

The TOAR-II Data Infrastructure

Martin G. Schultz, Sabine Schröder, Jianing Sun, Clara Betancourt, Ela Epp, Mathilde Romberg, Najmeh Kaffashzadeh, Lukas Leufen, Jan Vogelsang, Jessica Ahring, Amirpasha Mozaffari | Jülich Supercomputing Centre, Forschungszentrum Jülich, Germany



TOAR-II Quickstart Event (virtual)

16 Sept 2020



TOAR-II in a nutshell – the relevance of data for TOAR-II

TOAR-II continues the successful work of TOAR-I and lasts until 2024

TOAR-II will provide updated and extended metrics on tropospheric ozone.

TOAR-II will further enhance the TOAR data portal and web services

TOAR-II will extend the statistical toolbox and trend analyses

TOAR-II will maximize exploitation of the TOAR Surface Ozone Database

TOAR-II will provide an updated state of the science estimate of ozone's global distribution and trends relevant to climate, human health and vegetation

TOAR-II reaches out to the international scientific community

The TOAR database: foundation of the first assessment

Table 1: Ozone monitoring networks as defined in the TOAR database^a. DOI: <https://doi.org/10.1525/elementa.244.t1>

Network name	Short name	Regional coverage	Data center or data provider	Number of stations
Airbase	Airbase	Europe (incl. overseas locations)	European Environment Agency https://www.eea.europa.eu/data-and-maps/data/airbase-the-european-air-quality-database-8	3505
University of New Hampshire Air Quality and Climate Program	Airmap ^b	NE US	University of New Hampshire http://www.eos.unh.edu/observatories/data.shtml	6
US Air Quality System	AQS	US	United States Environmental Protection Agency https://ofmext.epa.gov/AQDMRS/aqdmrz.html	2963
Australia Air Quality Network	AUSAQN ^b	Australia	Ian Galbally; Rob Gillett, and Suzie Molloy, CSIRO, Australia	56
The Canadian Air and Precipitation Monitoring Network	CAPMoN	Canada	Environment and Climate Change Canada, 2016 Canadian Air and Precipitation Monitoring Network (CAPMoN), Toronto, Ontario, Canada. Data file: CAPMoN_O3_ALT_to2015.csv, generated 2016-11-09.	19
US Acid Deposition Monitoring Network in East Asia	EANET	East Asia	Asia Center for Air Pollution Research http://www.eanet.asia/	16
Global Atmosphere Watch	GAW	global	World Data Center for Greenhouse Gases http://ds.data.jma.go.jp/wdcgg	196
Korea Air Quality Network Surveillance	KRAQN ^b	South Korea	Meehye Lee, Korea University, and National Institute of Environmental Research, Korea	312
National Institute for Environmental Studies, Japan Quality Network	NIES ^b	Japan	Hiroshi Tanimoto, National Institute for Environmental Studies, Japan	1260
Umweltbundesamt	UBA	Germany	S. Feigenspan (UBA)	613
other ^c			various individuals and data from smaller national or regional networks	130

Schultz et al., *Elementa* (2017)

Martin G. Schultz on behalf of the TOAR-II SC
TOAR-II Quickstart Event (virtual), 16 Sept 2020



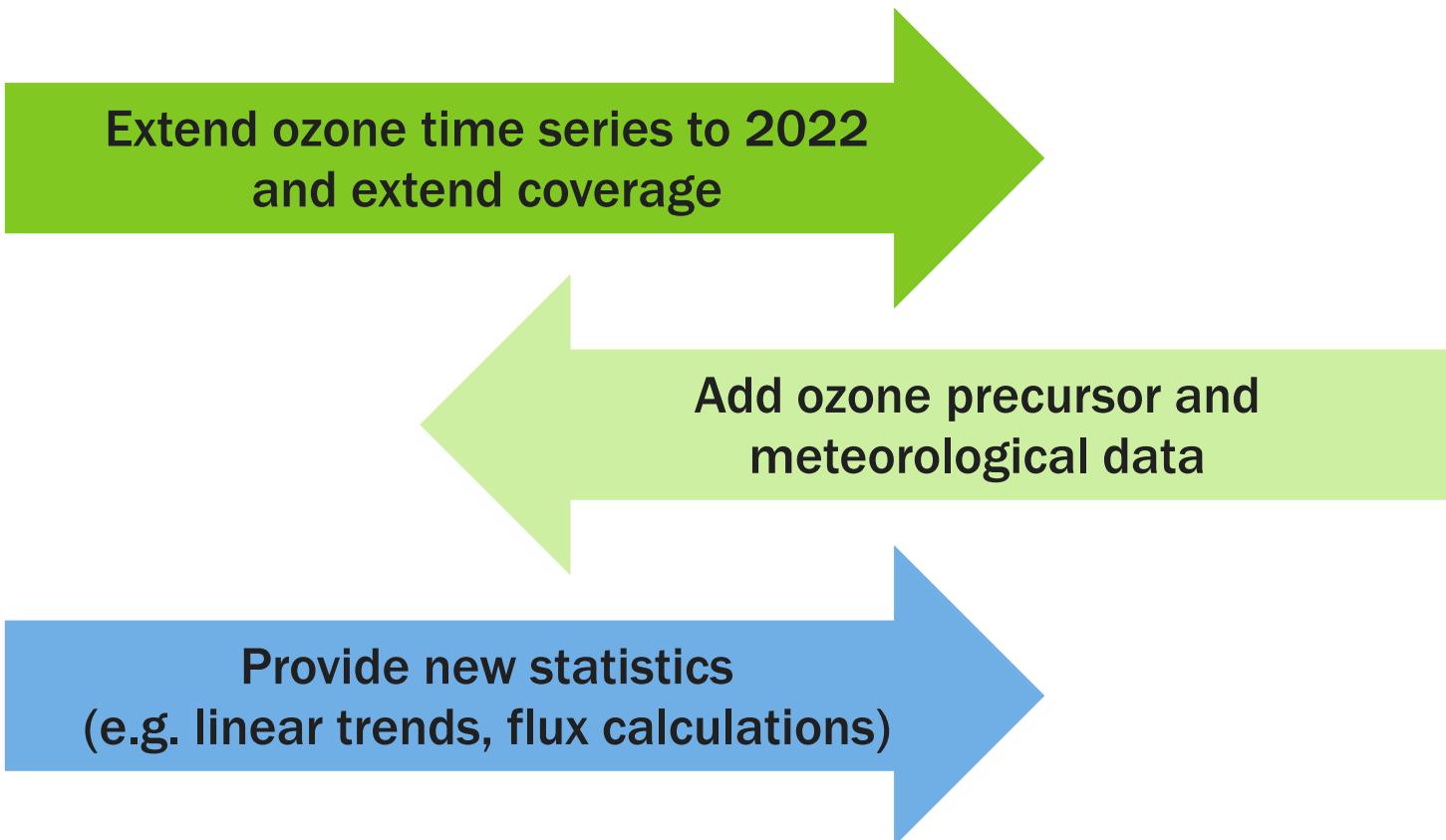
Ozone metrics	Daily	Monthly	Seasonal	Summer	Annual
Max. 1-h value	☒	☒	☒	☒	☒
Daily max, 8-h average (US EPA definition)	☒	☒	--	--	☒
Daily max, 8-h average (proposed new US EPA definition)	☒	--	--	☒	☒
Daily max, 8-h average (proposed new EU strict data capture)	☒	☒	--	☒	☒
SOMO35	☒	☒	☒	☒	☒
SOMO10	☒	☒	☒	☒	☒

> 30 ozone metrics
directly accessible
from database

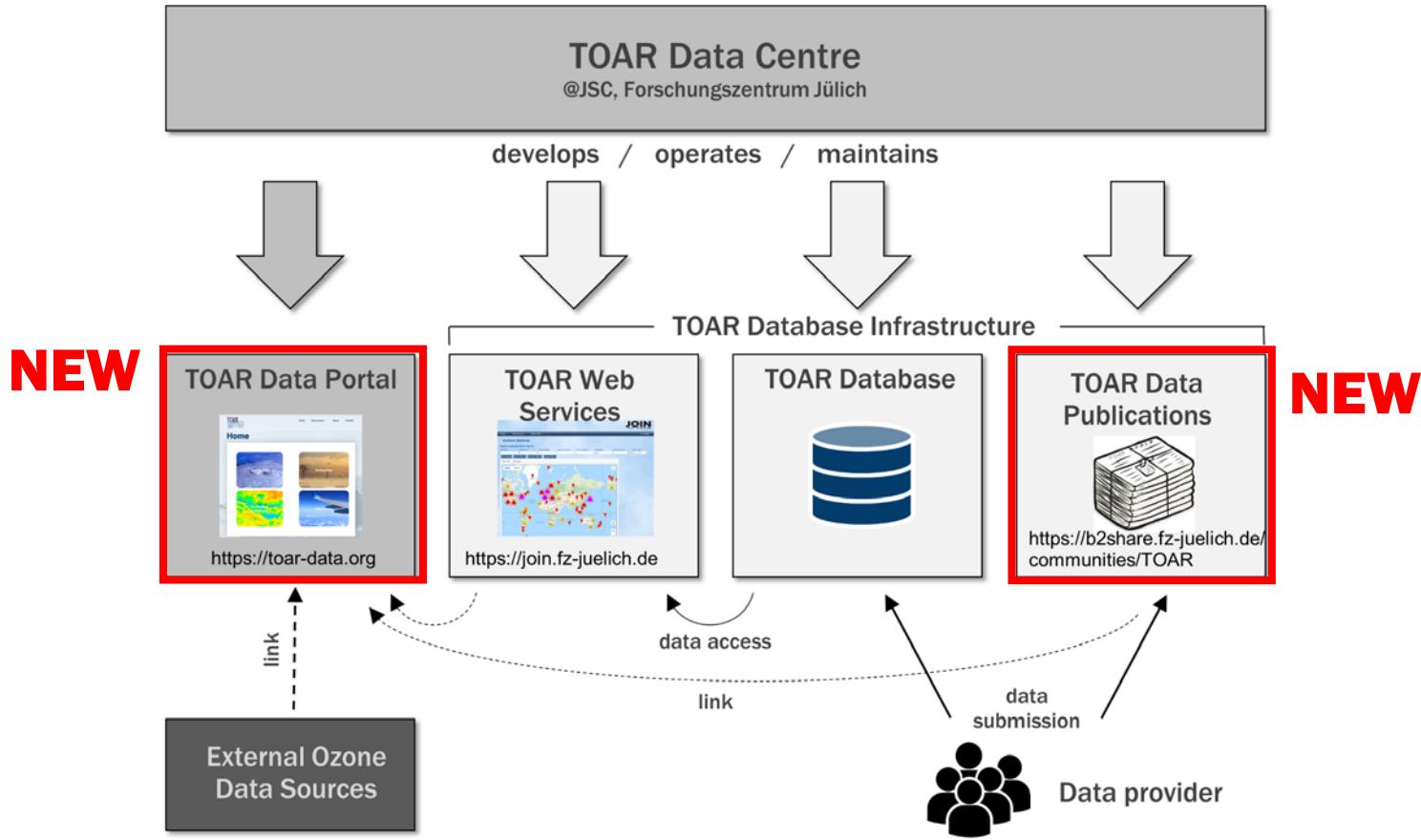


<https://igacproject.org/activities/TOAR/TOAR-II>

TOAR database: new content for TOAR-II



The TOAR data centre: more than a database



Design of a new TOAR data portal

<https://toar-data.org>

The screenshot shows the TOAR data portal homepage. At the top left is the TOAR logo: "TOAR" in large letters, "tropospheric ozone assessment report" in smaller letters below it. To the right are three navigation links: "HOME PAGE", "EXAMPLE PAGE", and "CUSTOM PAGE". Below the header are four main data categories: "Satellite Data" (image of a satellite in space), "Surface Data" (image of a city skyline), "Ozone sonde & LiDAR Data" (image of a weather map), and "Aircraft Data" (image of an airplane). To the right of these images is a "Latest News:" section with four news items:

- FAIRNESS in Air Quality and Weather Forecast (27 MARCH 2020 / 0 COMMENTS)
- News from Jülich (23 MARCH 2020 / 0 COMMENTS)
- New data from Colombia (23 MARCH 2020 / 0 COMMENTS)
- Hallo Welt! (13 MARCH 2020 / 0 COMMENTS)

A large red text box in the bottom right corner of the page area contains the text: "Make it easy to find ozone datasets".



Design of a new TOAR data portal

<https://toar-data.org>

Example: Ozone sonde & LIDAR data

The screenshot shows the homepage of THE WORLD OZONE AND ULTRAVIOLET RADIATION DATA CENTRE (WOUDC). It features the TOAR logo at the top left, followed by navigation links: Home, Data Access, About, and Contact. Below this is the WMO logo. A central section titled "THE WORLD OZONE AND ULTRAVIOLET RADIATION DATA CENTRE (WOUDC)" provides a brief description of the data centre's role in the Global Atmosphere Watch programme. Below this are three main data access sections: "SOUTHERN HEMISPHERE ADDITIONAL OZONESONDSES" (SHADOZ), "NETWORK FOR THE DETECTION OF ATMOSPHERIC COMPOSITION CHANGE" (NDACC), and "LIDAR". Each section includes a brief description, a "REST:API" link, a "Contact" link, and a "License" link.

Goal:
Provide easy access to
tropospheric ozone
datasets

- Direct link to data (as direct as possible)
- Descriptive text with link to home page
- Link to REST API (if available)
- Contact form or email (if available)
- Data use policy (if available)



TOAR data publications

<https://b2share.fz-juelich.de>

The screenshot shows the B2SHARE EUDAT portal interface. At the top, there are logos for B2SHARE and EUDAT. A search bar contains the placeholder "Search records for...". Below the search bar are navigation links: HELP, COMMUNITIES (which is highlighted in orange), UPLOAD, and CONTACT. To the right of the search bar are "SEARCH" and "Login" buttons. A decorative graphic of yellow hexagons is on the right side. The main content area displays a dataset entry for "TOAR timeseries of ground-level ozone at station GuCheng (GCH), China". The entry includes author information ("by Xu Xiaobin; Chinese Academy of Meteorological Sciences (CAMS), China Meteorological Administration (CMA);"), the date ("Jan 27, 2020"), and the last update date ("Last updated at Jun 18, 2020"). A detailed abstract follows, along with series information and technical details. The dataset is identified by a DOI: 10.34730/9f767ab1de5947a9ac7b0bbe501f6b41.

TOAR timeseries of ground-level ozone at station GuCheng (GCH), China

by Xu Xiaobin; Chinese Academy of Meteorological Sciences (CAMS), China Meteorological Administration (CMA);

Jan 27, 2020

Last updated at Jun 18, 2020

Abstract: Timeseries of 1-hourly measurements of ground-level ozone at station GuCheng (GCH), China.

These data were provided by Prof.Dr. Xiaobin Xu of Key Laboratory for Atmospheric Chemistry (KLAC), Chinese Academy of Meteorological Sciences (CAMS), China Meteorological Administration (CMA) in the context of the Tropospheric Ozone Assessment Report (TOAR).

For further information about TOAR see <http://www.igacproject.org/activities/TOAR>.

An up-to-date record of this data can be found in the TOAR database at https://toar-data.fz-juelich.de/api/getCoverage/?id=48237&sampling_method=hourly (not yet implemented)

SeriesInformation: This timeseries belongs to the collection <http://doi.org/10.34730/b26288bee23b4b1da6a97ce1fd64387e> of TOAR datasets from station GuCheng (GCH), China.

TechnicalInfo: The data are formatted as a csv file with variable header information. Each header line begins with # and contains key: value metadata elements.

The data file was generated by the TOAR database query https://toar-data.fz-juelich.de/api/getCoverage/?id=GCH&sampling_method=hourly&version_tag=1.0 (not yet implemented).

A detailed description of the file format and the metadata elements can be found at: https://toar-data-portal.fz-juelich.de/docs/surfacedata/TOAR_file_format_v2.pdf (not yet implemented).

Software tools for working with these data are available at <https://jugit.fz-juelich.de/m.schultz/toar-public-utilities>.

[View in B2SHARE](#) [View in EUDAT](#) [View in TOAR](#) [View in DataCite](#) [View in Zenodo](#)

Keywords: TOAR, air quality, tropospheric ozone;

DOI: [10.34730/9f767ab1de5947a9ac7b0bbe501f6b41](https://doi.org/10.34730/9f767ab1de5947a9ac7b0bbe501f6b41)

Name	Size
CAMS_GCH_O3_2006-2018_v1-0.csv	2.42MB
2 downloads	
Open Access	True ✓
Embargo Date	Tuesday, March 24, 2020 12:00 AM
License	Creative Commons Attribution (CC-BY)

Martin G. Schultz on behalf of the TOAR-II SC
TOAR-II Quickstart Event (virtual), 16 Sept 2020



<https://igacproject.org/activities/TOAR/TOAR-II>

We can now issue dataset
for your data
→
You get a citeable reference

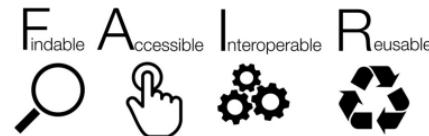


TOAR database: technical enhancements



Improved data model and new REST API

Implement FAIR and
Open Data principles

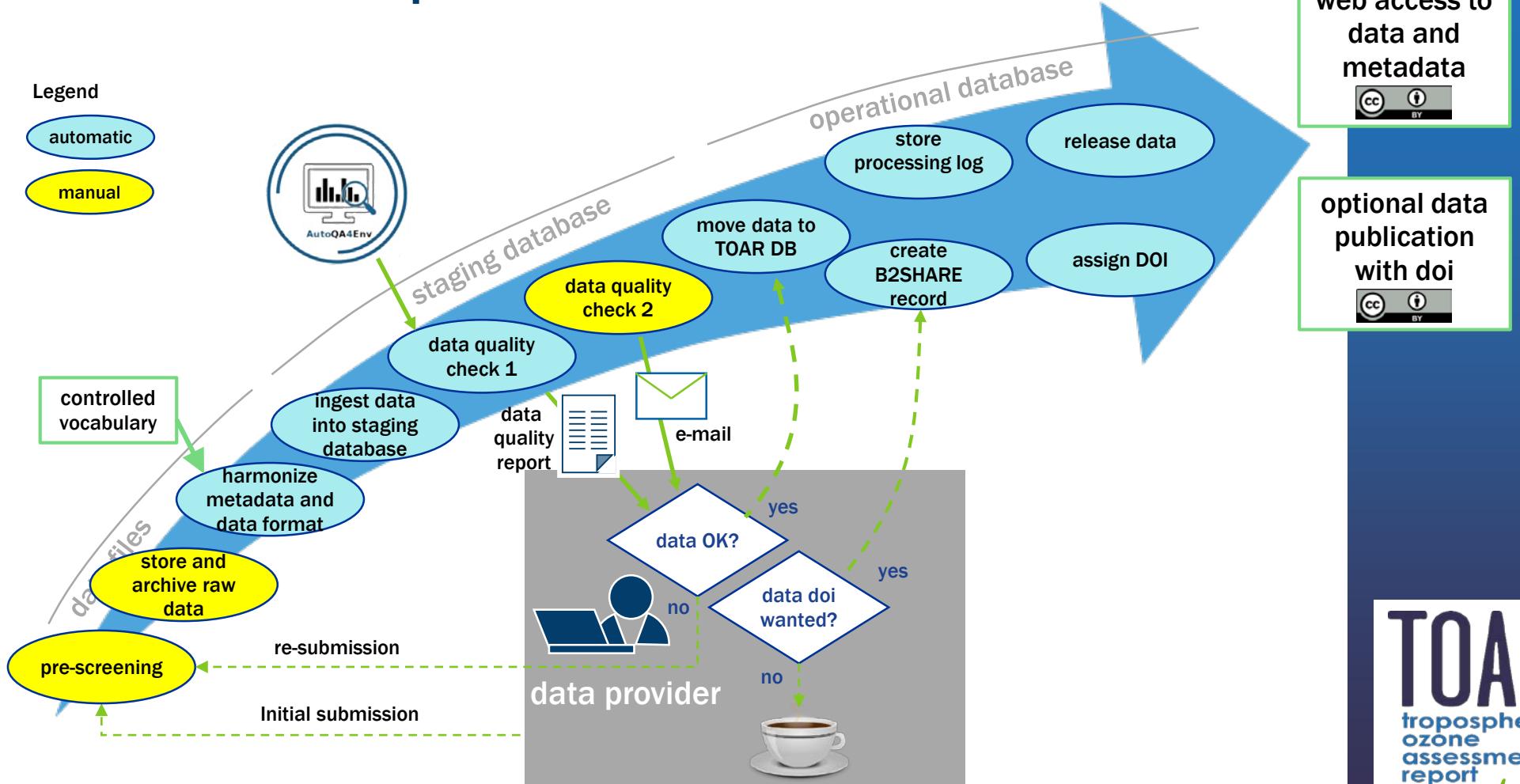


Performance and functionality
enhancements

Obtain Core Trust Seal certification

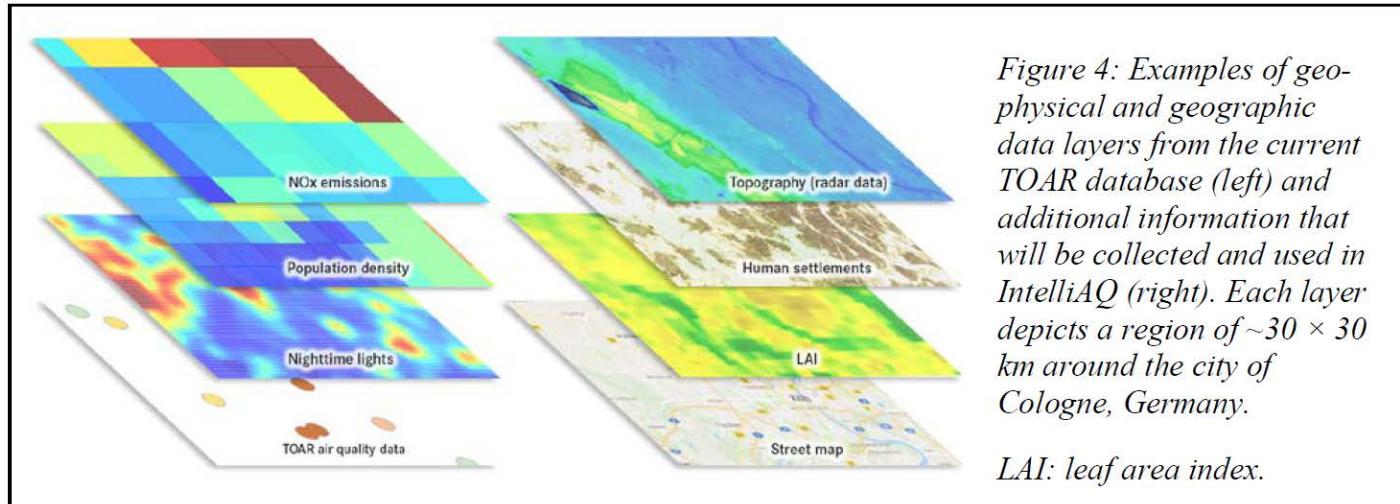


TOAR database: improved workflow and metadata



TOAR-II will make use of more and higher-resolution geospatial datasets

Such global datasets have turned out to be very useful for a globally consistent station characterisation



Source: IntelliAQ proposal

TOAR-II will integrate thousands of additional measurements from OpenAQ

openaq

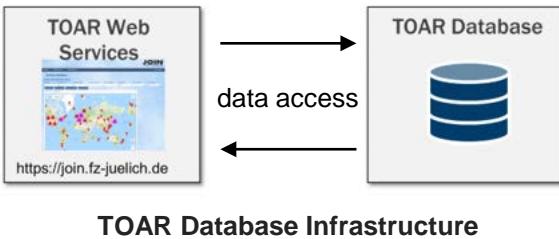
Programmatic, transparent access to historical + NRT station AQ data in a universal format:



- Primarily gov't sources
- ~594 million total measurements
~600,000 added daily
- 12000+ stations in 93 countries
- Data accessible via API, csv files, user-customized files / basic visualizations via openaq.org

openaq

TOAR-II will integrate thousands of additional measurements from OpenAQ



■ Challenges:

- Full / historical OpenAQ dataset stored as JSON dumps on Amazon S3 cloud data storage (API access only for last two years)
- Data mapping from source service (OpenAQ) to target service (TOAR database) is required

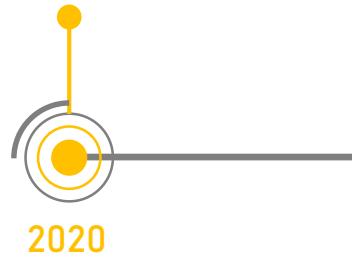
■ Solution:

- Use an intermediate PostgreSQL database for data staging
- Parse daily real-time JSON dumps into PostgreSQL
- Extract stations and parameters from parsed JSON records
- Create Timeseries in TOAR database based on station and parameter
- Match real-time data to Timeseries and import data
- Workflow automation with **Apache Airflow**



TOAR-II Data infrastructure roadmap

- Design data infrastructure
- Bring data backend online



2020

2021



- Optimize performance
- Design web services for TOAR-II
- Receive first data submissions(*)

(*) limited data submissions are always welcome

- Implement web applications
- Receive data submissions
- Update large network data
- Generate sample products

2022



2023



- Generate final products
- Publish data products

2024



Acknowledgements

The TOAR data centre is funded through



European Research Council

Established by the European Commission

Advanced Grant
ERC-2017-ADG
#787576

and



Quick links

TOAR-I

Web interface to database: <https://join.fz-juelich.de>

REST API for software access to database: <https://join.fz-juelich.de/services/rest/surfacedata/>

Pre-compiled data products: <https://doi.org/10.1594/PANGAEA.876108>

Git repository with (very limited) software: <https://jigit.fz-juelich.de/m.schultz/toar-public-utilities>

TOAR-II

TOAR data portal: <https://toar-data.org/>

Web interface to database: <https://join2.fz-juelich.de> (*TBD*)

REST API for software access to database: *TBD*

Pre-compiled data products: <https://b2share.fz-juelich.de/communities/TOAR>

TOAR data publications: <https://b2share.fz-juelich.de/communities/TOAR>

Git repository: <https://gitlab.version.fz-juelich.de/toar> (*currently only internal use*)