FAIRness in Air Quality and Weather Forecast

FAIRness in the multi-services data infrastructure of the Tropospheric Ozone Assessment Report (TOAR) and Artificial Intelligence for Air Quality (IntelliAQ) project

Amirpasha Mozaffari, Sabine Schröder, Sander Apweiler, Rajveer Saini, Björn Hagemeier, Martin Schultz
Contact: a.mozaffari@fz-juelich.de Jülich Supercomputing Centre (JSC), Forschungszentrum Jülich GmbH, Germany

Motivation

TOAR and IntelliAQ are building a multi-level data service infrastructure for air quality and weather data. FAIR principles and modern data science guide the design at every level.

TOAR$_1$ is a joint effort to provide an up-to-date scientific assessment of tropospheric ozone's global distribution and trends from the surface to the tropopause.
TOAR Phase I : (2014 – 2019)
TOAR Phase II : (2020 – 2024)

IntelliAQ is a European project developing novel deep learning approaches for the analysis and synthesis of global air quality data based on deep neural networks by building a linkage of several different types of data.

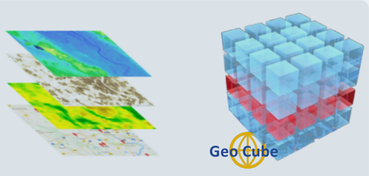
Data → **Data Analytics** → **Data Analytics** → **Products & Services**

Quality control of submitted data (manual / semi-automated)

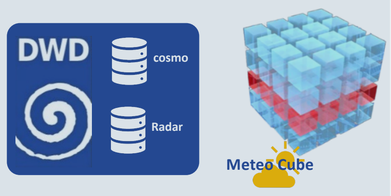
TOAR Database

TOAR database is a PostGIS database that includes multi-year time-series of ground-level ozone observations from over 30 providers at more than 10,000 sites around the world.

GeoCube includes satellite topographical data that are stored in an array database.

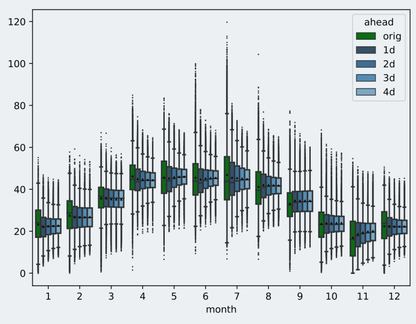


MeteoCube includes weather data and forecasts that are stored in an array database.



Ex.1 : Application of AI for near-surface ozone concentration prediction$_2$

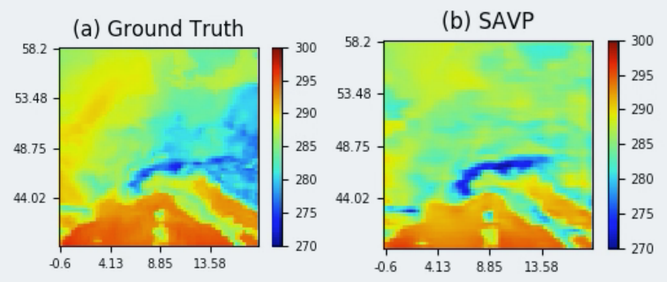
IntelliO3 is a data-driven forecasting model which aims to predict near-surface ozone concentrations (dma8eu) and generalizes well for approximately 330 measurement stations in Germany.



Monthly comparison of measured ozone concentration

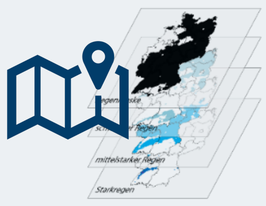
Ex.2 : Video frame prediction for weather and climate forecasting$_3$

Generative Adversarial Networks (GAN) - base Machine Learning (ML) architectures in the weather application.



2 meter above sea level temperature over Europe Region.

TOAR Data Portal



Geolocation and mapping services with API interface

A Representational State Transfer (API) allows querying all meta-data and data in TOAR.

Jülich Open Web Interface provides a graphical interface for TOAR database.

Data publishing services for datasets accompanying TOAR, in collaboration with EUDAT (B2SHARE)

Findable

- Standard data format :
  - ISO 19115
  - INSPIRE
- B2share metadata profile is developed by community extenstion of EUDAT
- Unique DOI for datasets

Accessible

- Free and open access services under the CC-BY 4.0 license for all IntelliAQ products.
- Metadata and provenance log will be available even when the data is not available anymore
- Https and REST API access

Interoperable

- Common self-describing data formats and standards (NetCDF 4.0, JSON)
- Following OGC coverage data structure
- Standardised vocabulary
- REST API access

Reusable

- Long-term usability of the air quality and climate data
- Available under CC-BY 4.0 license
- Manual and statistics-based automated quality control of the submitted data

1. Schultz, M.G., Schröder, S., Lyapina, O., … at al., (2017), Tropospheric Ozone Assessment Report: Database and Metrics Data of Global Surface Ozone Observations, Elem Sci Anth, 5, p.58
2. F. Kleinert, L. H. Leufen and M. G. Schultz (2020, in preparation)
   Current working title is: IntelliO3 v0.9: A Neural Network Approach to Predict Near Surface Ozone Concentrations in Germany
3. Bing Gong, Severin Hußmann, Scarlet Stadtler, … at al. , (2020, submitted). Temperature Prediction over Central Europe by Stochastic Adversarial Video Prediction, 2020 European Conference on Computer Vision

Outlook

We have started the process to have the TOAR data centre certified under the Core Trust Seal regulations. IntelliAQ and TOAR aim to produce datasets that can be reused for several decades. Besides its main role as a community data repository, the TOAR data centre acts as a platform to test novel, high-performance workflows for heterogeneous data sets, primarily in the context of machine learning applications.

IntelliAQ is funded by the EU's ERC programme, Grant Agreement 78576.