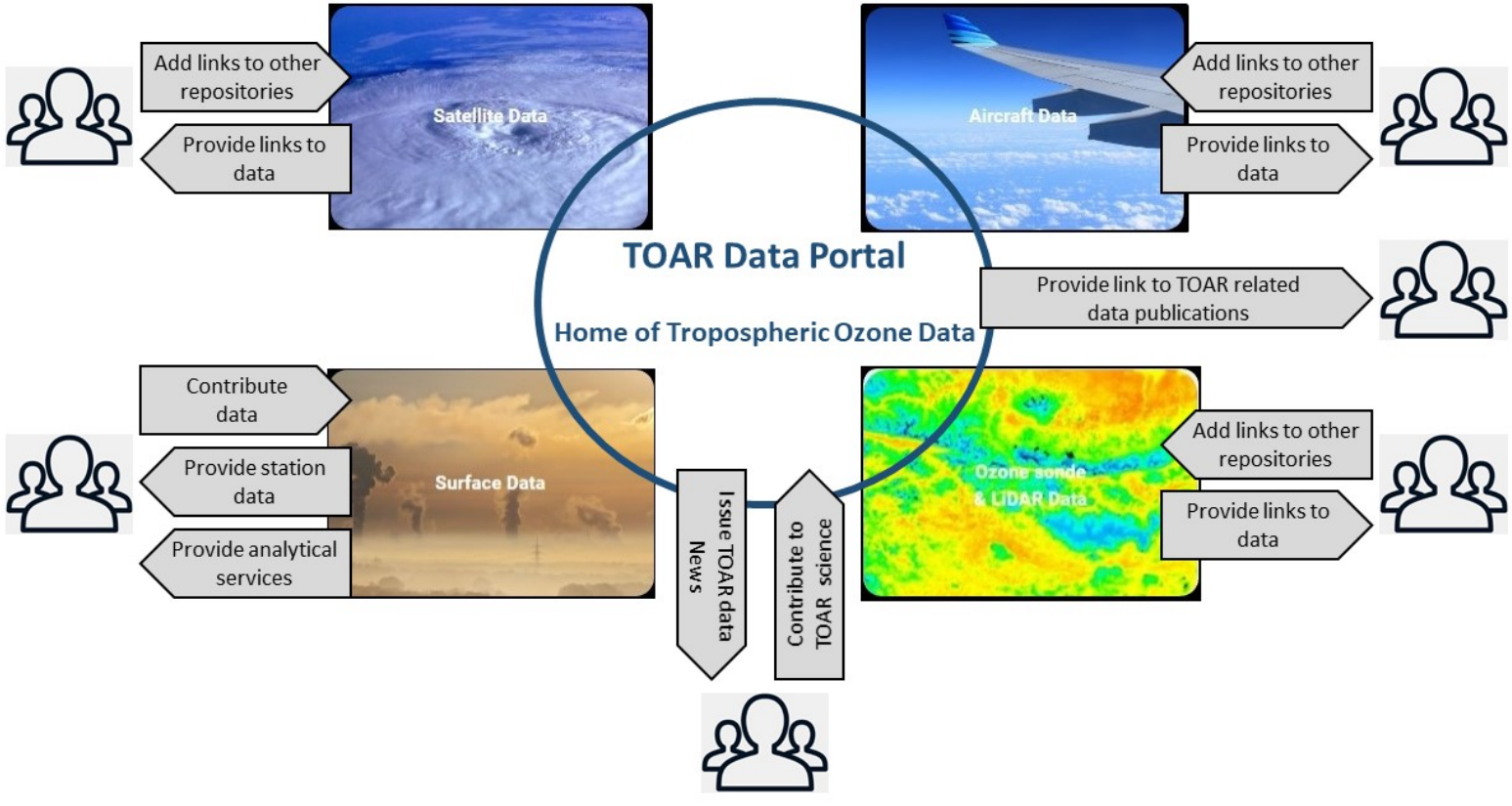
*after conversion

Owen R. Cooper and Martin G. Schultz

TOAR-II co-Chairs
on behalf of the TOAR-II Steering Committee

- > 200 scientists participating from > 30 countries
- First assessment published in 2017-2019: > 450 citations
- TOAR database with ~ 10 billion data points (13,000 stations)
- Second assessment phase: 2021 – 2024
- TOAR data infrastructure at JSC

The TOAR data portal



https://toar-data.org

Data services for TOAR data

Findable Accessible Interoperable Reusable

TOAR tropospheric ozone assessment report Phase II

TOAR database

12 bio data points
104,000 time series
13,000 stations
391 users

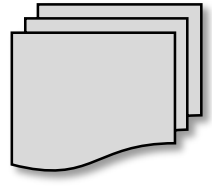
Graphical user interface → https://join.fz-juelich.de

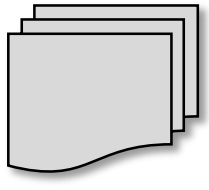
REST API → https://... (TBD)

Precompiled data products → https://b2share.fz-juelich.de

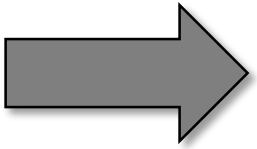
Team members:
Martin Schultz, Sabine Schröder, Niklas Selke, Björn Hagemeier, Rajveer Saini, Amirpasha Mozaffari, Mathilde Romberg, Ela Epp, Lukas Leufen, Sander Apweiler, Max Lensing, Jan Vogelsang, Clara Betancourt

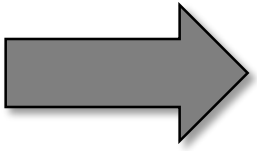
Processing of geospatial data for TOAR database and ML applications

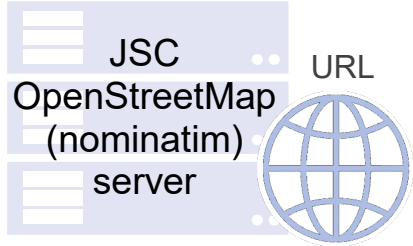


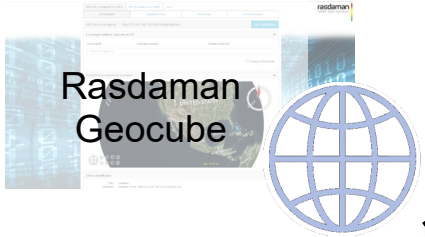


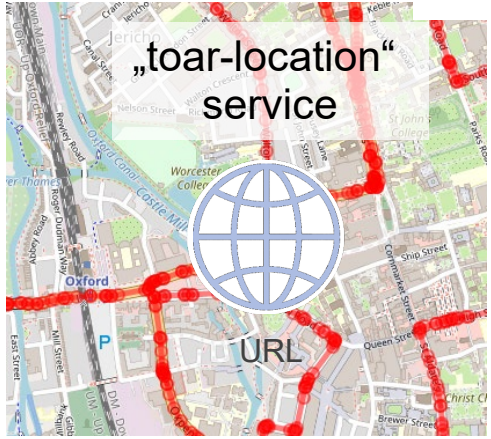
Original data:
Download from
various sources





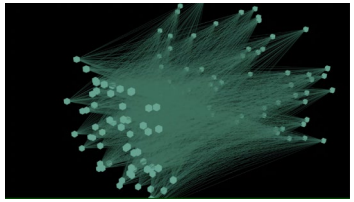


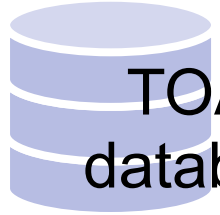




*

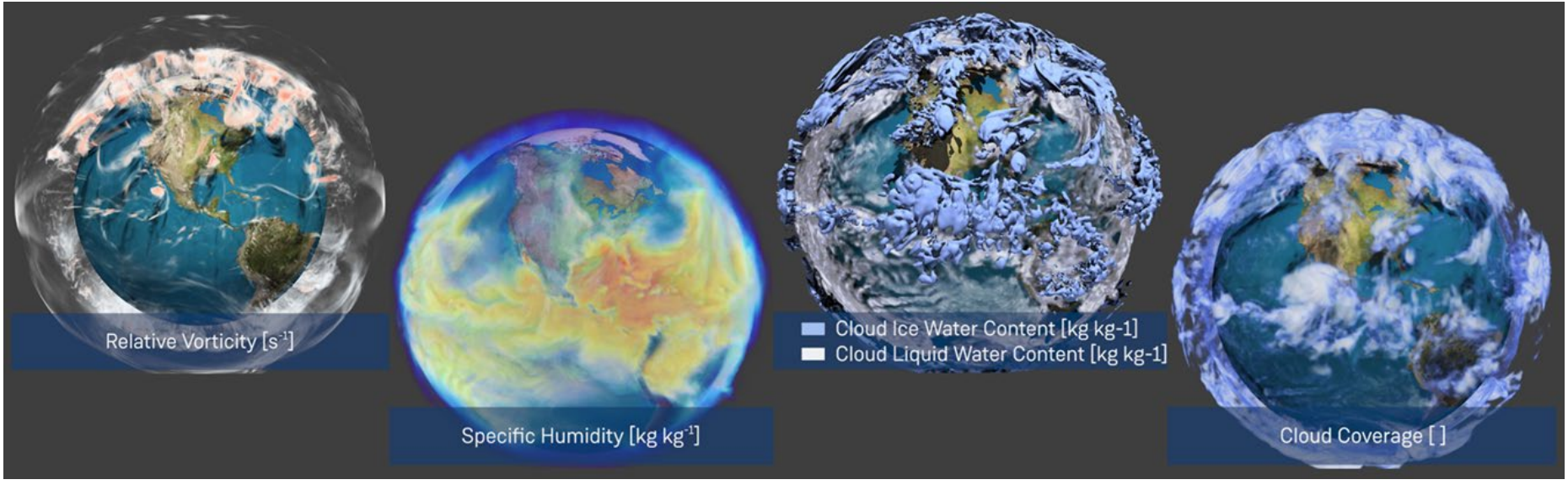
ML applications





* A whole story by itself…

The MeteoCloud @ JSC

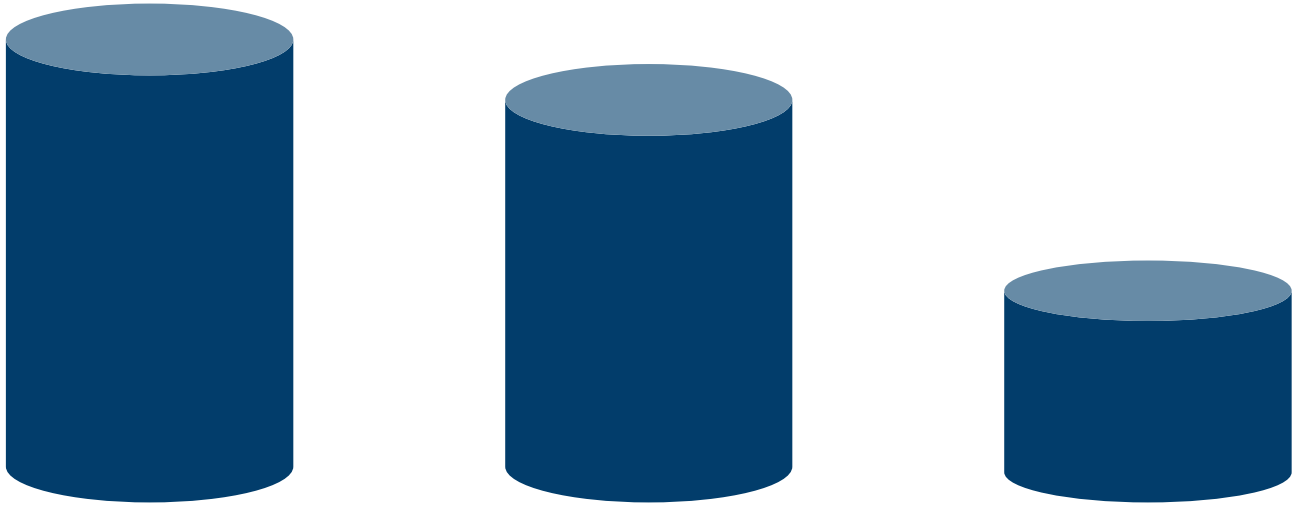


Community data project
Towards a collaborative analysis tool

Team members: Lars Hoffmann, Olaf Stein, Sabine Grießbach, Amirpasha Mozaffari, Gebhard Günther et al.

The MeteoCloud @ JSC

> 100 users

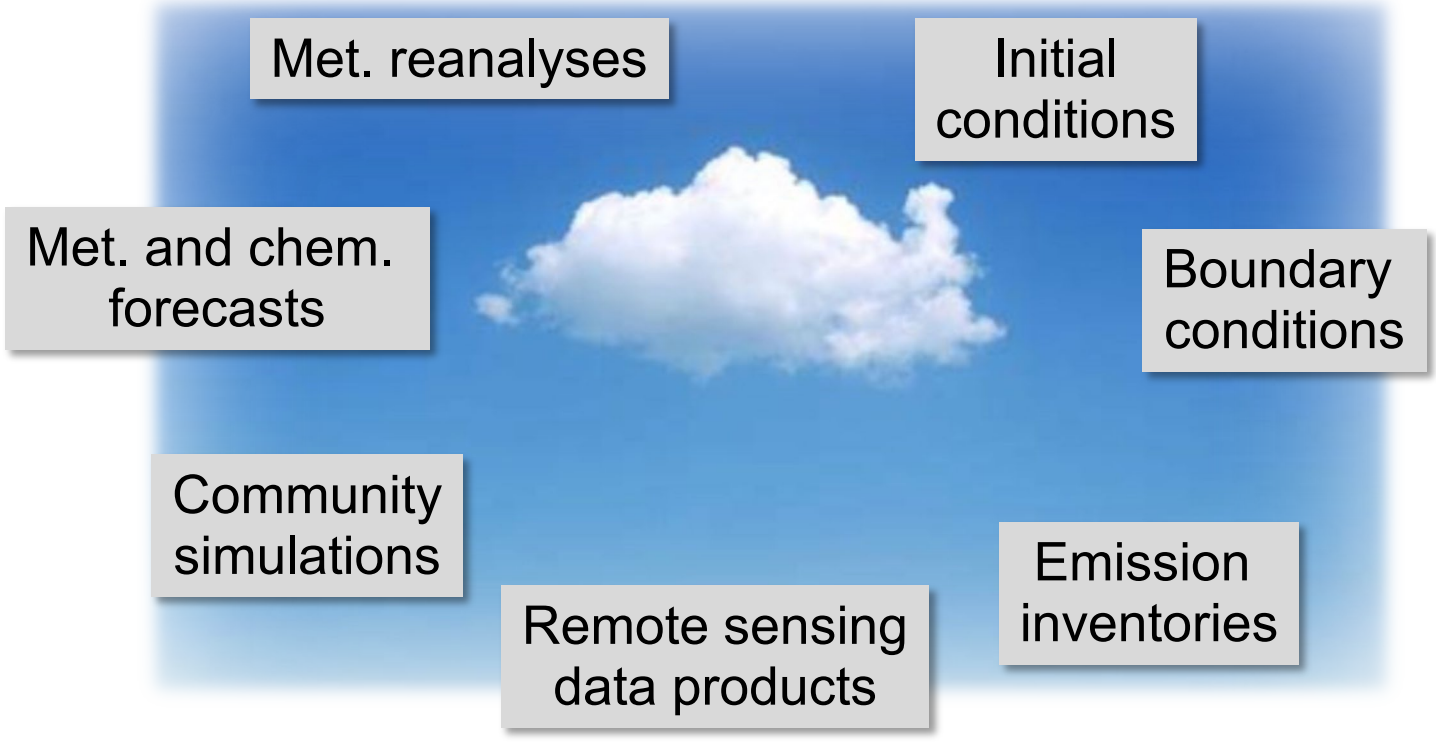


880 Tbytes
on „fastdata“

770 Tbytes
on „largedata“

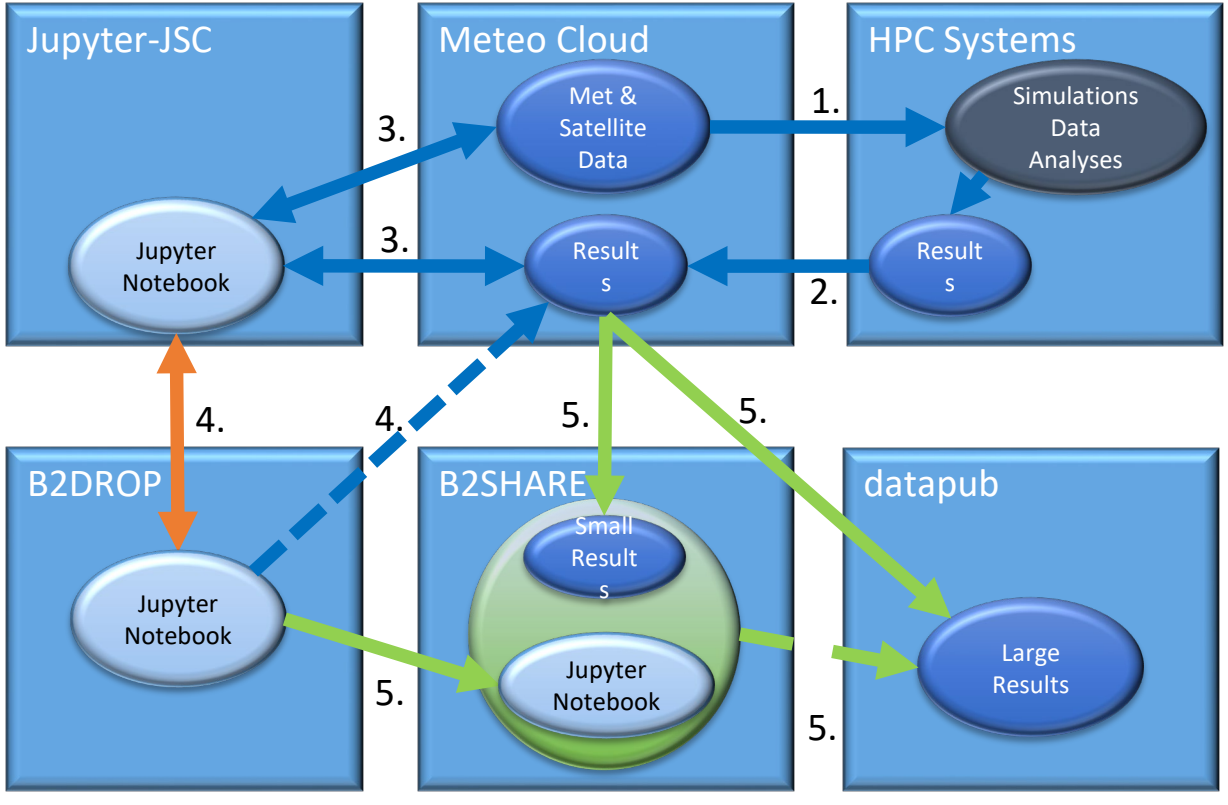
330 Tbytes
on „archive“

MeteoCloud Data Categories



Most data relate to Earth System Model simulations run at JSC

INTEGRATION OF DATA- AND ANALYTICS-SERVICES



1. Meteorological and satellite data is used in simulations and large analyses
2. Simulation results are stored on HPC systems
3. Results are visualized and prepared for publication with Jupyter Notebook (→ Jupyter-JSC)
4. Notebook is stored on B2DROP for sharing it with collaborators, linked to data on Meteo Cloud
5. Notebook and results are published on B2SHARE as one data set, large results are stored on datapub

METEOCLOUD: DeepRain use case

Amirpasha Mozaffari, Jessica Ahring,
Martin Schultz & DWD colleagues

DWD archive

Meteo cloud

655 TBytes

Former ensemble prediction & reanalysis, station data and radar data over Germany

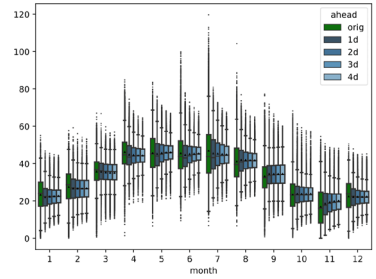
Rasdaman Earth Server
(data cube, OWS services)

rasdaman
raster data manager

$PROJECT (DeepACF)
$SCRATCH (DeepACF)

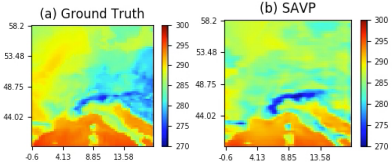
ML applications

JSC Jupyter



Ex.1: Monthly comparison of measured ozone concentration

JUWELS HDFML

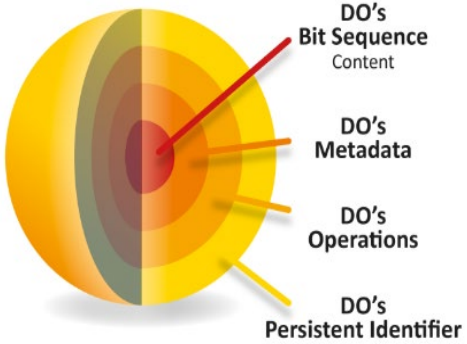


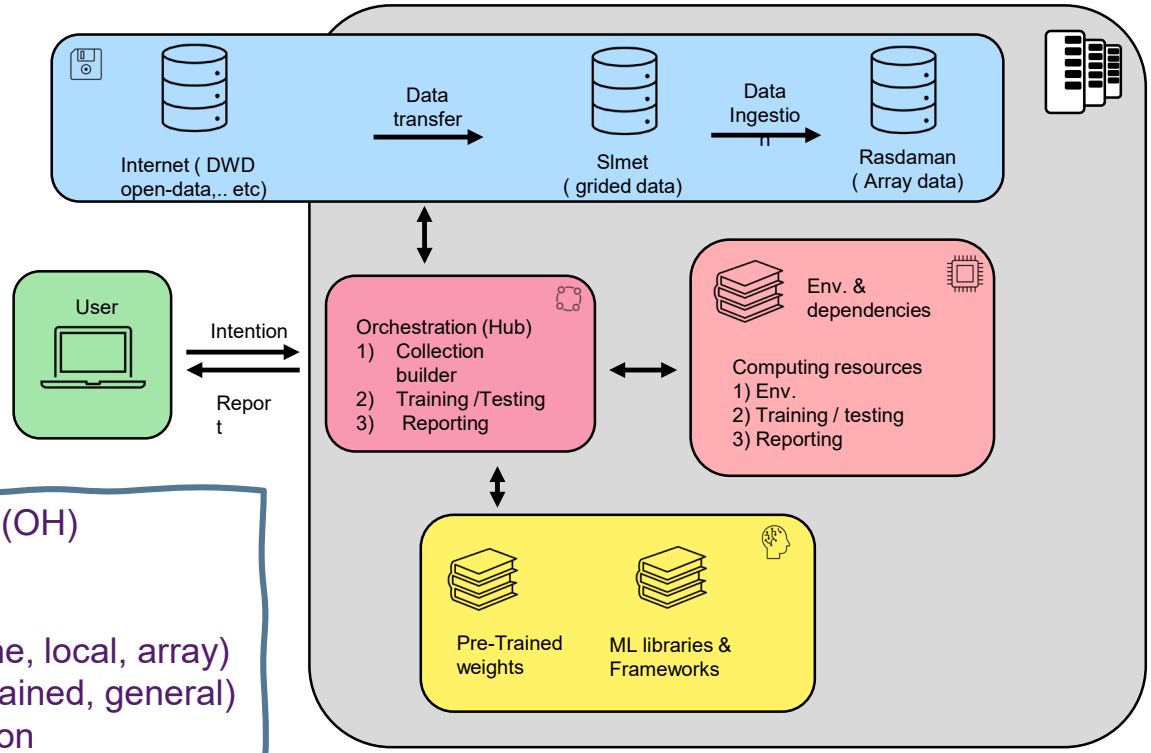
Ex.2: 2 meter above sea level temperature over Europe Region.

WORKFLOW

Conceptual DeepRain – ML- WF: Design

Digital Objects

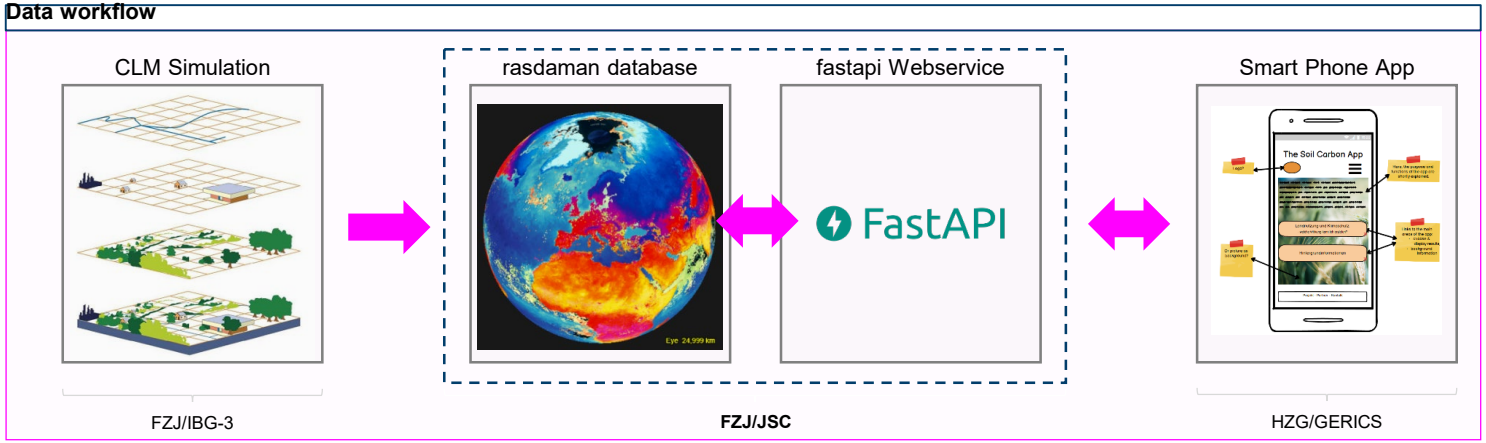




Orchestration Hub (OH)
4 interfaces

- User
- Data (online, local, array)
- ML (pre-trained, general)
- Computation

DATA WORKFLOW IN THE HI-CAM PROJECT
:: DEVELOPMENT OF PROTOTYPE SOIL-CARBON APP



Containerized multi-layer system with **rasdaman database** backend and **RESTful Webservice**

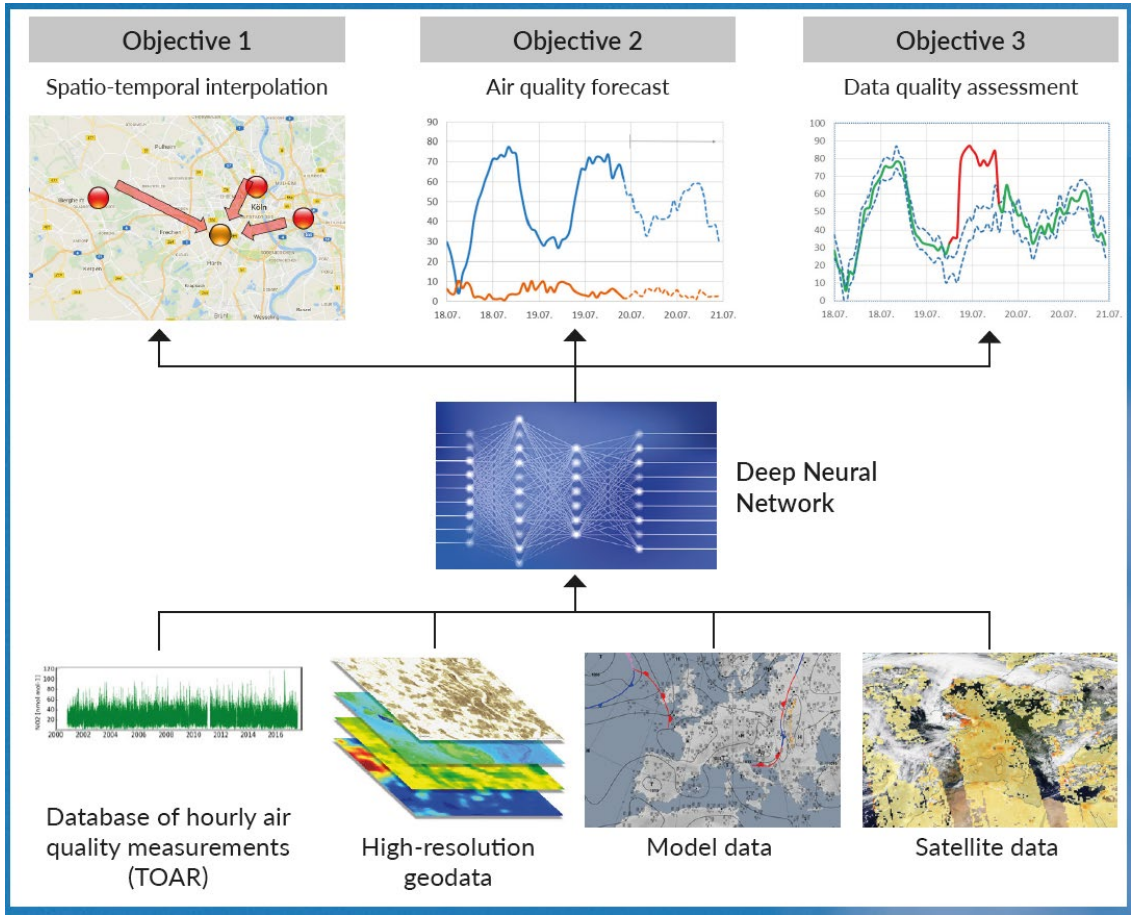
Team: Jianing Sun, Tanja Blome, Christian Dold, Michael Herbst, Diana Rechid, Martin Schultz

Data fusion: The core idea of IntelliAQ (and ESDE)

Relevant Science

Novel Methods

Big Data



Source: Schultz et al. (2020) link

IntelliAQ

European Research Council
Established by the European Commission

Advanced Grant
ERC-2017-ADG
#787576