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

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 reaches out to the international scientific community

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

TOAR tropospheric ozone assessment report Phase II

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/aqdmrs.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
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Acid Deposition Monitoring Network in East Asia	EANET	East Asia	Asia Center for Air Pollution Research http://www.eanet.asia/	16
[illegible]	[illegible]	[illegible]	[illegible]	196
Global Atmosphere Watch	GAW	global	World Data Center for Greenhouse Gases http://ds.data.jma.go.jp/wdcgg	123
[illegible]	[illegible]	[illegible]	Levana Kordova, Israeli Ministry of Environmental Protection	12
Korea Air Quality Network	KRAQN[b]	South Korea	Meehye Lee, Korea University, and National Institute for Environmental Research	312
[illegible] Surveillance	[illegible]	[illegible]	[illegible]	[illegible]
National Institute for Environmental Studies	NIES[b]	Japan	Hiroshi Tanimoto, National Institute for Environmental Studies	1260
[illegible] Quality Network	[illegible]	[illegible]	[illegible] tersrand, South Africa	22
Umweltbundesamt	UBA	Germany	S. Feigenspan (UBA)	613
[illegible]	[illegible]	[illegible]	[illegible] germany	[illegible]
[illegible]	other[b]	[illegible]	various individuals and data from smaller national or regional networks	130

Schultz et al., Elementa (2017)

https://join.fz-juelich.de

~350 users

Unique collection of global long-term surface ozone observations:

Now > 13,000 sites from 15 networks and many individual contributions

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 [illegible]		--	--		
Daily max. 8-h average (EU)					
Daily max. 8-h average (proposed new US EPA [illegible] capture)					
Daily max. 8-h average (EU), strict data capture		--	--		
3-months running mean of DMAX1H		--			
SOMO35					
SOMO10					

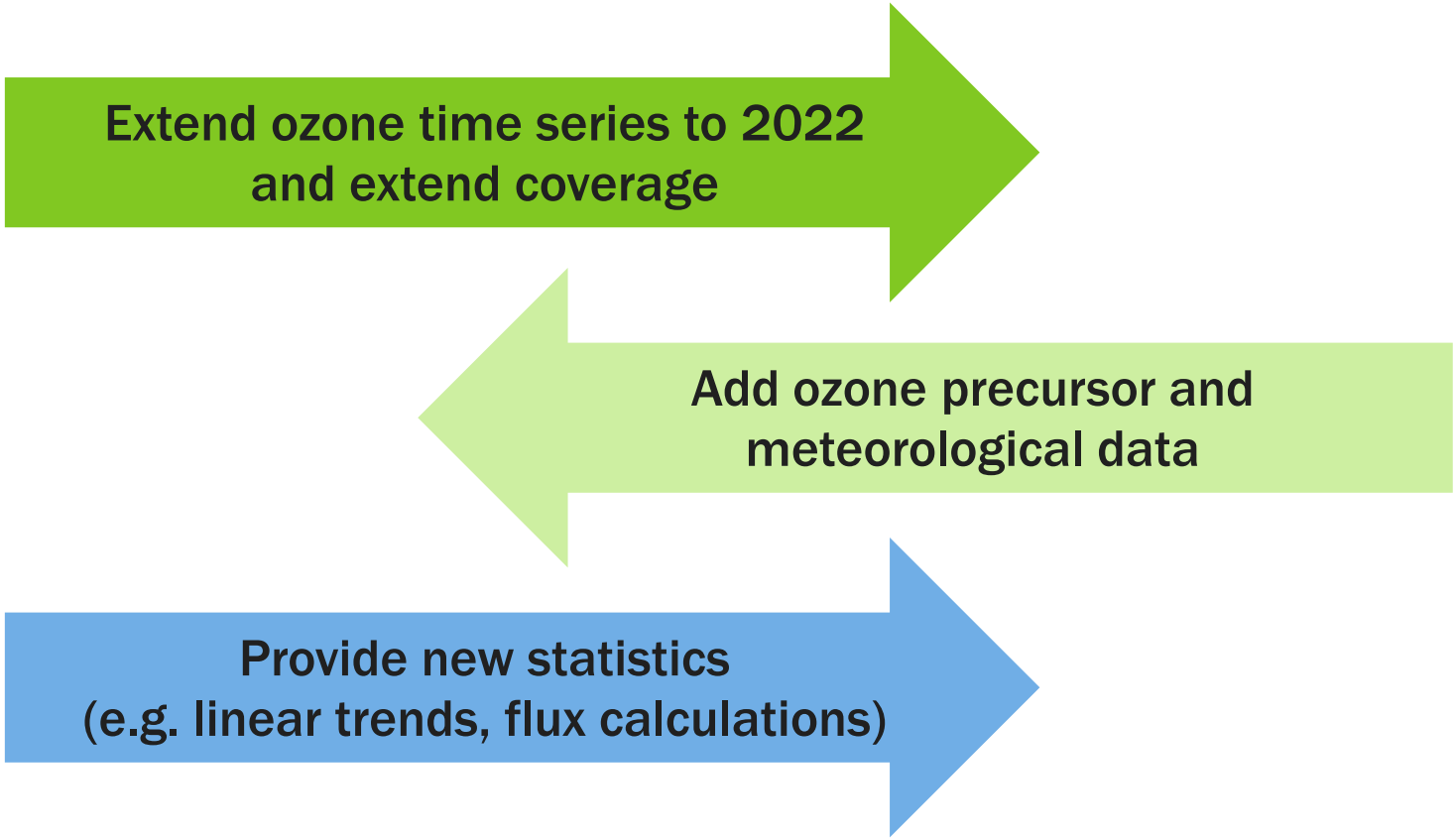
> 30 ozone metrics directly accessible from database

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

TOAR tropospheric ozone assessment report Phase II

TOAR database: new content for TOAR-II

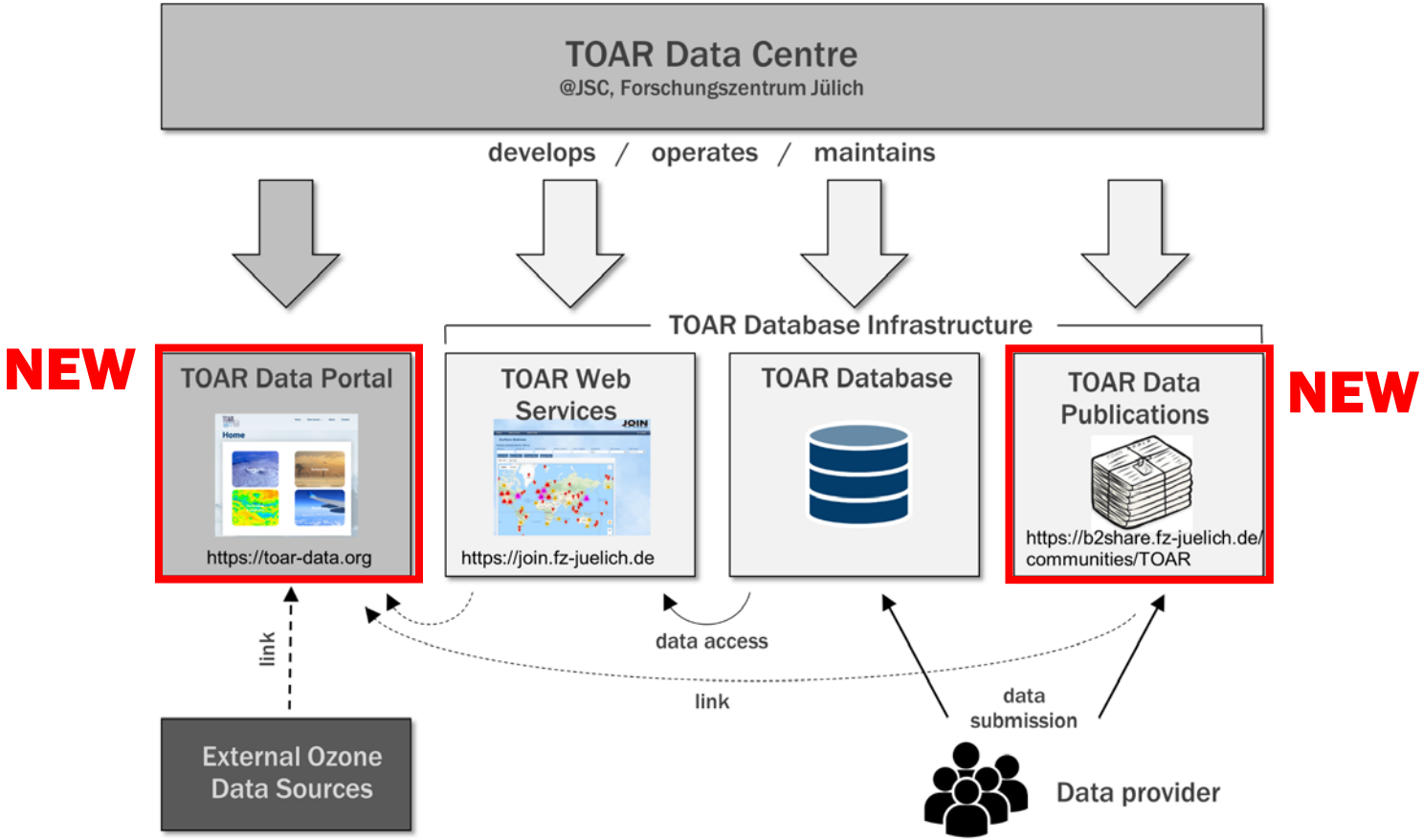


Martin G. Schultz on behalf of the TOAR-II SC
TOAR-II Quickstart Event (virtual), 16 Sept 2020

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

TOAR tropospheric ozone assessment report Phase II

The TOAR data centre: more than a database

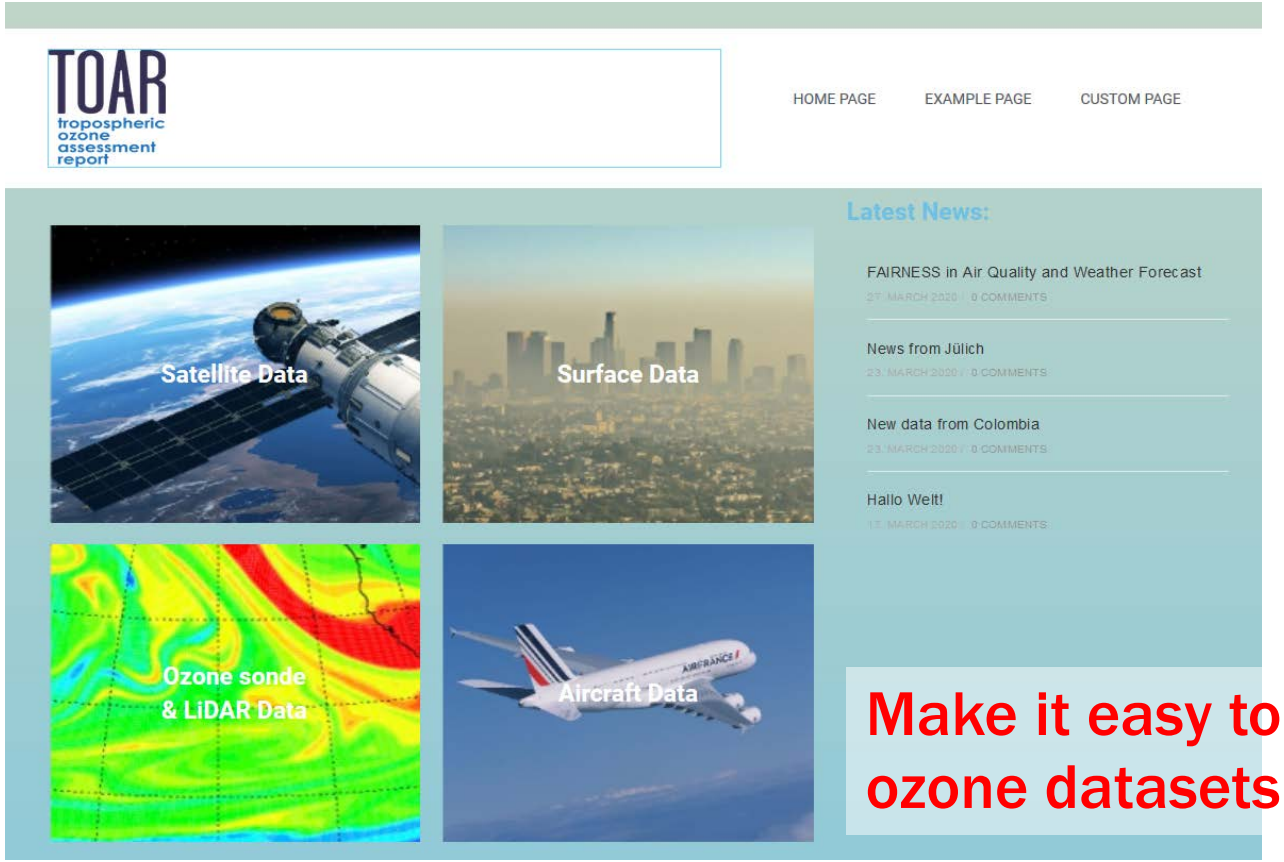


https://igacproject.org/activities/TOAR/TOAR-II

TOAR tropospheric ozone assessment report Phase II

Design of a new TOAR data portal

https://toar-data.org



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

TOAR tropospheric ozone assessment report Phase II

Design of a new TOAR data portal

https://toar-data.org

Example: Ozone sonde & LIDAR data

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)

REST:API Contact License

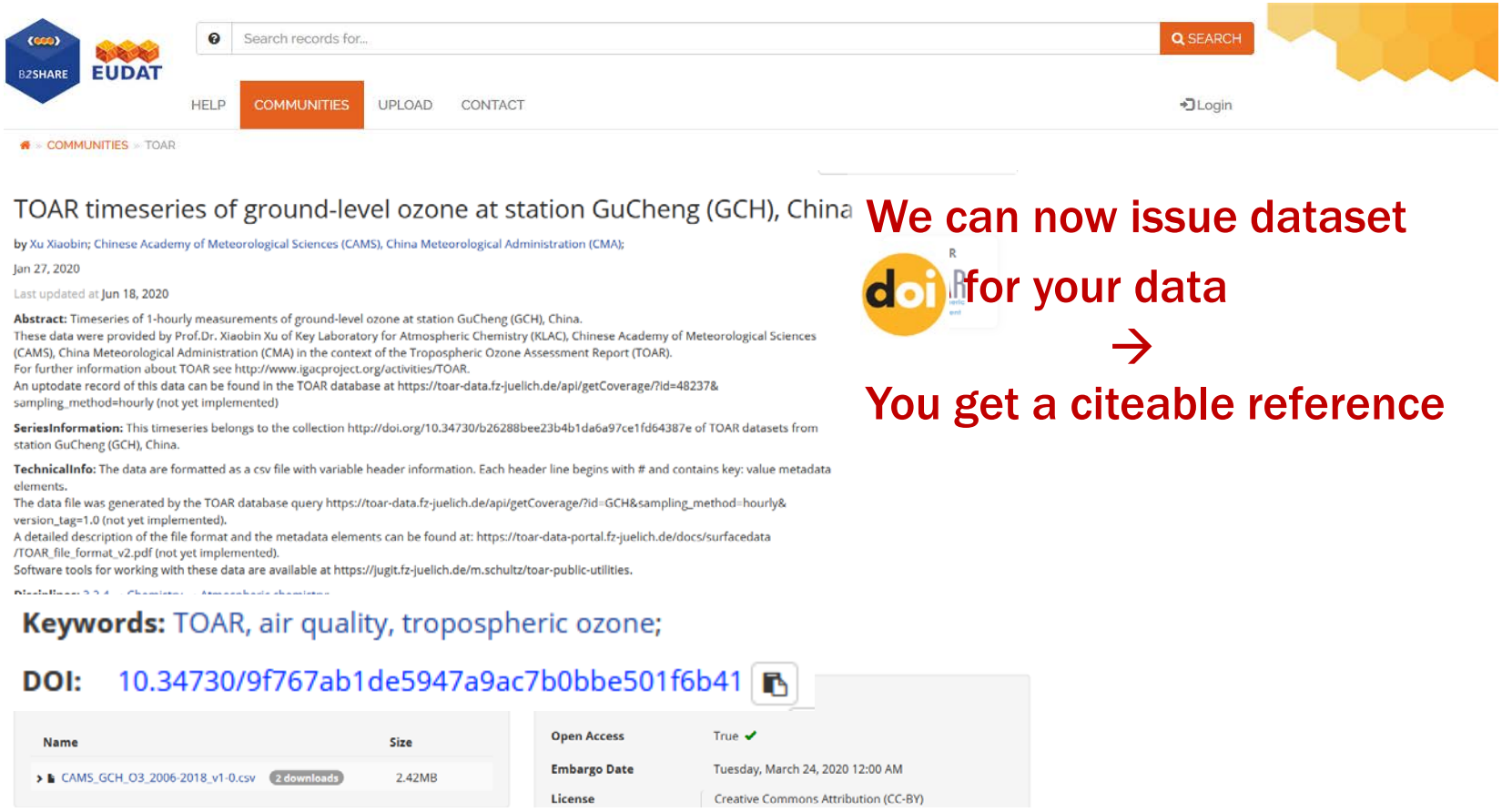
Martin G. Schultz on behalf of the TOAR-II SC
TOAR-II Quickstart Event (virtual), 16 Sept 2020

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

TOAR tropospheric ozone assessment report Phase II

TOAR data publications

https://b2share.fz-juelich.de



We can now issue dataset

for your data

→

You get a citeable reference

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

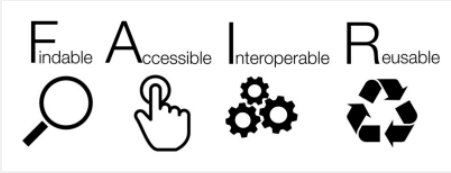
https://igacproject.org/activities/TOAR/TOAR-II

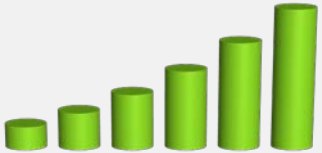
TOAR tropospheric ozone assessment report Phase II

TOAR database: technical enhancements

Improved data model and new REST API

Implement FAIR and
Open Data principles





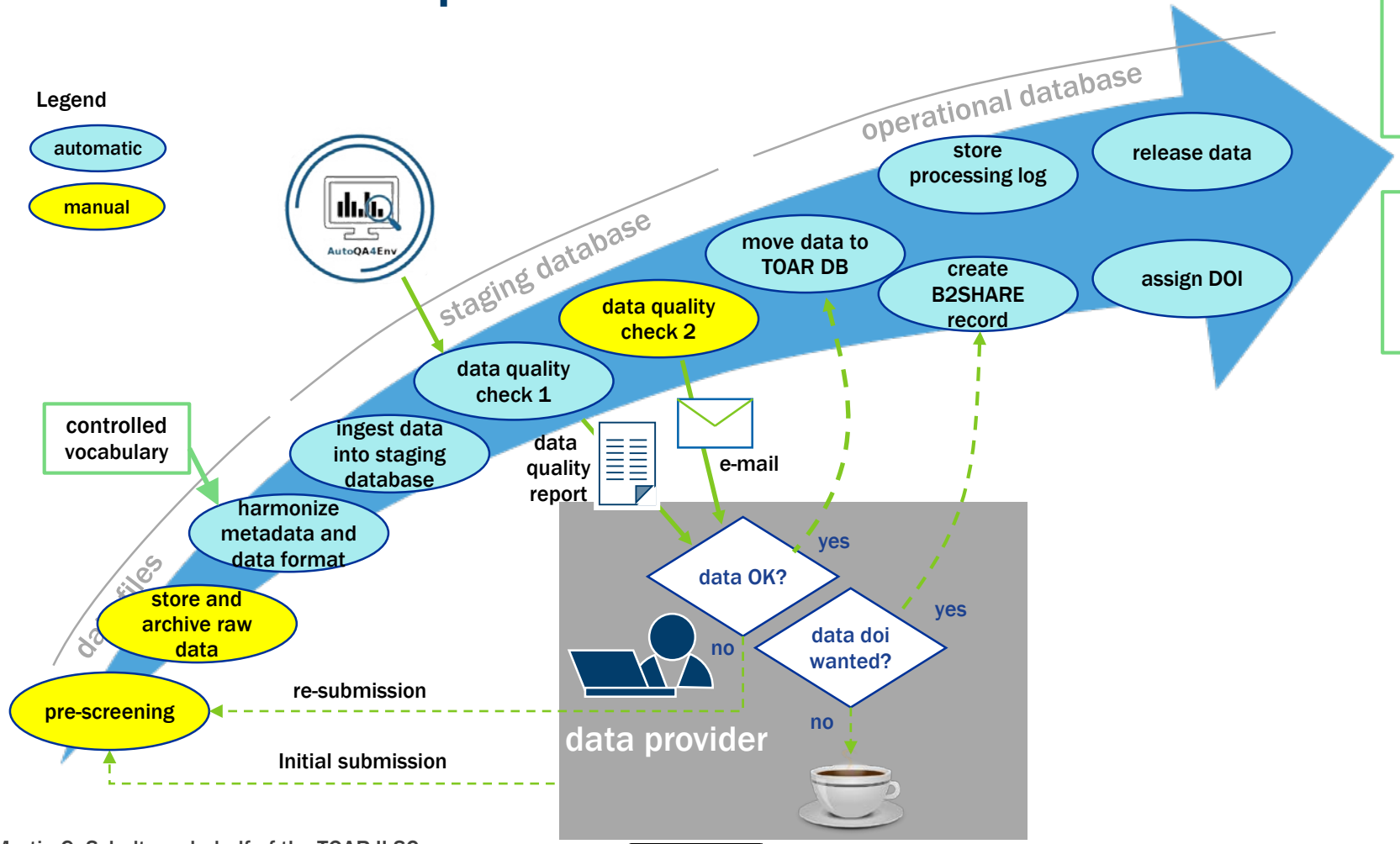
Performance and functionality
enhancements

Obtain Core Trust Seal certification

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

TOAR tropospheric ozone assessment report Phase II

TOAR database: improved workflow and metadata



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

TOAR tropospheric ozone assessment report Phase II

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

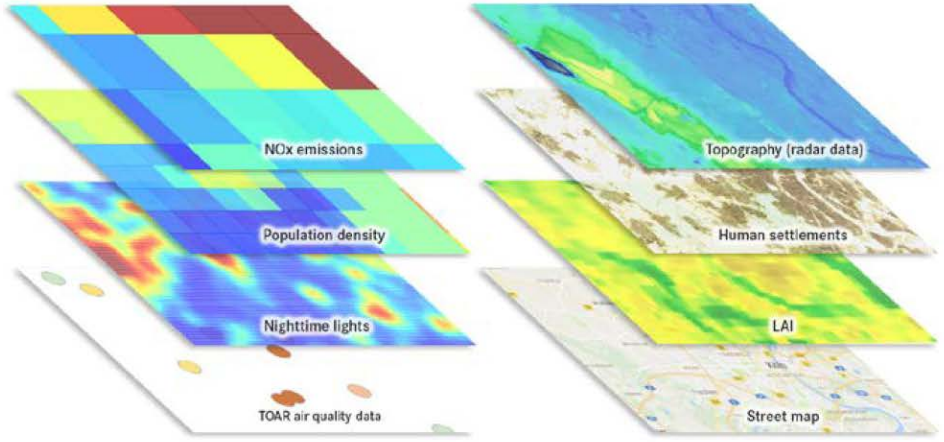


Figure 4: Examples of geo-physical and geographic data layers from the current TOAR database (left) and additional information that will be collected and used in IntelliAQ (right). Each layer depicts a region of ~30 × 30 km around the city of Cologne, Germany.

LAI: leaf area index.

Source: IntelliAQ proposal

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

TOAR tropospheric ozone assessment report Phase II

TOAR-II will integrate thousands of additional measurements from OpenAQ

openaq

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

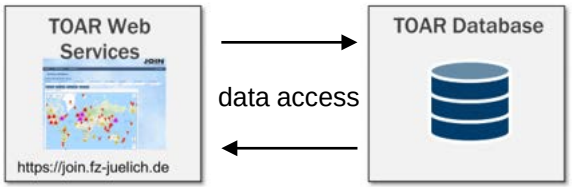
- Primarily gov't sources
- **~594 million** total measurements
  ~600,000 added daily
- **12000+** stations in **93** countries
- Data accessible via API, csv files, user-customized files / basic visualizations via openaq.org

openaq

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

TOAR tropospheric ozone assessment report Phase II

TOAR-II will integrate thousands of additional measurements from OpenAQ

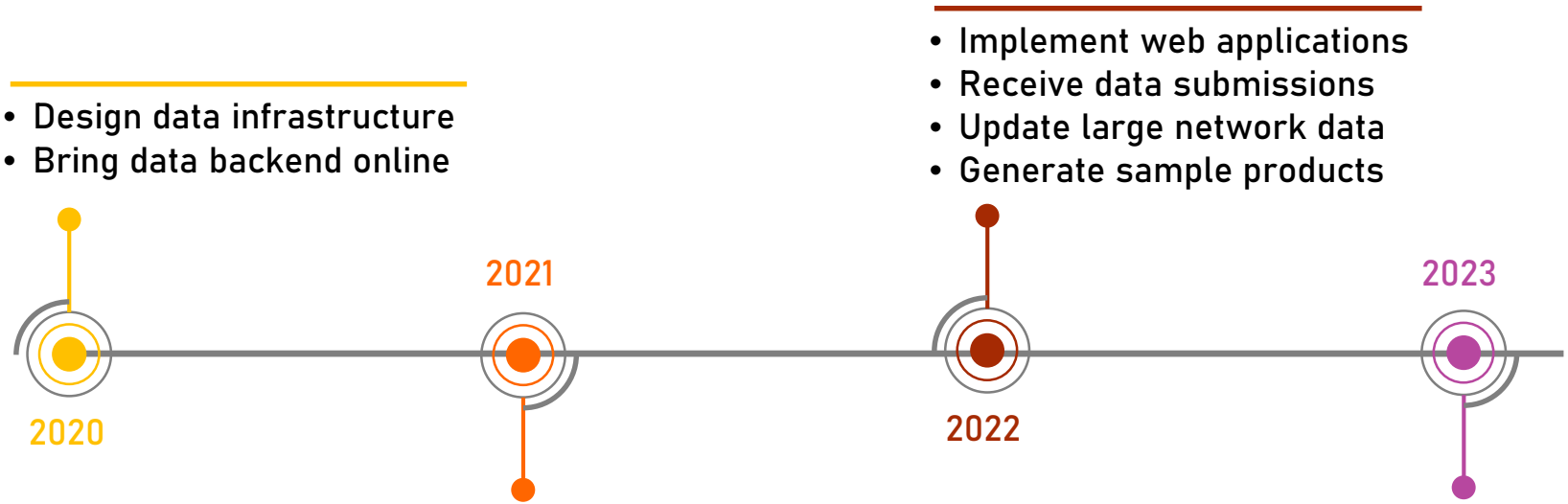


TOAR Database Infrastructure

- 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**

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

TOAR-II Data infrastructure roadmap



(*)limited data submissions are always welcome

2024

TOAR tropospheric ozone assessment report Phase II

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

Acknowledgements

The TOAR data centre is funded through

Advanced Grant
ERC-2017-ADG
#787576

and

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

TOAR tropospheric ozone assessment report Phase II

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://jugit.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)***

TOAR tropospheric ozone assessment report Phase II