

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 | SCHRÖDER, S.; APWEILER, S.; SAINI, R.; HAGEMEIERS, B.; SCHULTZ, M. G.

JÜLICH SUPERCOMPUTING CENTRE



Diese Präsentation steht unter der Lizenz Creative Commons Namensnennung 4.0 International (CC BY 4.0).
This presentation is licensed under a Creative Commons Attribution 4.0 International License (CC-BY 4.0).

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.

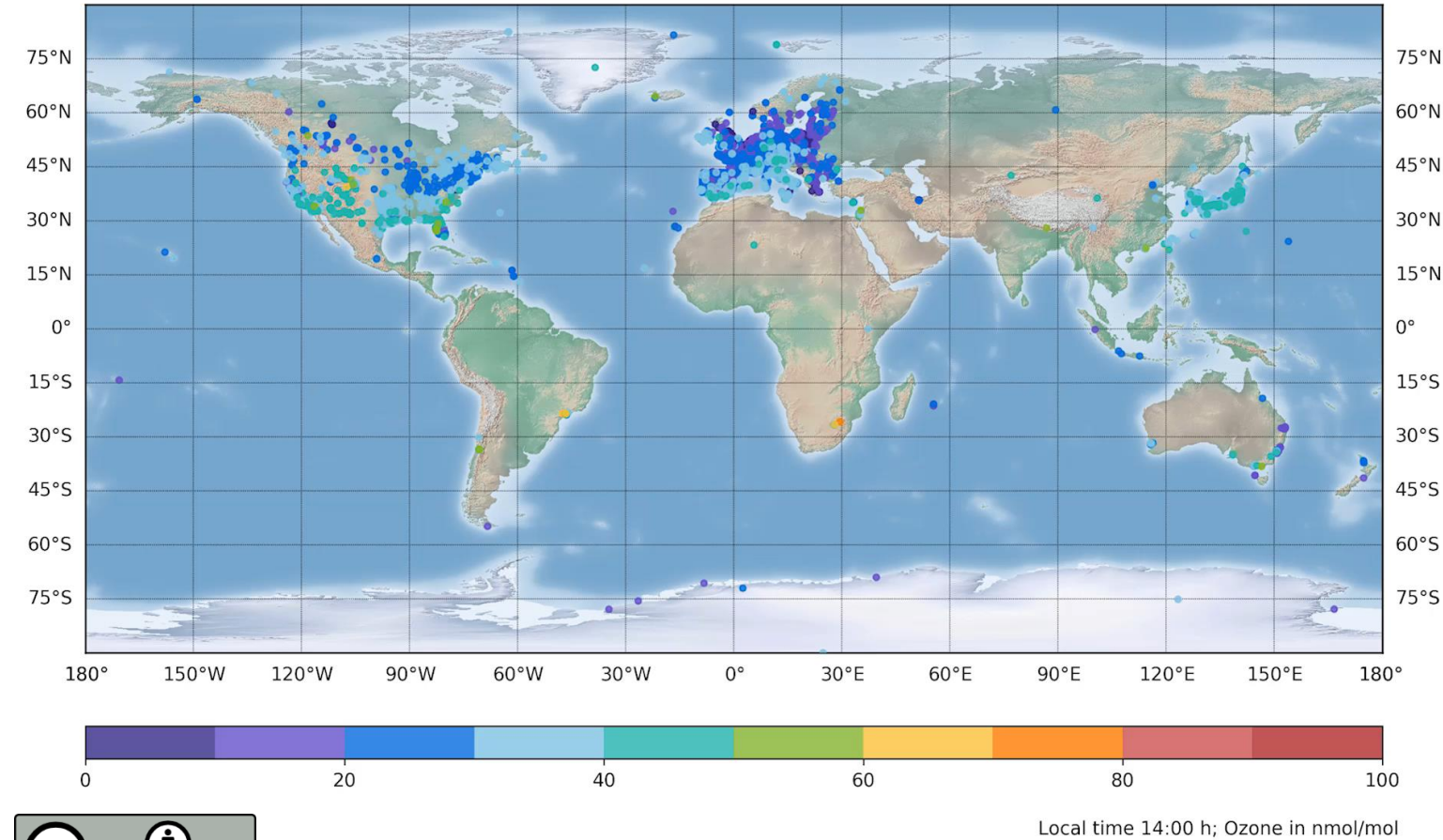
Web link: <http://www.igacproject.org/activities/TOAR>

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



Motivation

01 Jan 2008



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 NO₂
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

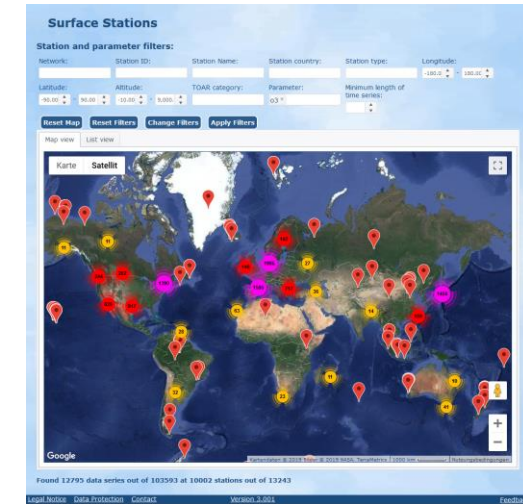


TOAR database services

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

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

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



<https://join.fz-juelich.de/services/rest/surfacedata/parameters/>
→
["albedo", "aswdifu", "aswdir", "benzene",
"ch4", "cloudcover", "co", "ethane",
"humidity", "no", "no2", ..., "wspeed"]



TOAR database

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

Data extraction options

Variable: O3

Date: to

Station information: Station_id: jp08402010, Network_name: NIES, Station_name: Hokotahokenjo, Parameter_PI: H.Tanimoto, Parameter_pi_email: tanimoto@nies.go.jp

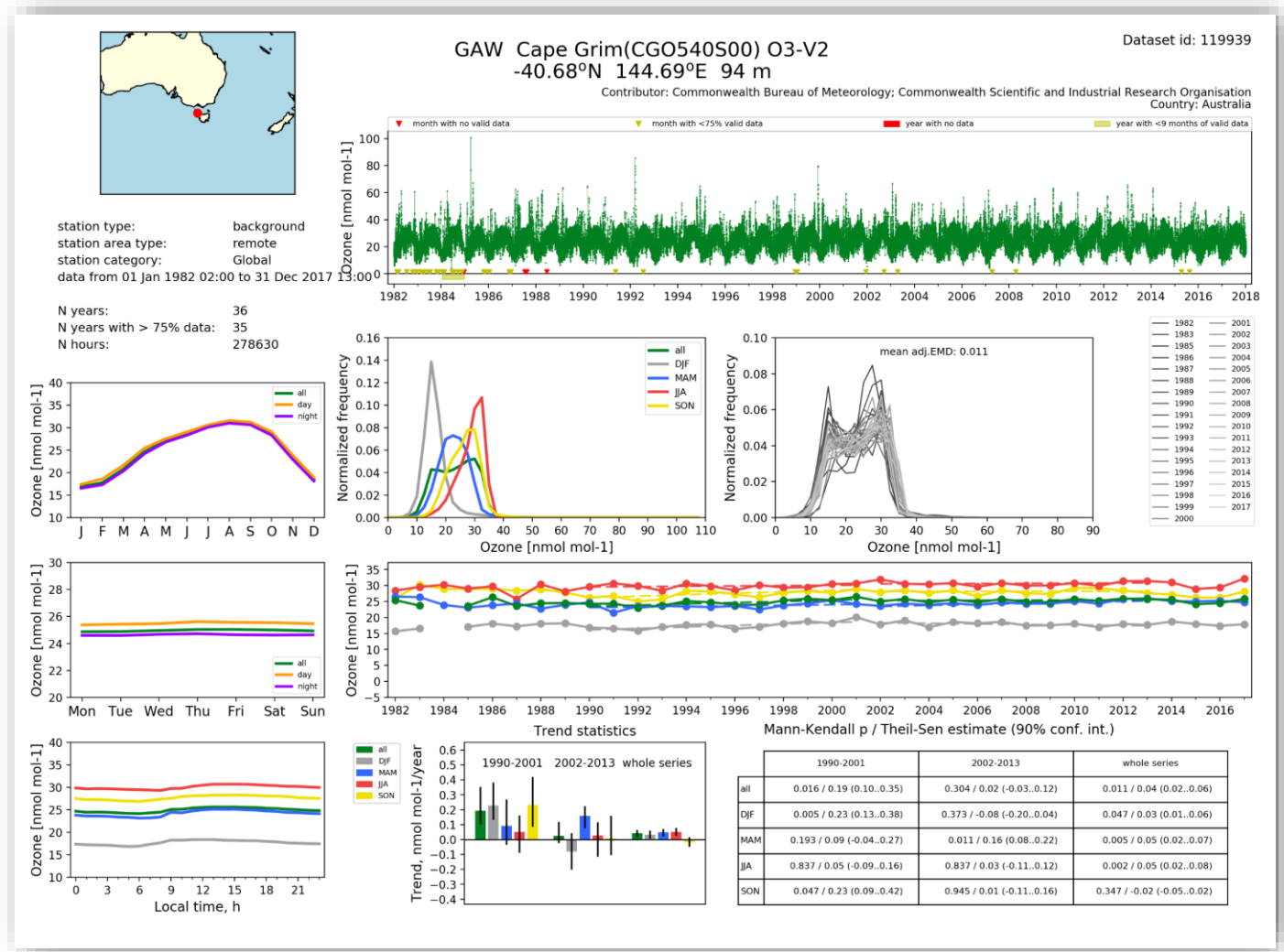
Data quality flag: OK not_checked

Hourly values:

Data summary:

Statistics	Daily	Monthly	Seasonal	Summer	Annual
Count	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Data capture	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Average values	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Standard deviation	--	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Ozone metrics	Daily	Monthly	Seasonal	Summer	Annual
Max. 1-h value	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Daily max. 8-h average (US EPA definition)	<input type="checkbox"/>	--	--	<input type="checkbox"/>	<input type="checkbox"/>
Daily max. 8-h average (proposed new US EPA definition)	<input type="checkbox"/>	--	--	<input type="checkbox"/>	<input type="checkbox"/>
Daily max. 8-h average (EU)	<input type="checkbox"/>	--	--	<input type="checkbox"/>	<input type="checkbox"/>
Daily max. 8-h average (US EPA definition, strict data capture)	<input type="checkbox"/>	--	--	<input type="checkbox"/>	<input type="checkbox"/>
Daily max. 8-h average	<input type="checkbox"/>	--	--	<input type="checkbox"/>	<input type="checkbox"/>



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

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

[columns=network_name,station_id,station_name](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/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

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

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

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

```
▼ 0:
0: "AIRBASE"
1: "DENW141"
2: "Münster Steinfurter Straße"
▼ 1:
0: "UBA"
1: "DENW141"
2: "Münster Steinfurter Straße"
▼ 2:
0: "AIRBASE"
1: "DENW095"
2: "Münster-Geist"
▼ 3:
0: "UBA"
1: "DENW260"
2: "Münster Weseler Straße"
▼ 4:
0: "UBA"
1: "DENW199"
2: "Münster Weseler Straße 22"
▼ 5:
...
▼ 9:
0: "UBA"
1: "DENW095"
2: "Münster-Geist"
```

le query arguments to the base URL.



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



- 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

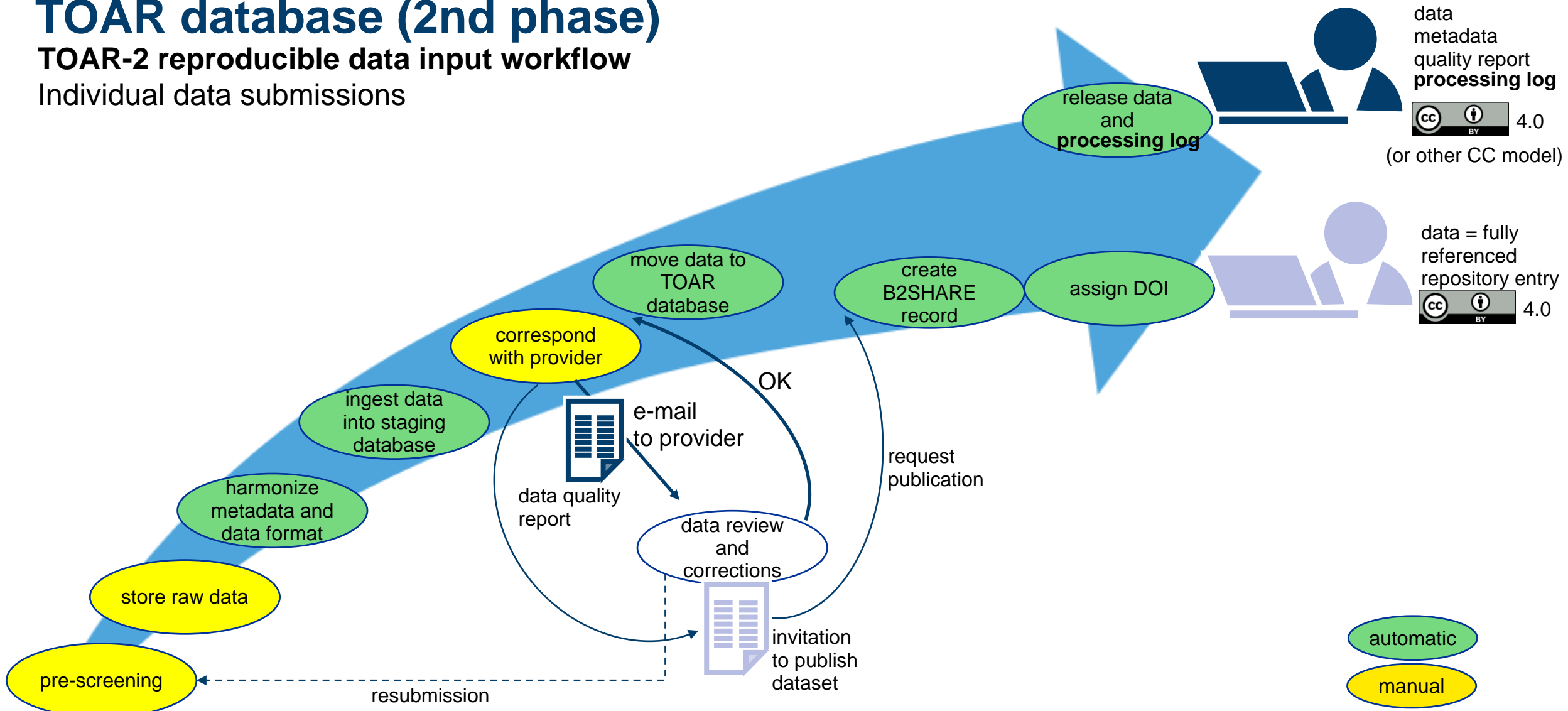
in preparation



TOAR database (2nd phase)

TOAR-2 reproducible data input workflow

Individual data submissions



data
metadata
quality report
processing log
CC BY 4.0
(or other CC model)

data = fully
referenced
repository entry
CC BY 4.0

automatic
manual



TOAR database (2nd phase)

metadata

Station_id: ...
Station_name: ...

predefined keywords

...
Comments: Some example text: any number of arbitrary free text with additional information

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': '...'}

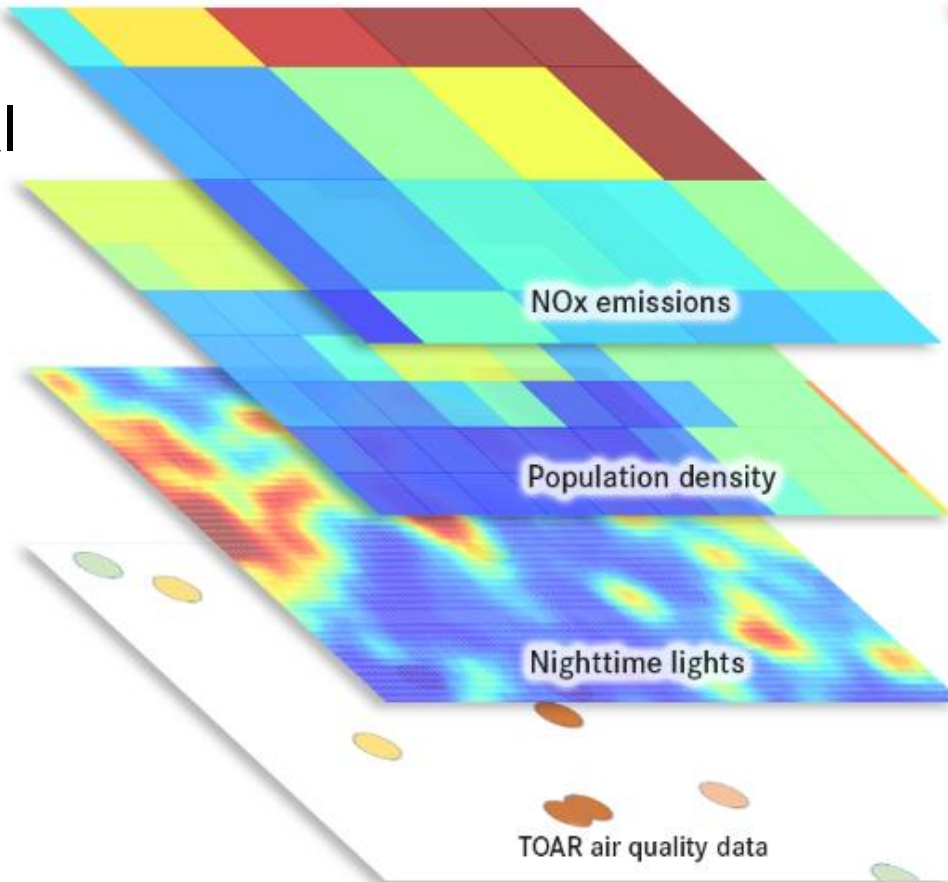
→ All other metadata stored in DB as json structure!



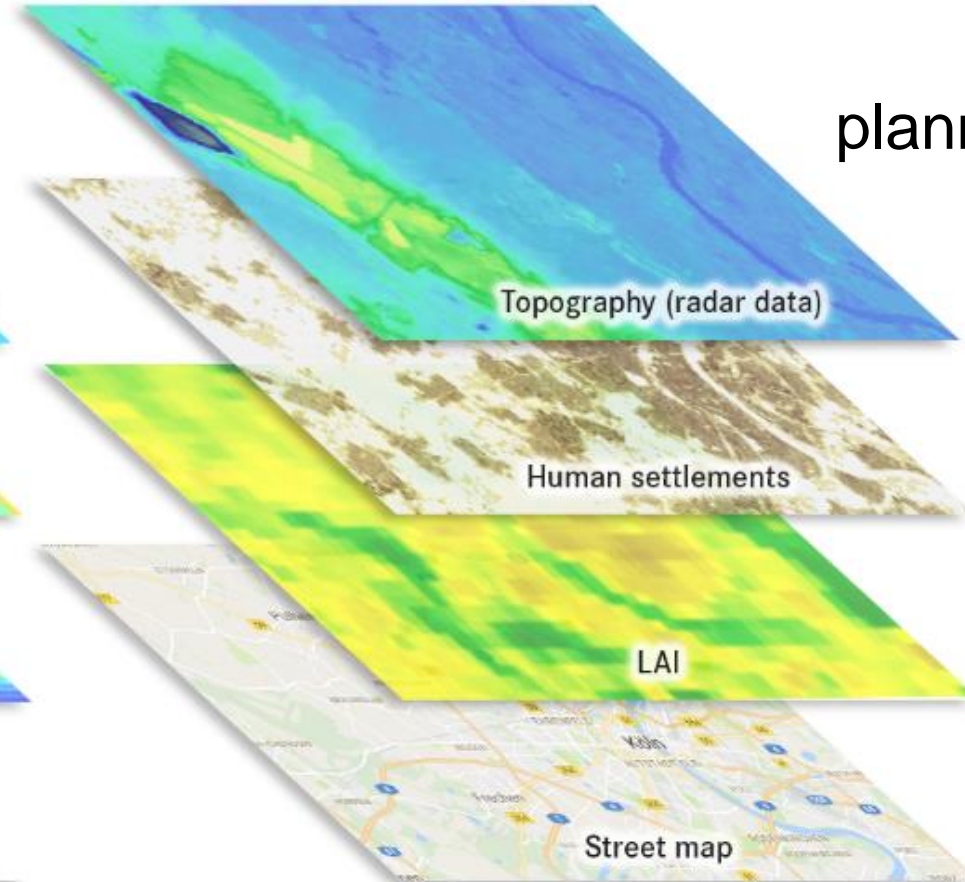
TOAR database (2nd phase)

metadata and web service infrastructures

actual

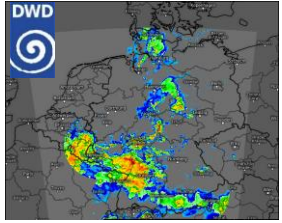


planned



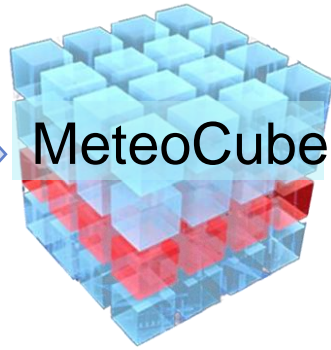
TOAR database (2nd phase)

web service infrastructures



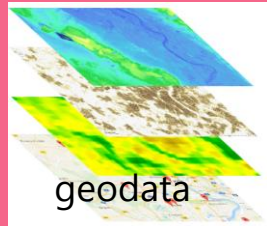
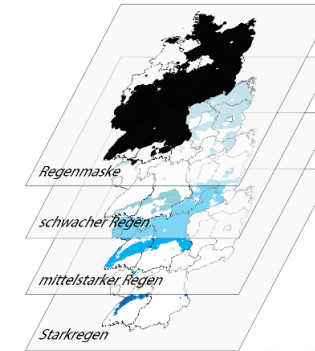
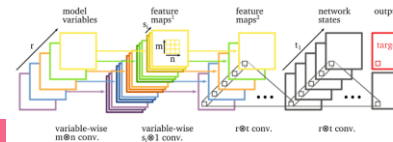
meteoCloud

Workflow Performance



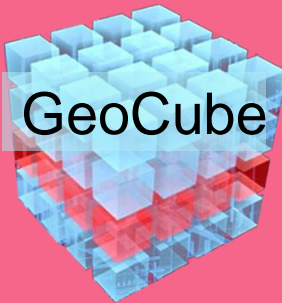
Workflow Performance

Data Analytics



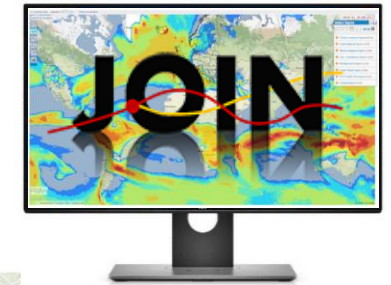
geodata

Workflow

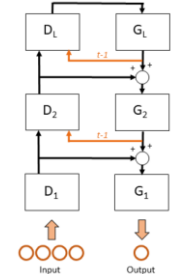


Workflow Services

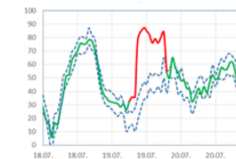
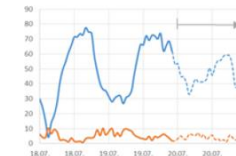
Products Services Usability



Metadata generation



Deep Learning Clustering Validation



Workflow Data quality Provenance



Workflow Services



Conclusions: How far are we with respect to FAIRness?

FAIR Principles (FORCE 11)

TO BE **F**INDABLE:

- 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 **A**CCESIBLE:

- 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 **I**NTEROPERABLE:

- 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 **R**E-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)

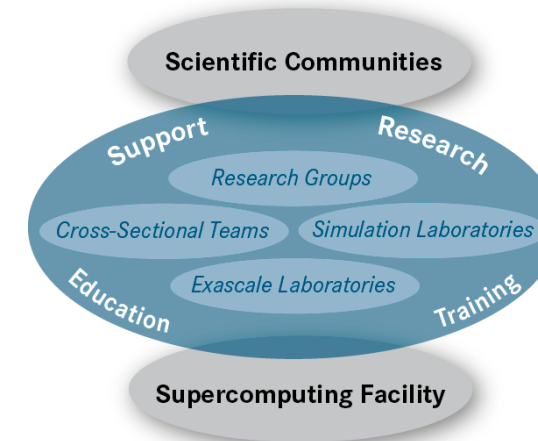
Research & Development

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

Home Data access About JOIN s.schroeder

Surface Stations

Station and parameter filters:

Station ID: Station Name: Station country:

Reset Map Reset Filters Change Filters Apply Filters

Map view List view

Karte Satellit

Julich Supercomputing Centre (JSC)

Kleine Füchse e.V.

Seecasino

Forschungszentrum Jülich Besucheranmeldung

Google

Found 103593 data series out of 103593 at 13243 stations out of 13243

Legal Notice Data Protection Contact Version 3.001 Feedback

Every feedback is welcome:

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

