RESEARCH DATA AND SOFTWARE MANAGEMENT IN TIMES OF FAIR AND OPEN DATA

Enhancing FAIRness of global air quality data: The Tropospheric Ozone Assessment Report database

25 SEP 2019 I SCHRÖDER, S.; APWEILER, S.; SAINI, R.; HAGEMEIER, B.; SCHULTZ, M. G.

JÜLICH SUPERCOMPUTING CENTRE

What is TOAR?

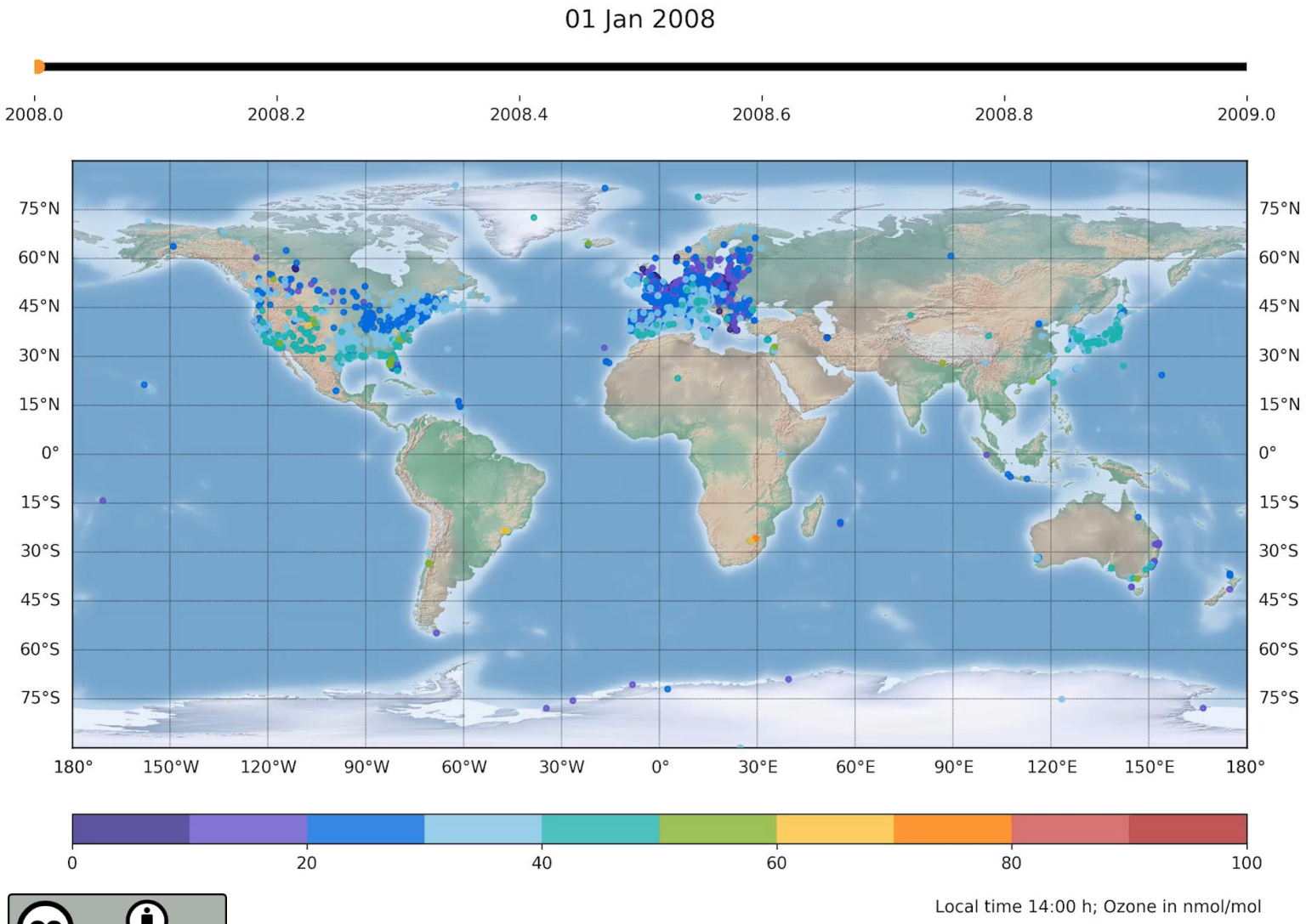
The „Tropospheric Ozone Assessment Report“ has been created by ~220 scientists from 36 countries to:

1. **Produce the first tropospheric ozone assessment report** based on the peer-reviewed literature and new analyses.
2. **Generate easily accessible, documented data** on ozone exposure and dose metrics at hundreds of measurement sites around the world (urban and non-urban), **freely accessible** for research on the global-scale impact of ozone on climate, human health and crop/ecosystem productivity.

Web link: http://www.igacproject.org/activities/TOAR

Movie about TOAR (in German): https://youtu.be/-k5mvtbq4M4

Motivation



JOIN (Juelich Open Web Interface)

as of 06 August 2019:
12,795 Ozone data series at 10,002 stations

Precursors and meteorology:
9278 data series NO_2
6253 data series NO
396 data series PM10
2744 data series CO
7560 data series temperature
…

TOAR database

sources of original data

- datacenter (networks: local/global)
- individual
- NRT

TOAR database
services

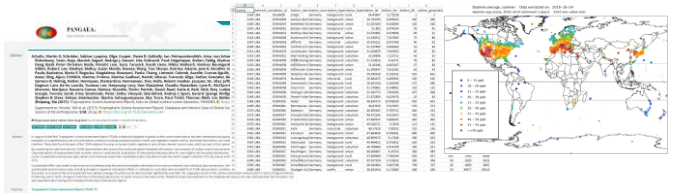
https:/join.fz-juelich.de
(graphical web interface)

https:/join.fz-juelich.de/services/rest/surfacedata/
(REST service)

https://join.fz-juelich.de/services/rest/surfacedata/parameters/
→

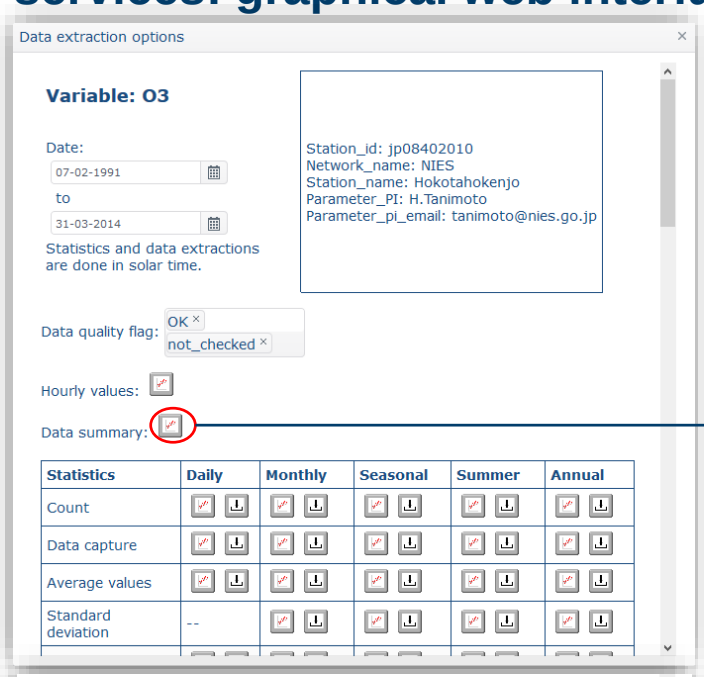
```
["albedo", "aswdifu", "aswdir", "benzene",
"ch4", "cloudcover", "co", "ethane",
"humidity", "no", "no2", …, „wspeed"]
```

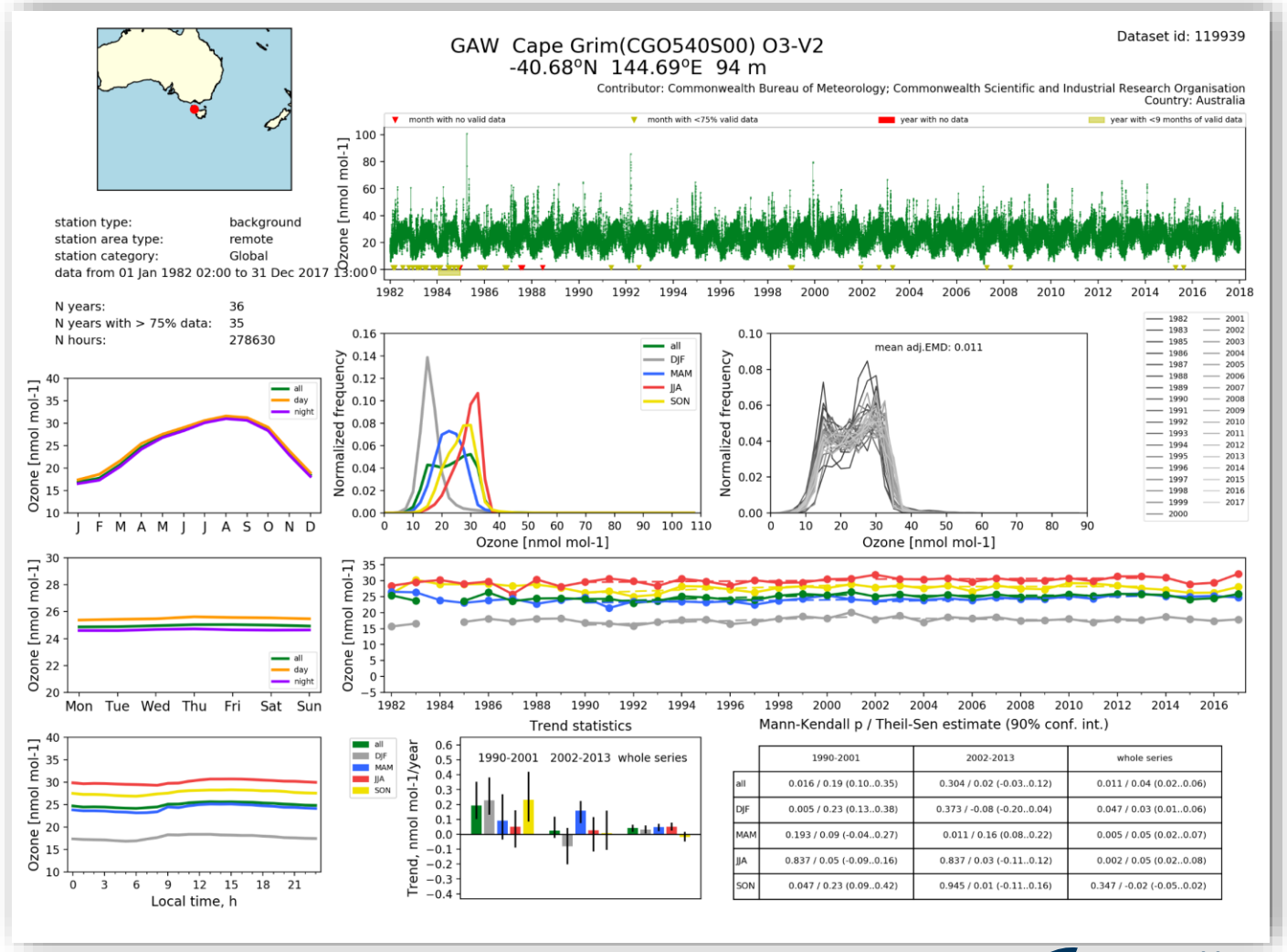

https://doi.pangaea.de/10.1594/PANGAEA.876108
(pre-processed data products)



TOAR database

services: graphical web interface: https://join.fz-juelich.de





TOAR database

services: REST Service: https:/join.fz-juelich.de/services/rest/surfacedata/

the default response format is json (currently, only json and html are supported)

https://join.fz-juelich.de/services/rest/surfacedata/

Response: Description and documentation of available REST services (this document)

1.2 Services

The following information services are available and described individually below. Each service is invoked by appending its name and possible query arguments to the base URL.

parameters: query the parameter (i.e. variable) names and properties of the database
networks: query the observation networks and their description from the database
stations: query station ids, station names, and station location from the database
series: query the data series id and specific metadata of a series from the database
search: execute a flexible database search query on stations and data series
stats: retrieve TOAR data products for a selected data series

Example:

https://join.fz-juelich.de/services/rest/surfacedata/search/?
station_name=Münster&station_state=Nordrhein-Westfalen&
columns=network_name,station_id,station_name

TOAR database

services: REST Service: https:/join.fz-juelich.de/servic

the default response format is json (currently, only json and html are suppo

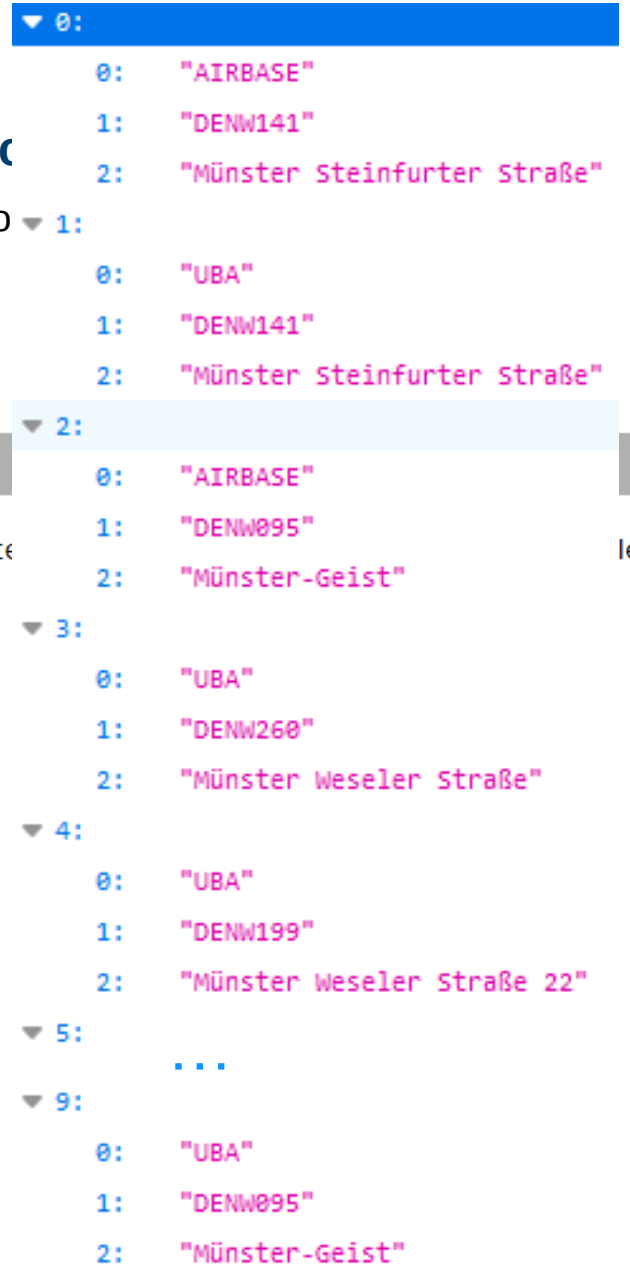
https://join.fz-juelich.de/services/rest/surfacedata/

Response: Description and documentation of available REST services (this document)

1.2 Services

The following information services are available and described individually below. Each servic le query arguments to the base URL.

parameters: query the parameter (i.e. variable) names and properties of the database
networks: query the observation networks and their description from the database
stations: query station ids, station names, and station location from the database
series: query the data series id and specific metadata of a series from the database
search: execute a flexible database search query on stations and data series
stats: retrieve TOAR data products for a selected data series



Example:

https://join.fz-juelich.de/services/rest/surfacedata/search/?station_name=Münster&station_state=Nordrhein-Westfalen&columns=network_name,station_id,station_name

TOAR database (2nd phase)

TOAR database roadmap

European Research Council
Established by the European Commission

Advanced Grant
ERC-2017-ADG
#787576
Oct 2018 – Sep 2023

IntelliAQ

- Enhance database model to capture more metadata
- Interface with openaq to obtain near realtime data globally
- Implement automatic quality control procedures
- Introduce data quality scores
- Document all data modifications

in preparation

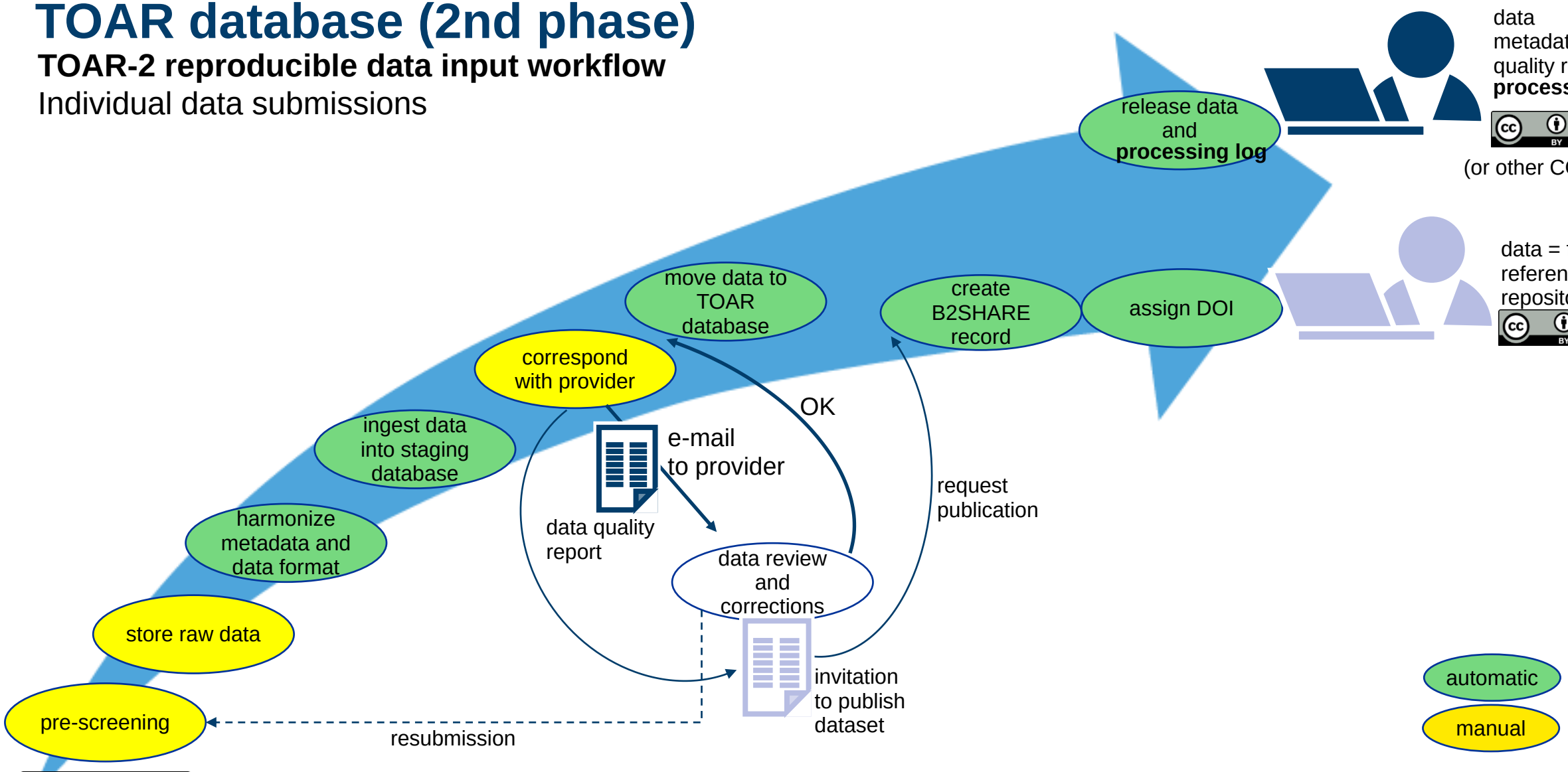
doi

CORE TRUST SEAL

TOAR database (2nd phase)

TOAR-2 reproducible data input workflow

Individual data submissions



TOAR database (2nd phase)
metadata

Station_id: …
Station_name: …
…
Comments: Some example text: any number of arbitrary free text with additional information

predefined keywords

Inlet_tube_material: ...
Inlet_tube_outer_diameter: ...
Flow_rate: …

arbitrary keywords

actual:

Station_id: …
Station_name: …
…
Comments: Some example text: any number of arbitrary free text with additional information

➔ All other metadata not stored in DB!

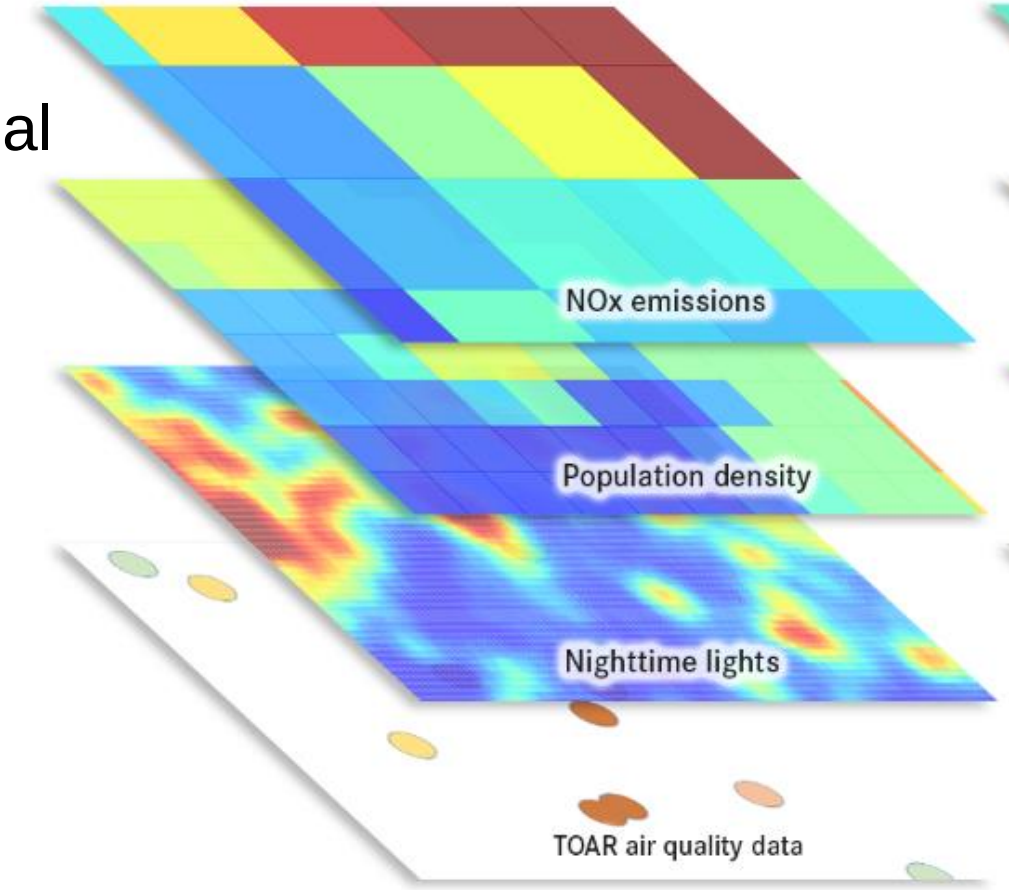
planned:

...
additional_metadata: {‘**Inlet_tube_material**‘:‘…‘, ‘**Inlet_tube_outer_diameter**‘:‘…‘ ,**Flow_rate**‘:‘…‘}

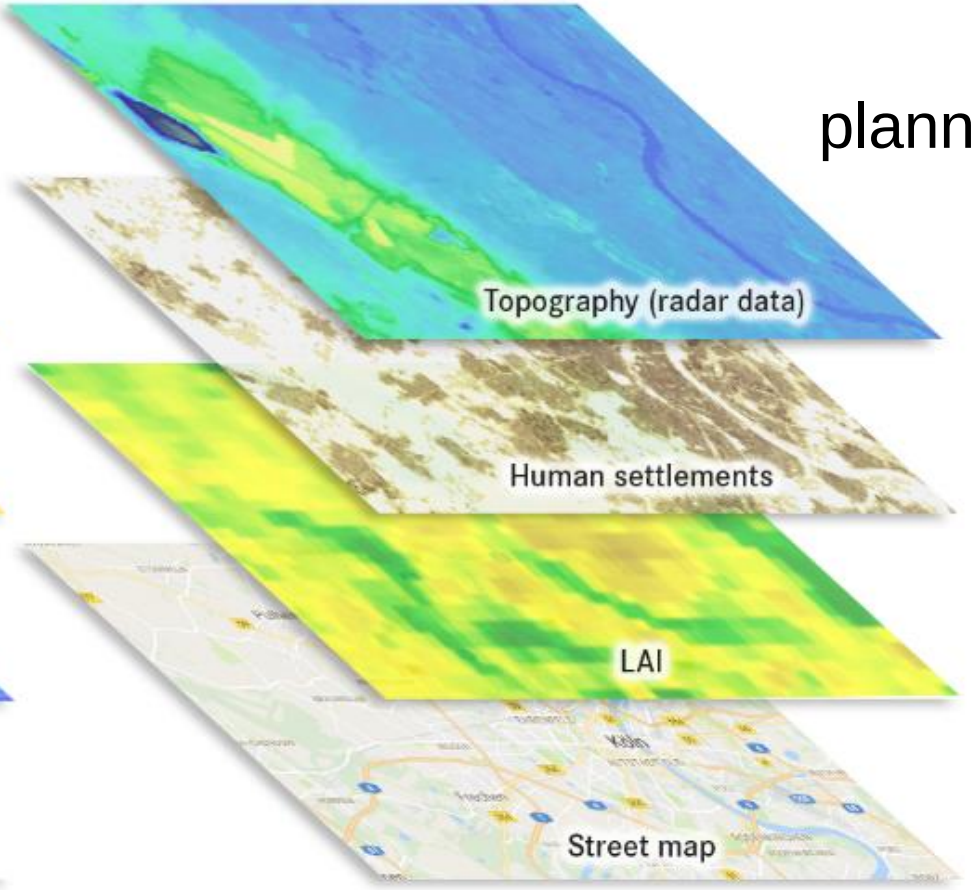
➔ All other metada stored in DB as json structure!

TOAR database (2nd phase)
metadata and web service infrastructures

actual

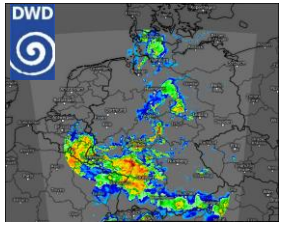


planned



TOAR database (2nd phase)

web service infrastructures



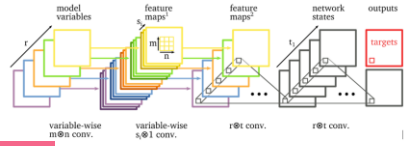
meteoCloud

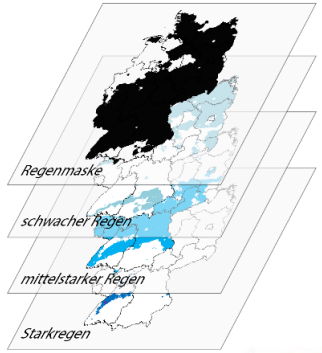
Workflow Performance

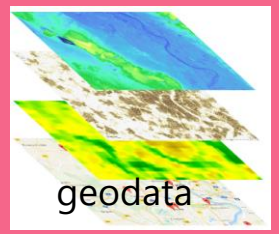
MeteoCube

Workflow Performance

Data Analytics



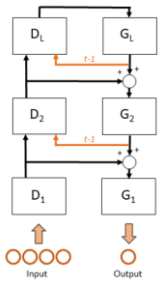




Workflow

GeoCube

Workflow Services



Products Services Usability

Metadata generation

openaq

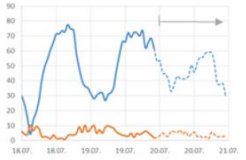
Umwelt Bundesamt

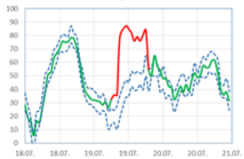
WMO GAW

Workflow Data quality Provenance

TOAR DB

Workflow Services





Deep Learning
Clustering
Validation

Conclusions: How far are we with respect to FAIRness?

FAIR Principles (FORCE 11)

TO BE FINDABLE:

F1. (Meta)data are assigned globally unique and persistent identifiers ✓

F2. Data are described with rich metadata ✓

F3. Metadata clearly and explicitly include the identifier of the data they describe ✓

F4. (Meta)data are registered or indexed in a searchable resource ✓

TO BE INTEROPERABLE:

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation ✓

I2. (Meta)data use vocabularies that follow the FAIR principles ✓

I3. (Meta)data include qualified references to other (meta)data ✓

TO BE ACCESSIBLE:

A1. (Meta)data are retrievable by their identifier using a standardised communication protocol (✓)

A1.1 The protocol is open, free and universally implementable ✓

A1.2 The protocol allows for an authentication and authorisation where necessary ✓

A2. Metadata should be accessible even when the data is no longer available ✓

TO BE RE-USABLE:

R1. (Meta)data are richly described with a plurality of accurate and relevant attributes ✓

R1.1. (Meta)data are released with a clear and accessible data usage license ✓

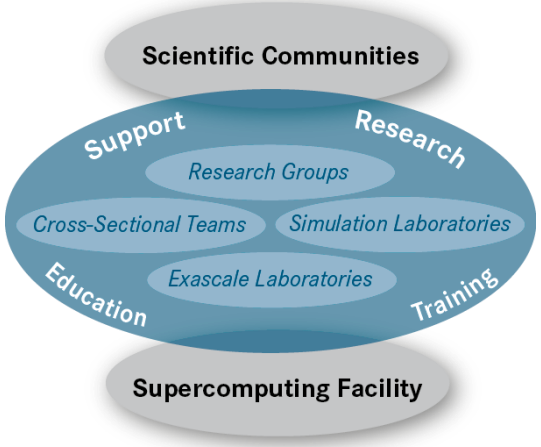
R1.2. (Meta)data are associated with detailed provenance ✓

R1.3. (Meta)data meet domain-relevant community standards ✓

acknowledgement

Jülich Supercomputing Centre (JSC)

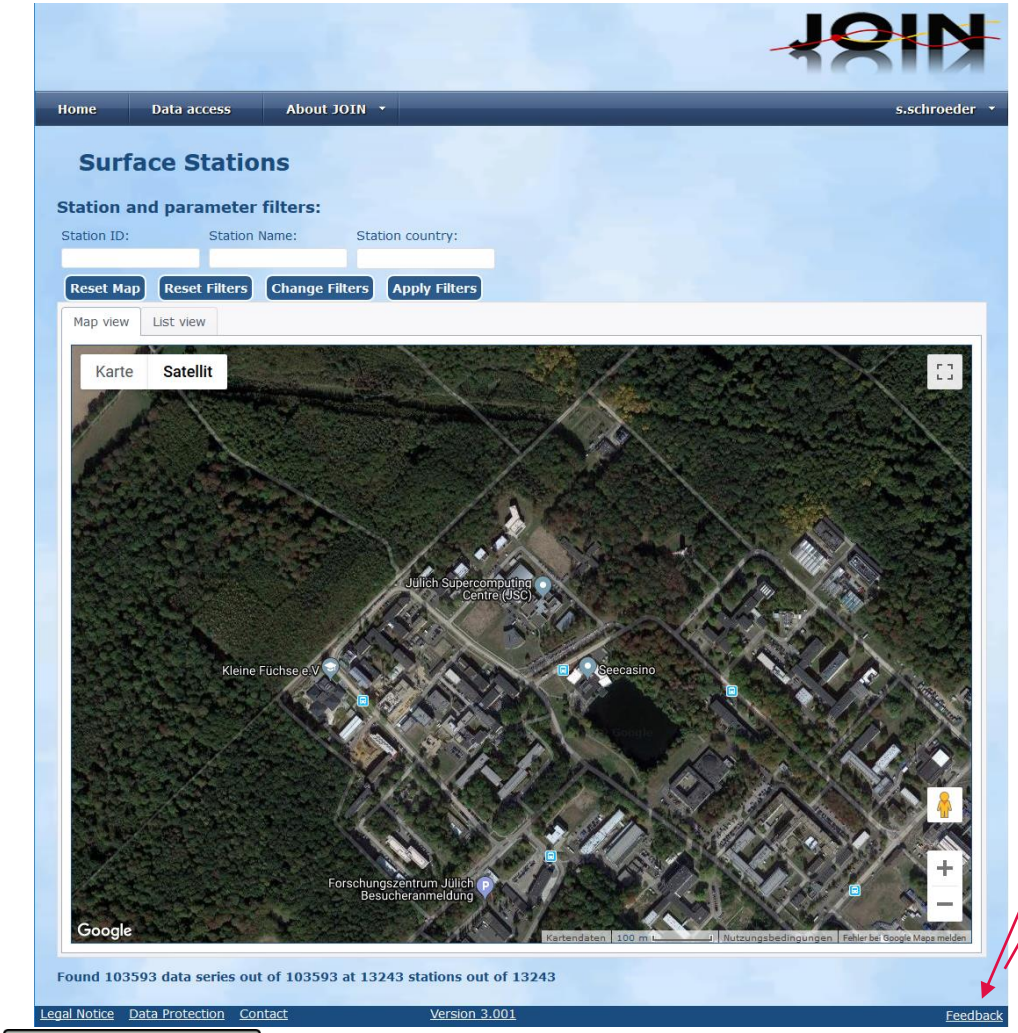
Research & Development
Supercomputer operation for Centre/Region/Germany/Europe
Application support
Education and Training



for support and help using the infrastructure provided at Jülich.

Thank you

join.fz-juelich.de



Every feedback is welcome:

- click on „Feedback“ link
- email to join-support@fz-juelich.de