

# The TOAR data infrastructure:

## A generalised database infrastructure for environmental time series


Sabine Schröder\*, Niklas Selke\*, Sander Apweiler, Clara Betancourt, Eleonora Epp, Björn Hagemeyer, Enxhi Kreshpa, Max Lensing, Lukas H. Leufen, Amirpasha Mozaffari, Tom Ohlmeyer, Mathilde Romberg, Rajveer Saini, Martin G. Schultz

\*equal contributions


### What we do



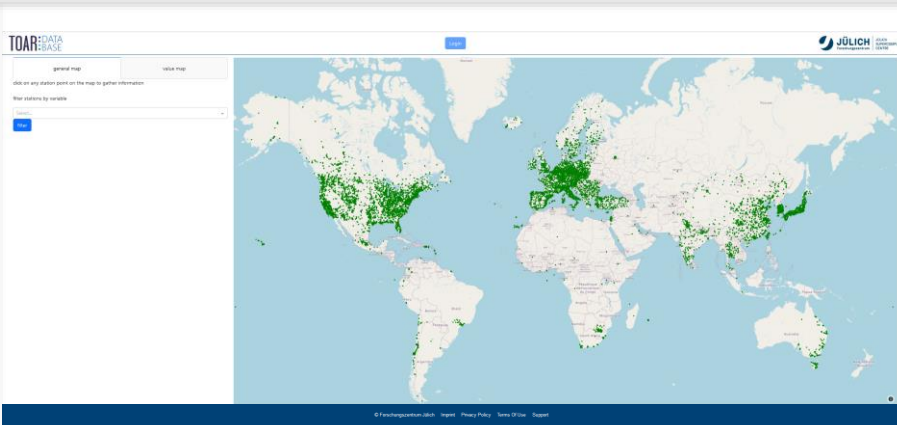

Collect and harmonize global air quality observations



Ingest near real time global air quality data



Use geospatial information to provide globally uniform station characterization

> 3.4 B. data points  
> 22796 stations  
150 users

### What you can do

Easy access to air quality data and analyses



Information about FAIR data, documentation, CoreTrust Seal



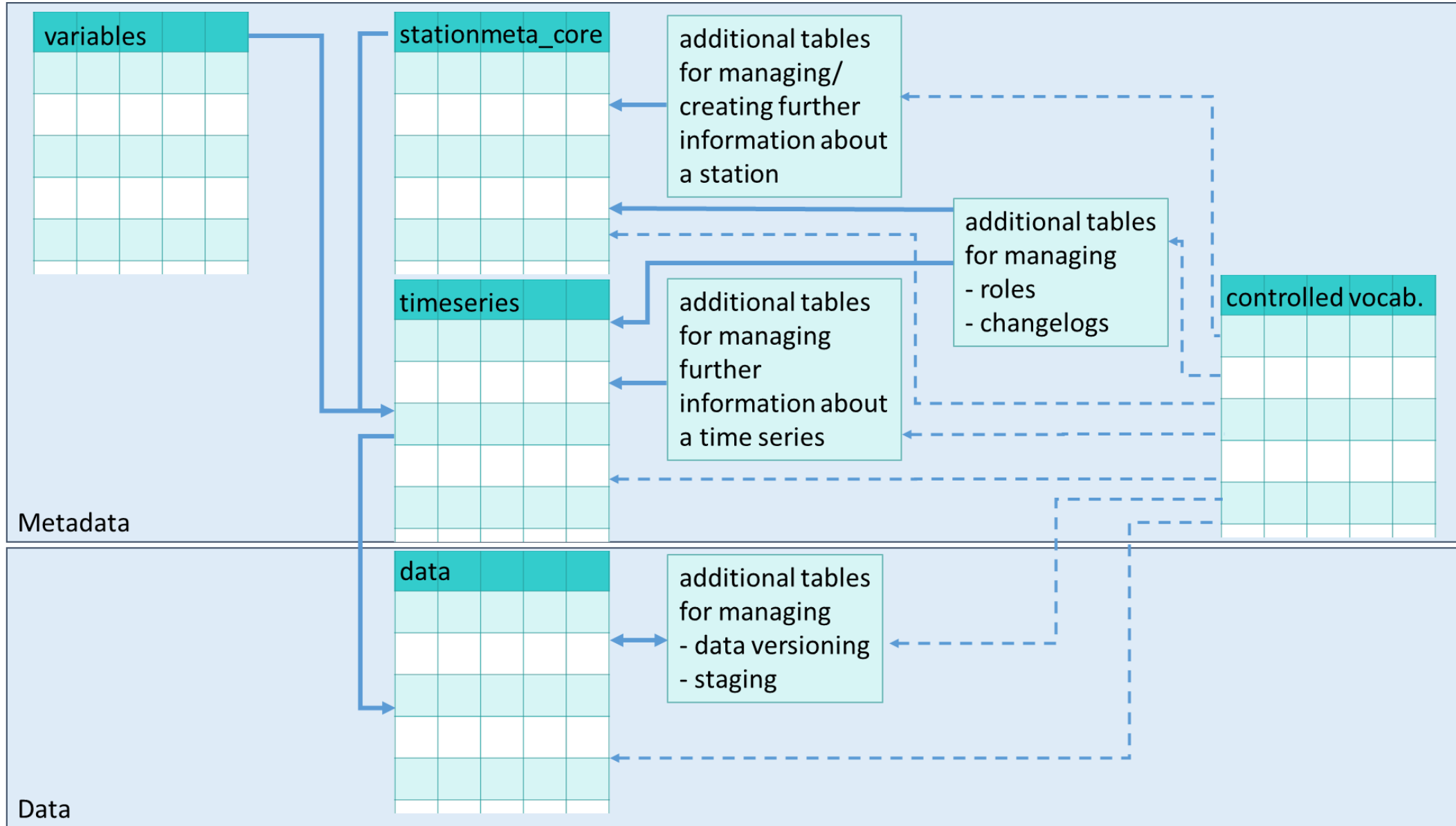
Use of TOAR data in ML applications



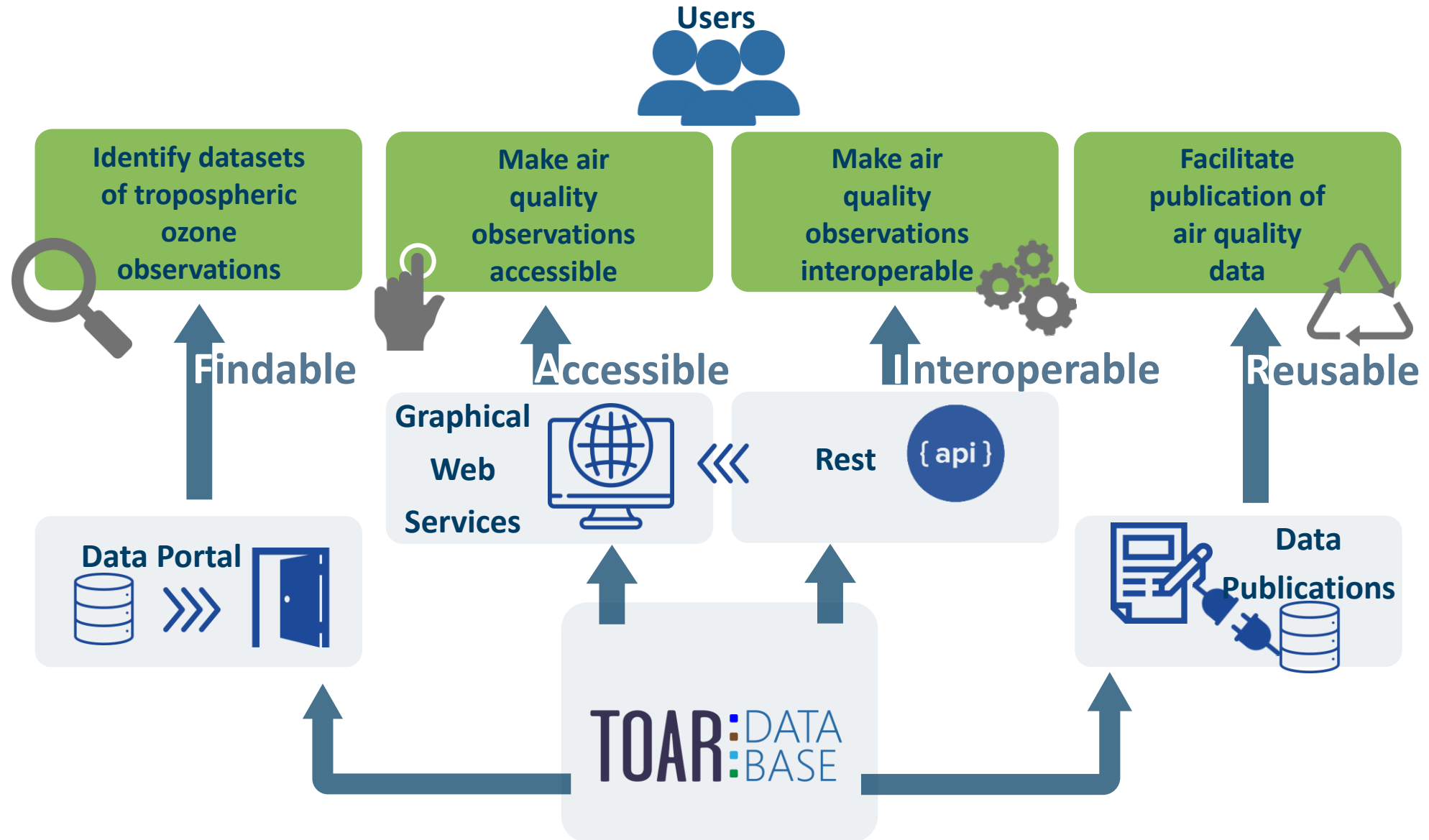
This is a nonlinear presentation. Please use the buttons to click through the content. Thanks!

Back to  
title slide

# The TOAR-II database schema



# FAIR data principles in the TOAR-II database



# The TOAR Database Infrastructure documentation

TOAR  
tropospheric  
ozone  
assessment  
report  
Phase II

JÜLICH  
Forschungszentrum | JÜLICH  
SUPERCOMPUTING  
CENTRE

TOAR Data User Guide #3

## The TOAR Database User Guide

Version 1.0 | 25 August 2021

Forschungszentrum Jülich GmbH  
ESDE | JSC - FSD  
52

**User guides**

TOAR  
tropospheric  
ozone  
assessment  
report  
Phase II

JÜLICH  
Forschungszentrum | JÜLICH  
SUPERCOMPUTING  
CENTRE

TOAR Data Technical Guide #2

## TOAR Data Input and Processing

Version 1.0 | 30 August 2021

Forschungszentrum Jülich GmbH  
ESDE | JSC - FSD

**Technical guides**

toardb\_fastapi

## Detailed documentation (gitlab pages)

TOAR II

**Models**

**Composite**

Name	Type	Description	Required
metadata		all metadata available to the data	Yes
data	[ <a href="#">Data</a> ]	the data itself	Yes

**Contact**

Name	Type	Description	Required
id	integer	for internal use only	Yes
person	<a href="#">Person</a>	A contact is either a person or an organisation	No
organisation	<a href="#">Organisation</a>	A contact is either a person or an organisation	No

**Coordinates**

Name	Type	Description	Required
lat	number	latitude coordinate of station (decimal degrees_north). This is our best estimate of the station location which is not always identical to the official station coordinates (see potential changelog entry).	Yes
lng	number	longitude coordinate of station (decimal degrees_east). This is our best estimate of the station location which is not always identical to the official station coordinates (see potential changelog entry).	Yes
alt	number	altitude of station (in m above sea level). This is our best estimate of the station altitude, which is not always identical to the reported station altitude, but frequently uses the elevation from google earth instead (see potential changelog entry).	Yes

**TOAR II Database**

- Models
  - Composite
  - Contact
  - Coordinates
  - Data
  - Organisation
  - Person
  - Stationmeta
  - StationmetaAnnotation
  - StationmetaAuxDoc
  - StationmetaAuxImage
  - StationmetaAuxUri
  - StationmetaChangelog
  - StationmetaGlobal
  - StationmetaRole
  - Timeseries
  - TimeseriesAnnotation
  - TimeseriesChangelog
  - TimeseriesProgramme
  - TimeseriesRole
  - Variable
    - Geolocation URLs
- Controlled Vocabulary
  - Role Code
  - Role Status
  - Kind Of Annotation
  - Kind Of Organization
  - Data Access Right
  - Sampling Frequency



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

**Destination Earth**

<https://toar-data.org>  
Contact: [info@toar-data.org](mailto:info@toar-data.org)

# Metadata change logs

By example:

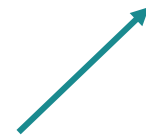
```
▼ changelog:
  ▼ 0:
    datetime: "2022-02-04T18:19:58.120019+00:00"
    ▼ description: "add climatic_zone_year2016 from geolocation service"
    old_value: '{"climatic_zone_year2016': 'Undefined'}"
    new_value: '{"climatic_zone_year2016': 'WarmTemperateDry'}"
    station_id: 2
    author_id: 1
    type_of_change: "single value correction in metadata"
  ▼ 1:
    datetime: "2022-04-14T12:55:47.800754+00:00"
    description: "add htap_tier1_region from GEO PEAS"
    ▼ old_value: '{"htap_region_tier1_year2010': 'HTAPTier1Undefined'}"
    new_value: '{"htap_region_tier1_year2010': 'HTAPTier1SAF'}"
    station_id: 2
    author_id: 1
    type_of_change: "single value correction in metadata"
```



## Time series versioning

As part of our curation efforts we may apply changes to original data. All changes are logged, timeseries are versioned, and our data quality flags are traceable.

Version = **major**.**minor**.**micro**



Used to store a date label (especially for realtime data)





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

Destination Earth

<https://toar-data.org>  
Contact: [info@toar-data.org](mailto:info@toar-data.org)

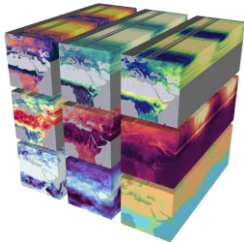




**GEO**spatial  
Point **E**xtraction and **A**ggregation **S**ervice

We store temporally high-resolved geo-data (such as population density or stable night lights) in a rasdaman-database, which offers access using geographic coordinates or time specifications.

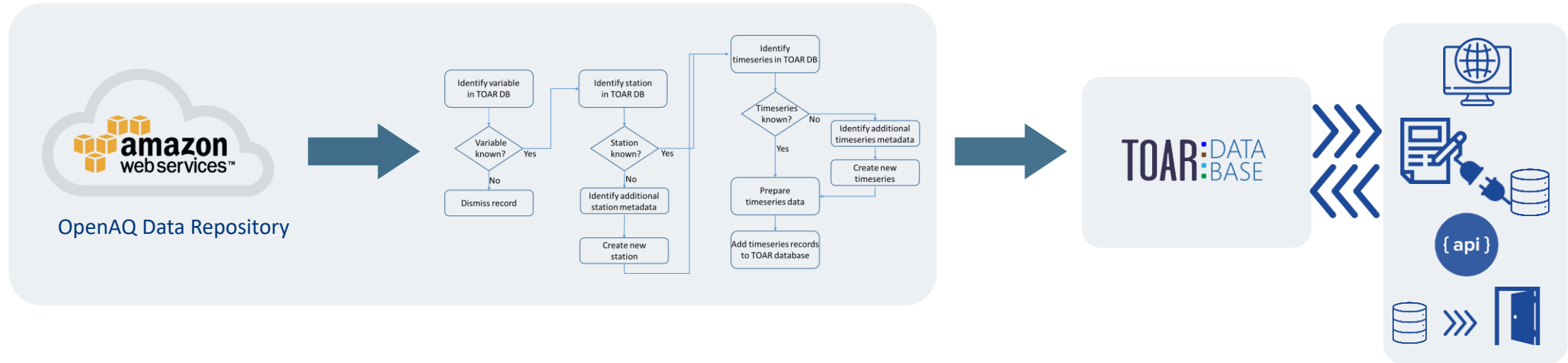
GEO-PEAS provides end-points for different geographical information and can proceed with requests given geographical coordinates (and a radius around this location) and year. The information can either be requested as a point extraction or as spatial aggregation (mean, std dev, median, min, max, sum, or different percentiles).



We included the service call in the automated data ingestion workflow, so that already during the data ingestion for each station the extended metadata is calculated and made available.



# Ingesting Real-Time Data (example: OpenAQ)



**OpenAQ is the world's first open, real-time and historical air quality data platform:**  
 Mainly government sources  
 Research-grade data

**Collective impact through community:**

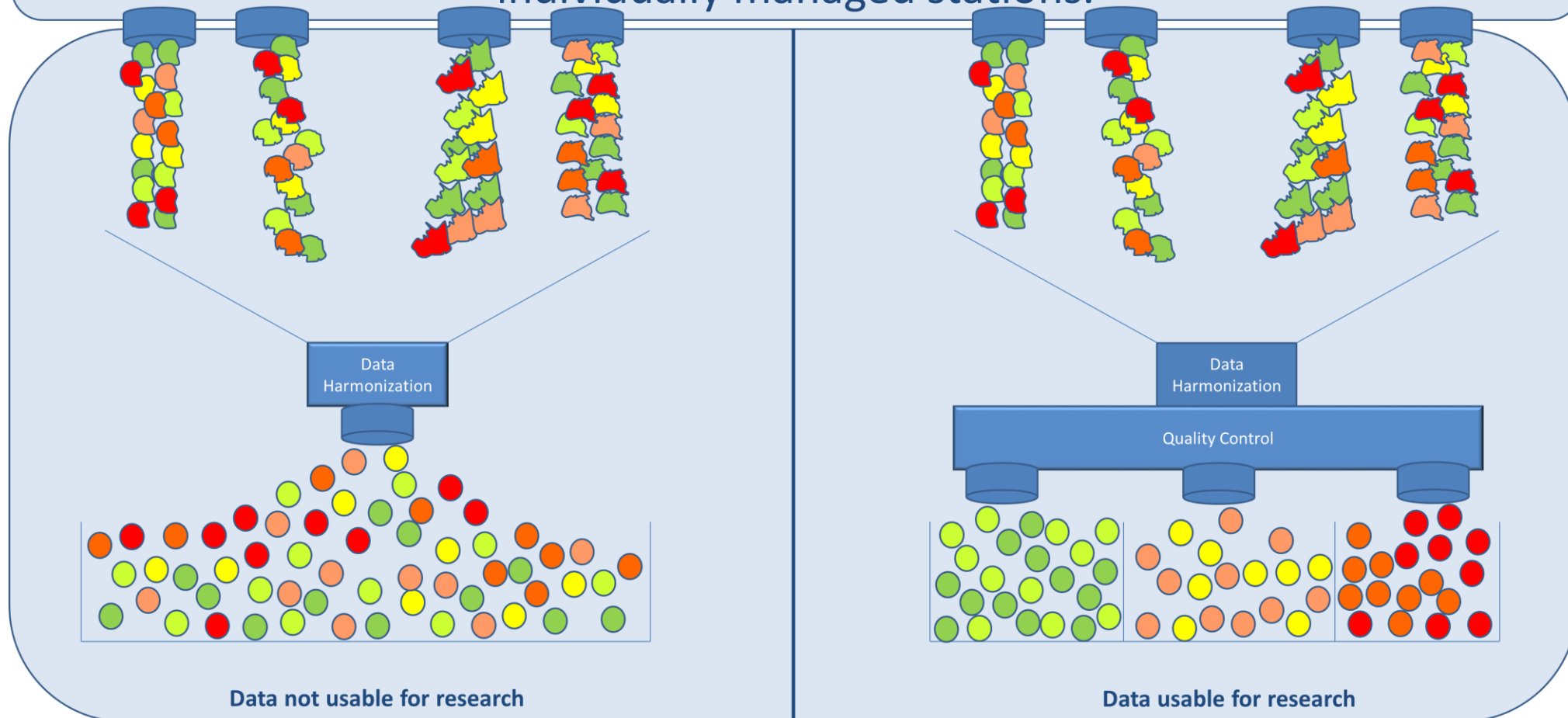
- Data platforms
- Developer tools
- Education tools
- Low-cost sensors
- Media
- Models
- Research
- Software



# Data Collection, Harmonization, and Quality Control

## Data Sources

Data is provided by various providers from large measurement networks to individually managed stations.



# Controlled Vocabulary / Ontology

toar-data.fz-juelich.de/api/v2/contr X

https://toar-data.fz-juelich.de/api/v2/controlled\_vocabulary/

JSON Rohdaten Kopfzeilen

Speichern Kopieren Alle einklappen Alle ausklappen (langsam) JSON durchsuchen

▼ Role Code:

- 0: 0
- 1: "pointofcontact"
- 2: "point of contact"

▶ 1: [-]

- ▶ 2: [-]
- ▶ 3: [-]
- ▶ 4: [-]
- ▶ 5: [-]
- ▶ 6: [-]
- ▶ 7: [-]
- ▶ 8: [-]

▶ Role Status: [-]

▶ Kind Of Annotation: [-]

▶ Kind Of Organization: [-]

▶ Sampling Frequency: [-]

▶ Aggregation Type: [-]

▶ Data Origin Type: [-]

▶ Data Origin: [-]

▶ Climatic Zone 2019: [-]

▶ Country Code: [-]

▶ Timezone: [-]

▶ Station Coordinate Validity: [-]

▶ Station Type: [-]

▶ Station Type Of Area: [-]

▶ Station TOAR Category: [-]

▶ Station HTAP Region: [-]

▶ Station Landcover Type: [-]

▶ Station ECO Region Type: [-]

▶ Result Type: [-]

▶ Data Flag: [-]

▶ Type Of Change: [-]

▶ Absorption Cross Section: [-]

▶ Sampling Type: [-]

▶ Calibration Type: [-]

OWLDoc

https://toar-data-dev.fz-juelich.de/documentation/

Ontologies Classes Object Properties Data Properties Annotation Properties Individuals Datatypes Clouds

Class: role

**Annotations (4)**

- rdfs:isDefinedBy "https://esde.pages.jsc.fz-juelich.de/toar-data/toar\_db\_fastapi/docs/toar\_db\_fastapi.html#role-code"(xsd:string)
- rdfs:isDefinedBy "https://www.ngdc.noaa.gov/wiki/index.php/ISO\_19115\_and\_19115-2\_CodeList\_Dictionaries#CI\_RoleCode"(xsd:string)
- rdfs:label "role" @en
- skos:definition "Role of contact (see controlled vocabulary: Role Code)"(xsd:string)

**Superclasses (1)**

- roles

**Members (9)**

Collaborator, Contributor, Custodian, Originator, PointOfContact, PrincipalInvestigator, ResourceProvider, RightsHolder, Stakeholder

OWLDoc

https://toar-data-dev.fz-juelich.de/documentation/

Ontologies Classes Object Properties Data Properties Annotation Properties Individuals Datatypes Clouds

ontologies

<https://toar-data.fz-juelich.de/documentation/ontologies/v1.0>

**Annotations (1)**

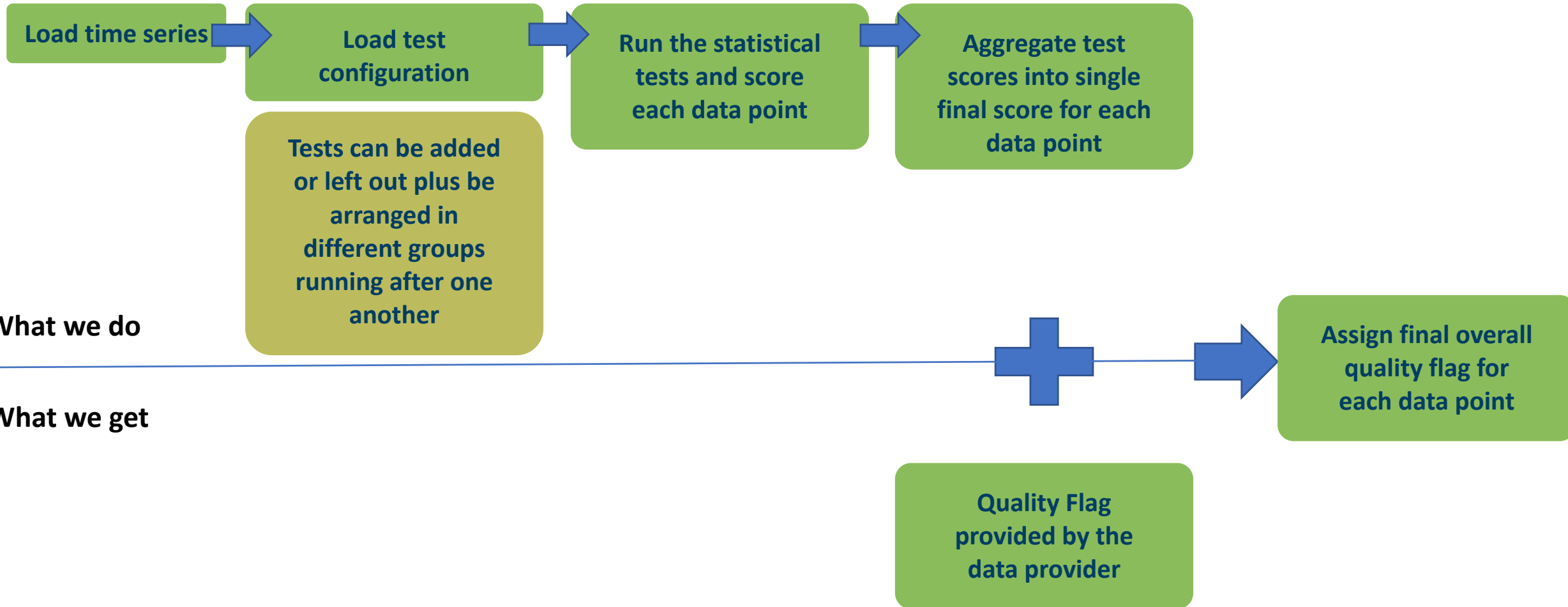
- rdfs:comment "definition of Metadata fields for the TOAR database" @en

**References**

- Classes (169)
- Object Properties (2)
- Data Properties (2)
- Annotation Properties (8)
- Individuals (1946)
- Datatypes (3)



# Data Quality Control



# Easy access to air quality data and analyses

TOAR Data at JSC

toar-data.fz-juelich.de

TOAR:DATA BASE

CORE TRUST SEAL

JÜLICH Forschungszentrum JÜLICH SUPERCOMPUTING CENTRE

## TOAR Data Infrastructure

### Mission Statement

The Tropospheric Ozone Assessment Report (TOAR) data centre is the central hub for data access in support of research assessing the impacts of ozone air pollution on human health, vegetation, and climate. Besides maintaining a [data portal](#) with links to ozone data sets from research organisations all over the world, we operate a database of harmonised surface ozone measurements and related data. This is one of the largest collections of quality controlled air pollution measurements in the world. All data in the database are easily accessible through open, freely available and well documented web services. The TOAR data centre team is committed to the FAIR principles and aims to achieve the highest standards with respect to data curation, archival, and re-use.

[Terms of Use](#)

## TOAR V2

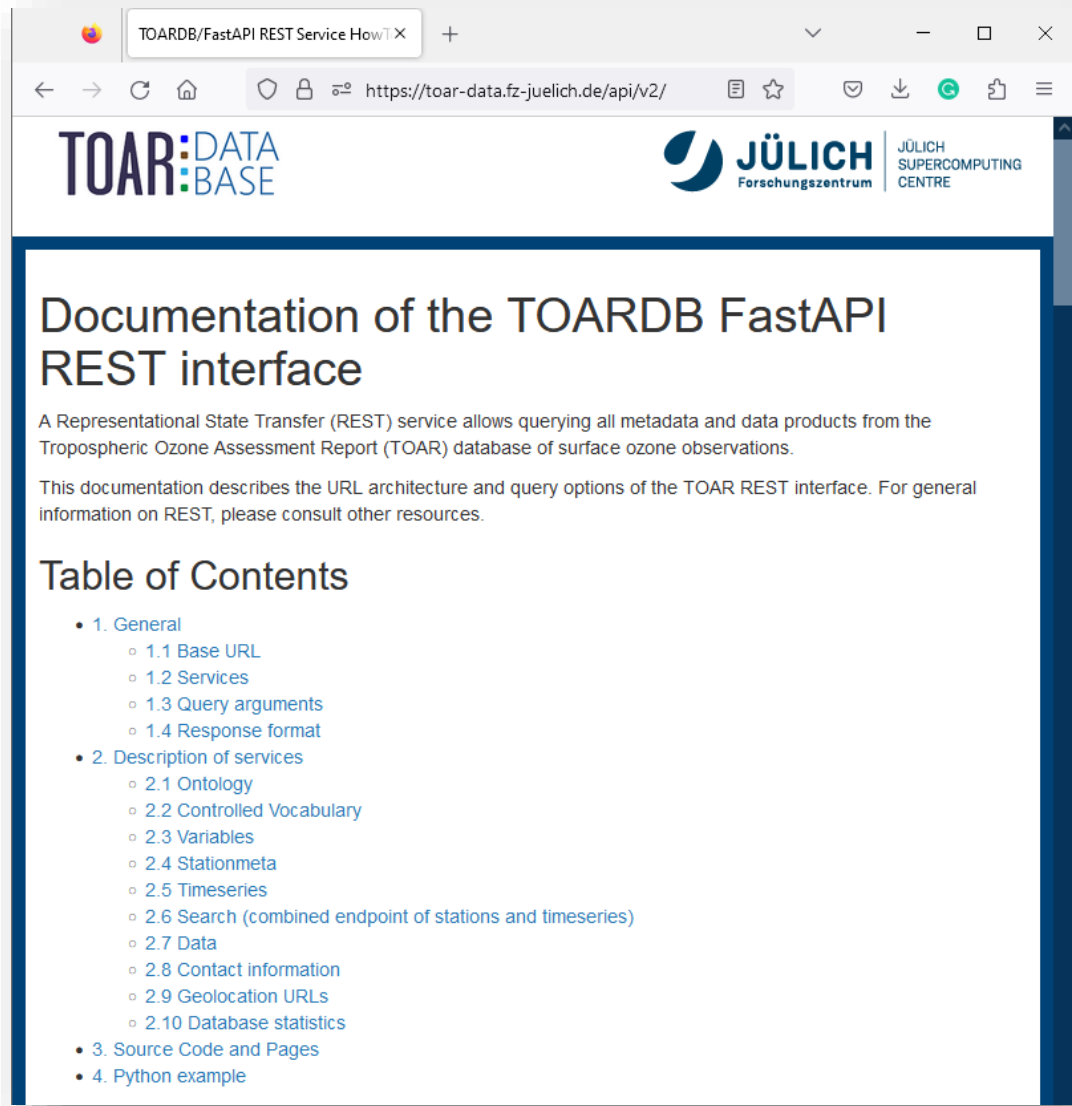
TOAR V2 supports the [second phase of the TOAR activity](#)

### Services

- [REST API to access TOAR V2](#)
- [GUI \(graphical user interface\) to access TOAR V2 \(under development\)](#)
- [DO3SE \(beta version\) REST API service for flux-based vegetation damage assessment](#)
- [analysis services REST API service for bulk data downloads and statistics on TOAR data](#)
- [GEO PEAS REST API to GEO PEAS \(GEOdata Point Extraction and Aggregation Service\)](#)

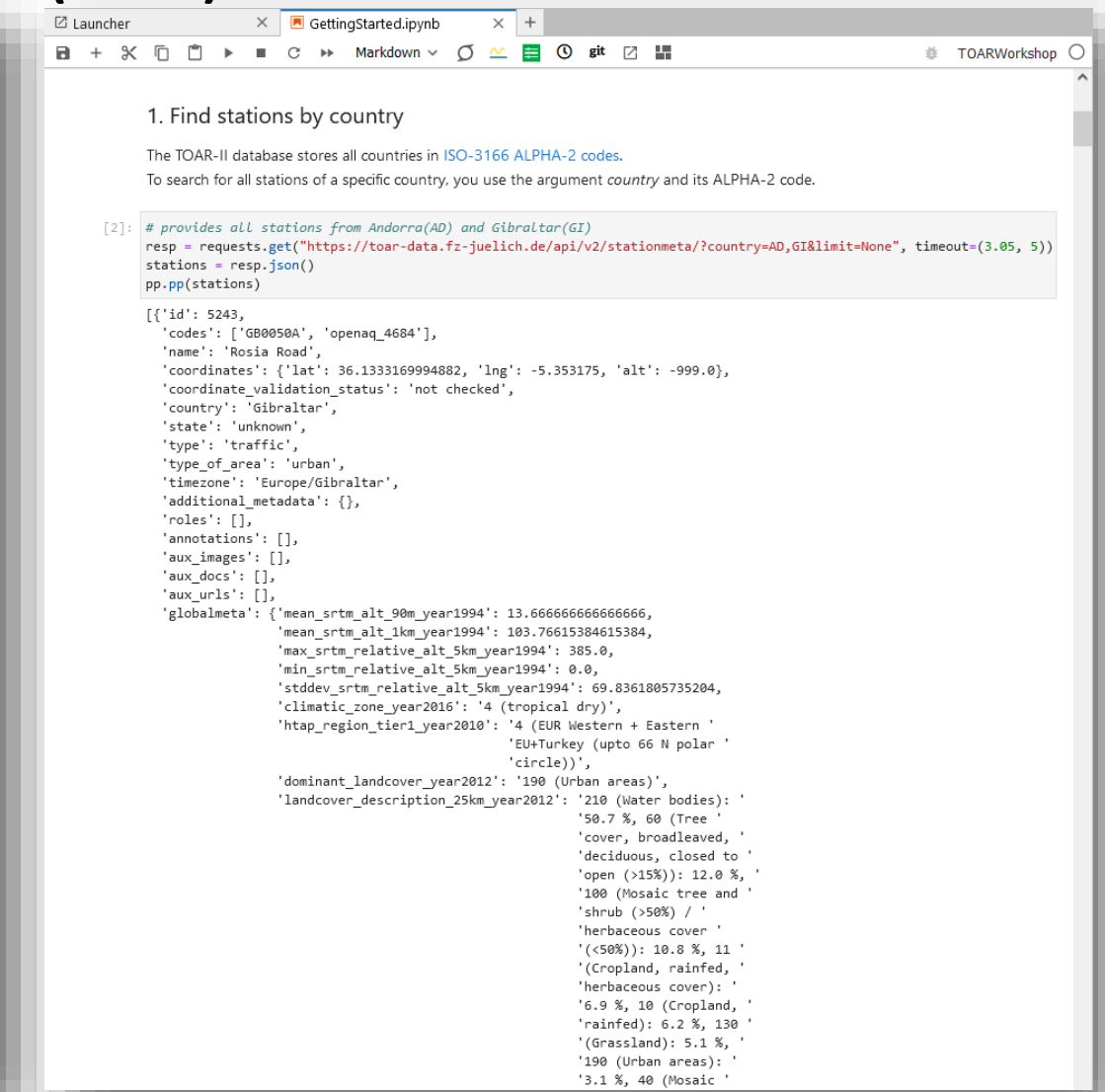


# Access TOAR-II (meta) data



The screenshot shows a web browser window displaying the TOARDB FastAPI REST interface documentation. The page title is "Documentation of the TOARDB FastAPI REST interface". It includes a table of contents with sections for General, Description of services, Source Code and Pages, and Python example. The TOAR logo and Jülich Supercomputing Centre logo are visible at the top.

from documentation





The screenshot shows a Jupyter Notebook window with a Python script that uses the requests library to call the TOARDB API. The script retrieves station data for Gibraltar and prints the response. The output shows a list of station details, including coordinates, country, and various metadata fields.

to Python notebook



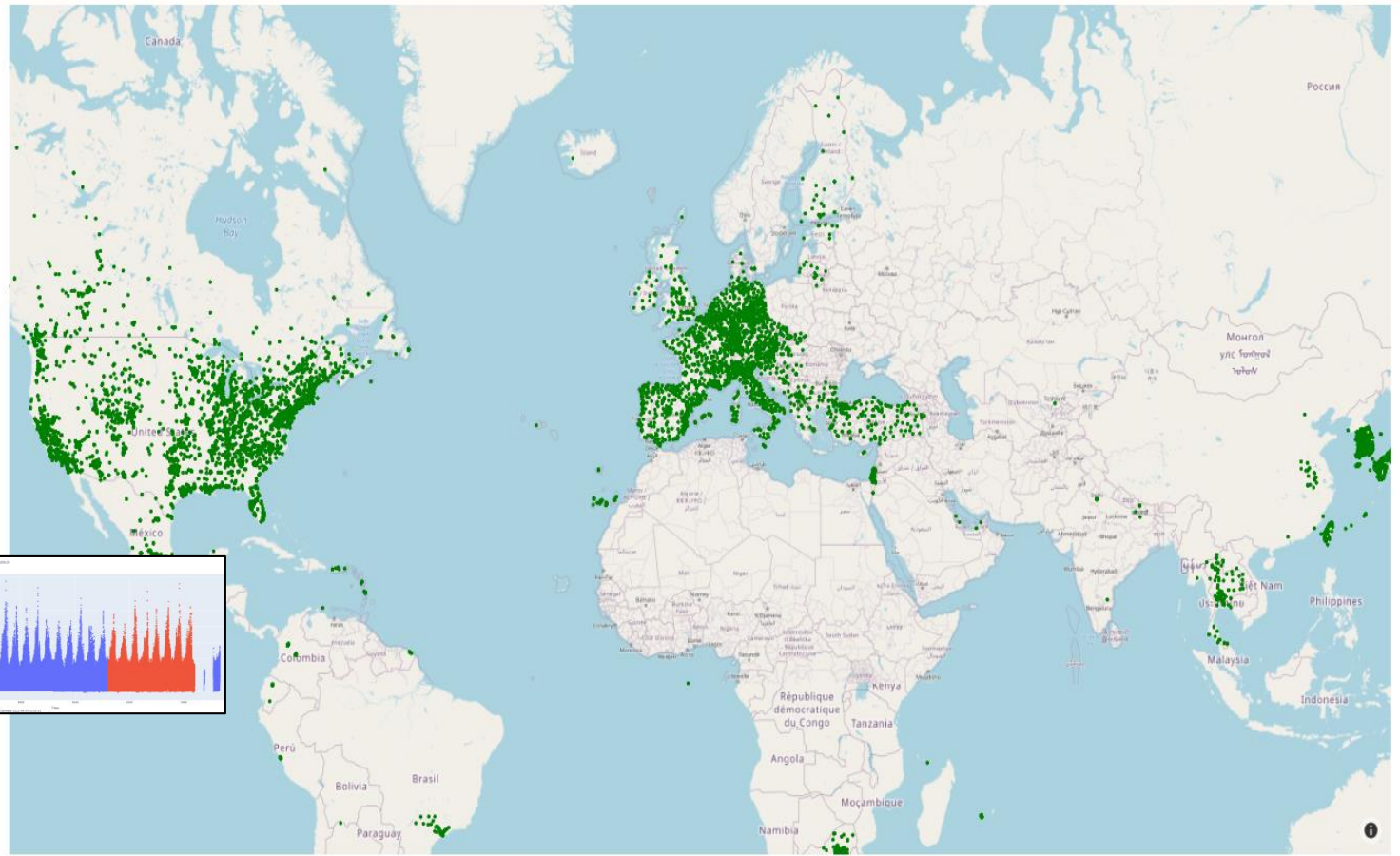
# GUI – general map





JÜLICH  
SUPERCOMPUTING  
CENTRE

	general map	value map
name:	Niederzier	
id:	46	
code:	DENW074	
country:	Germany	
state:	Nordrhein-Westfalen	
type_of_area:	rural	
variables:	<input type="checkbox"/> wspeed (UBA) <input type="checkbox"/> wdir (UBA) <input type="checkbox"/> o3 (UBA) <input type="checkbox"/> pm10 (UBA) <input type="checkbox"/> pm2p5 (UBA) <input type="checkbox"/> o3 (EEA) <input type="checkbox"/> pm2p5 (EEA) <input type="checkbox"/> pm10 (EEA) <input type="checkbox"/> pm10 (EEA) <input type="checkbox"/> pm10 (EEA) <input type="checkbox"/> temp (ECMWF) <input type="checkbox"/> press (ECMWF) <input type="checkbox"/> dcloudcover (ECMWF) <input type="checkbox"/> pblheight (ECMWF) <input type="checkbox"/> u (ECMWF) <input type="checkbox"/> v (ECMWF) <input type="checkbox"/> totprecip (ECMWF) <input type="checkbox"/> tdew2m (ECMWF) <input type="checkbox"/> wdir (ECMWF) <input type="checkbox"/> wspeed (ECMWF) <input type="checkbox"/> relhum (ECMWF) <input type="checkbox"/> humidity (ECMWF) <input type="checkbox"/> irradiance (ECMWF)	
action:	<div style="border: 1px solid black; padding: 5px;"> <input type="button" value="create plot (seperated timeseries)"/>  <input checked="" type="button" value="create plot (combined timeseries)"/>  <input type="button" value="create quality control plot"/>  <input type="button" value="create data summary plot"/> </div>	
filter stations by variable	<input type="text" value="o3"/> <input type="button" value="filter"/>	



© Forschungszentrum Jülich [Imprint](#) [Privacy Policy](#) [Terms Of Use](#) [Support](#)



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

**Destination Earth**

<https://toar-data.org>  
Contact: [info@toar-data.org](mailto:info@toar-data.org)



# GUI – value map

TOAR:DATA  
BASE

Login

JÜLICH  
Forschungszentrum  
JÜLICH SUPERCOMPUTING  
CENTRE

general map
value map

o3 x

07/23/2020

14:00

confirm

color

- < 10
- 20 - 30
- 30 - 40
- 40 - 50
- 50 - 60
- 60 - 70
- 70 - 80
- 80 - 90
- 90 - 100
- > 100

© Forschungszentrum Jülich Imprint Privacy Policy Terms Of Use Support



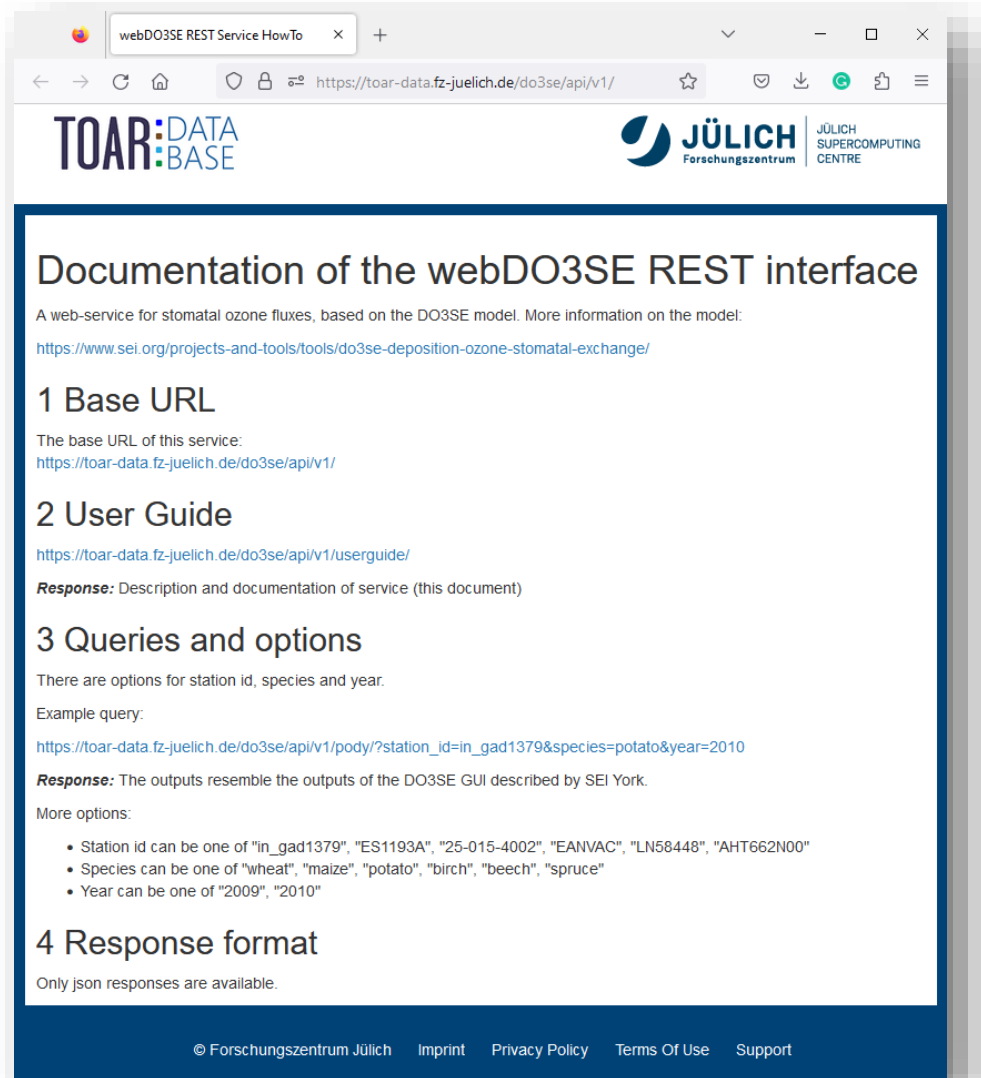
IntelliAQ is funded by the EU's ERC programme, Grant Agreement 78576.

**Destination Earth**

<https://toar-data.org>  
Contact: [info@toar-data.org](mailto:info@toar-data.org)

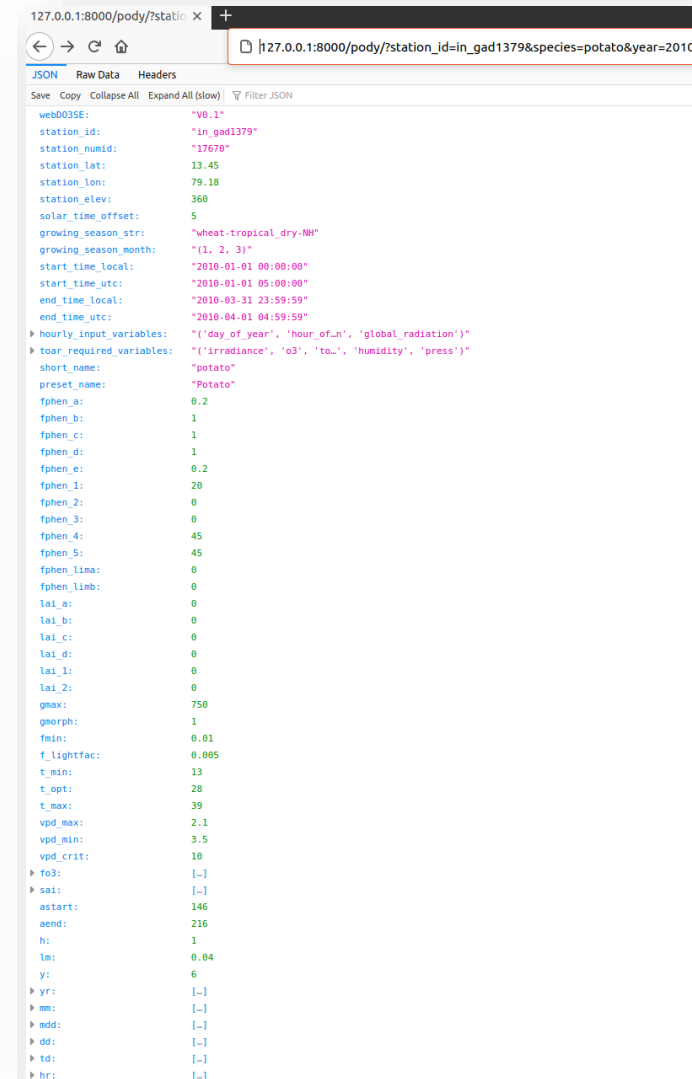
# webDO<sub>3</sub>SE

## REST API service for flux-based vegetation damage assessment



The screenshot shows the documentation page for the webDO3SE REST interface. The page title is "Documentation of the webDO3SE REST interface". It includes a sub-header "1 Base URL" with the base URL `https://toar-data.fz-juelich.de/do3se/api/v1/`. Under "2 User Guide", it lists a "Response" description and a link to the user guide. Under "3 Queries and options", it provides an example query: `https://toar-data.fz-juelich.de/do3se/api/v1/pody/?station_id=in_gad1379&species=potato&year=2010`. Under "4 Response format", it states that only JSON responses are available. The footer contains navigation links for "Forschungszentrum Jülich", "Imprint", "Privacy Policy", "Terms Of Use", and "Support".

from documentation



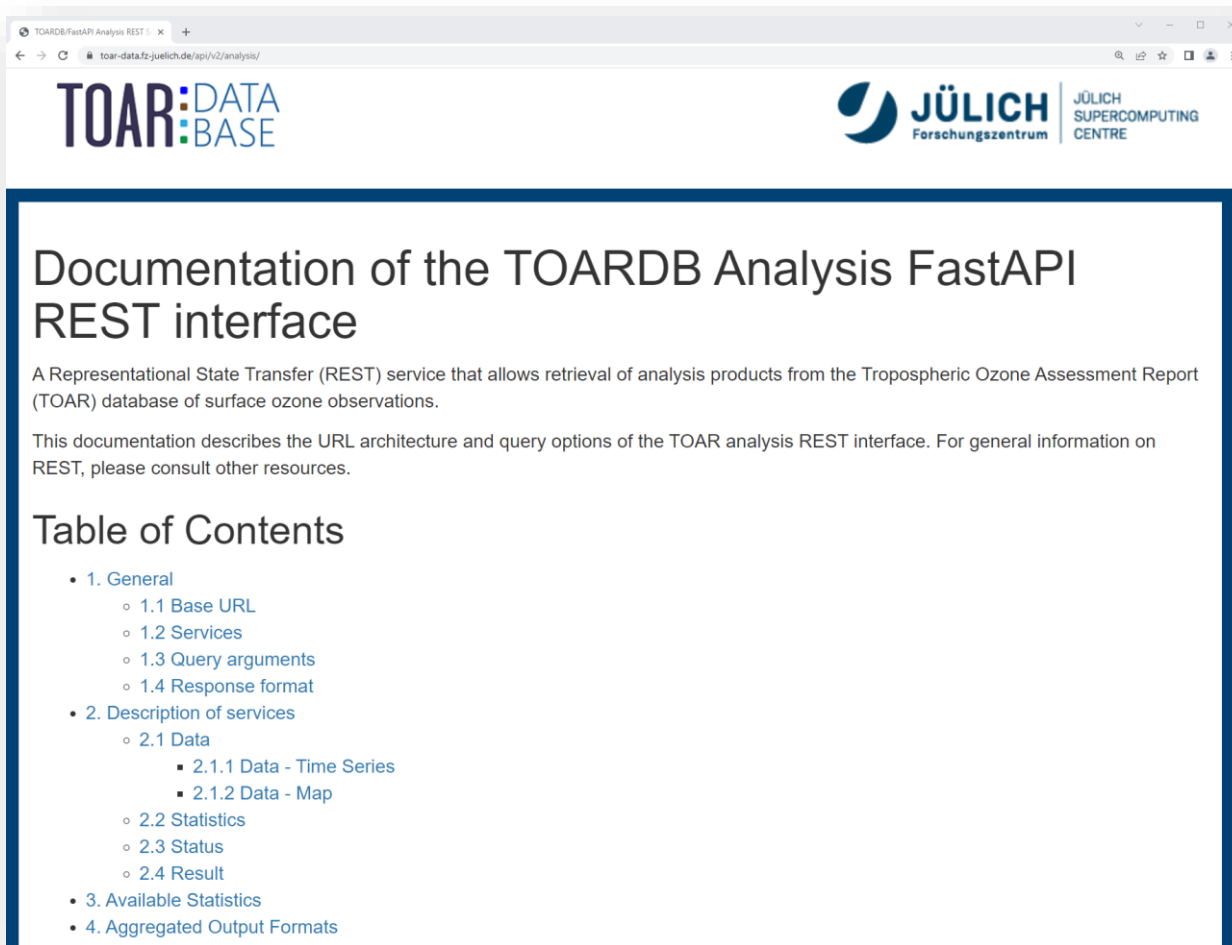
The screenshot shows a REST client interface displaying a JSON response for a request to `127.0.0.1:8000/pody/?station_id=in_gad1379&species=potato&year=2010`. The response is a detailed JSON object containing various parameters such as station information, growing season, and input variables.

```
webDO3SE: "V0.1"  
station_id: "in_gad1379"  
station_numid: "17670"  
station_lat: 13.45  
station_lon: 79.18  
station_elev: 360  
solar_time_offset: 5  
growing_season_str: "wheat-tropical-dry-NH"  
growing_season_month: "(1, 2, 3)"  
start_time_local: "2010-01-01 00:00:00"  
start_time_utc: "2010-01-01 05:00:00"  
end_time_local: "2010-03-31 23:59:59"  
end_time_utc: "2010-04-01 04:59:59"  
hourly_input_variables: ("day_of_year", "hour_of_n", "global_radiation")  
toar_required_variables: ("irradiance", "o3", "to", "humidity", "press")  
short_name: "potato"  
preset_name: "Potato"  
fphen_a: 0.2  
fphen_b: 1  
fphen_c: 1  
fphen_d: 1  
fphen_e: 0.2  
fphen_1: 20  
fphen_2: 0  
fphen_3: 0  
fphen_4: 45  
fphen_5: 45  
fphen_lima: 0  
fphen_limb: 0  
lai_a: 0  
lai_b: 0  
lai_c: 0  
lai_d: 0  
lai_1: 0  
lai_2: 0  
gmax: 750  
gmorph: 1  
fmin: 0.01  
f_lightfac: 0.005  
t_min: 13  
t_opt: 28  
t_max: 39  
vpd_max: 2.1  
vpd_min: 3.5  
vpd_crit: 10  
fo3: [-]  
sai: [-]  
astart: 146  
aend: 216  
h: 1  
lm: 0.04  
y: 6  
yr: [-]  
mm: [-]  
mdd: [-]  
dd: [-]  
td: [-]  
hr: [-]
```

to REST API request



# REST API analysis



TOAR:DATA  
BASE

JÜLICH  
Forschungszentrum

JÜLICH  
SUPERCOMPUTING  
CENTRE

## Documentation of the TOARDB Analysis FastAPI REST interface

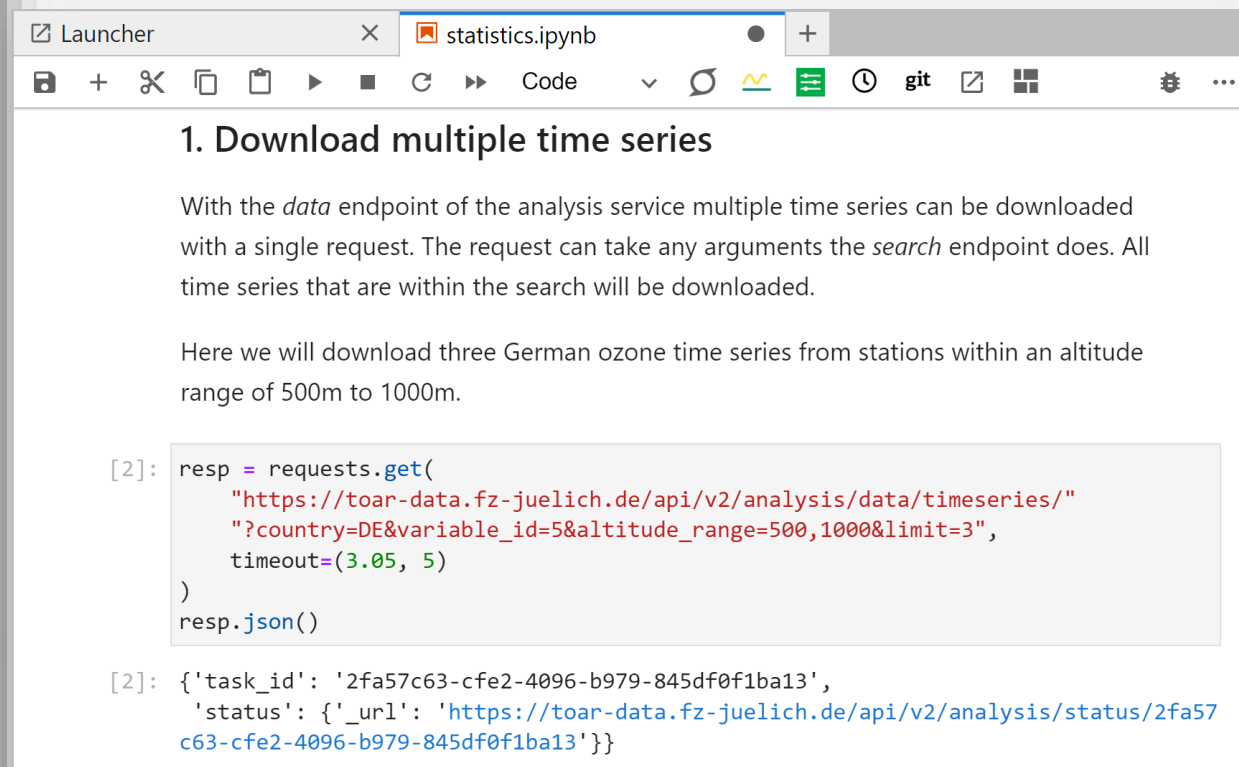
A Representational State Transfer (REST) service that allows retrieval of analysis products from the Tropospheric Ozone Assessment Report (TOAR) database of surface ozone observations.

This documentation describes the URL architecture and query options of the TOAR analysis REST interface. For general information on REST, please consult other resources.

### Table of Contents

- 1. General
  - 1.1 Base URL
  - 1.2 Services
  - 1.3 Query arguments
  - 1.4 Response format
- 2. Description of services
  - 2.1 Data
    - 2.1.1 Data - Time Series
    - 2.1.2 Data - Map
  - 2.2 Statistics
  - 2.3 Status
  - 2.4 Result
- 3. Available Statistics
- 4. Aggregated Output Formats

from documentation



Launcher statistics.ipynb

## 1. Download multiple time series

With the *data* endpoint of the analysis service multiple time series can be downloaded with a single request. The request can take any arguments the *search* endpoint does. All time series that are within the search will be downloaded.

Here we will download three German ozone time series from stations within an altitude range of 500m to 1000m.

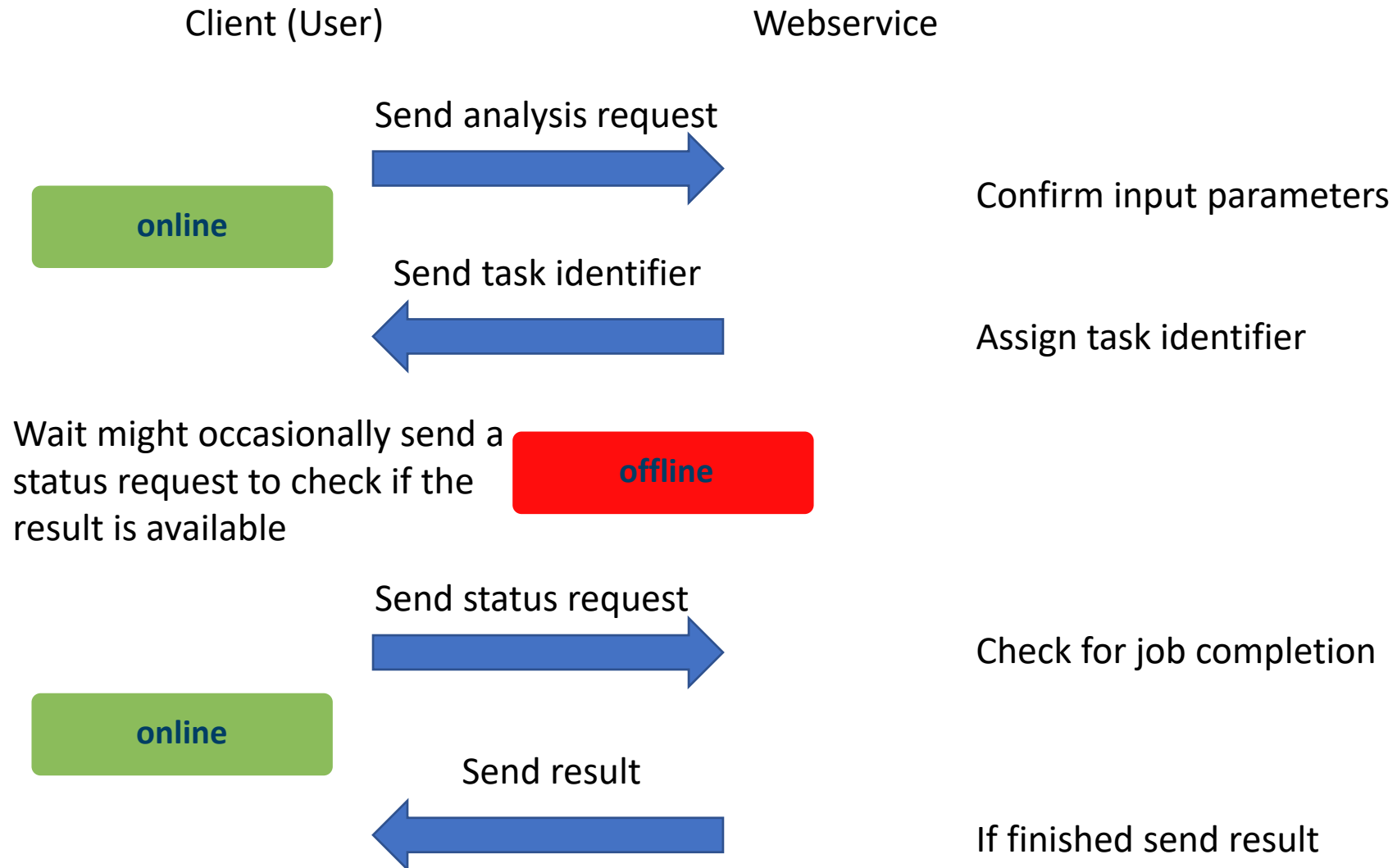
```
[2]: resp = requests.get(
      "https://toar-data.fz-juelich.de/api/v2/analysis/data/timeseries/"
      "?country=DE&variable_id=5&altitude_range=500,1000&limit=3",
      timeout=(3.05, 5)
    )
    resp.json()
```

```
[2]: {'task_id': '2fa57c63-cfe2-4096-b979-845df0f1ba13',
      'status': {'_url': 'https://toar-data.fz-juelich.de/api/v2/analysis/status/2fa57c63-cfe2-4096-b979-845df0f1ba13'}}
```

to Python notebook

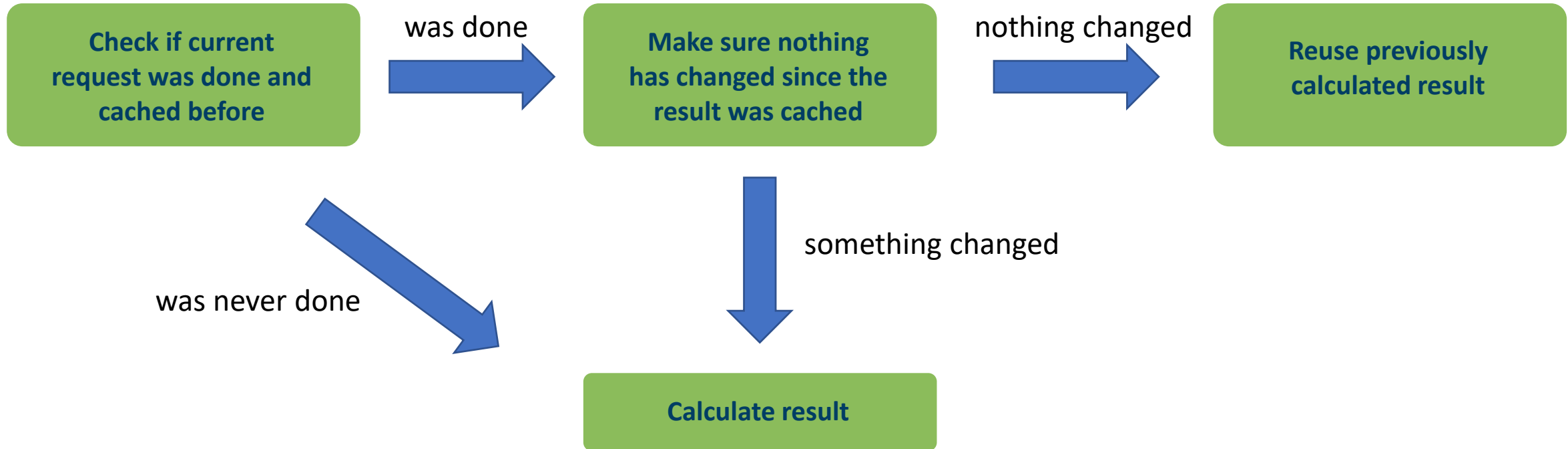


# Asynchronous Request-Reply Pattern



# Caching

To save time and energy we employ caching for the analysis products



# GEO PEAS GUI

## GEOspatial Point Extraction and Aggregation Service

Population Density

Modify:

Latitude and Longitude (lat,lon)

50.90899,6.41161

All Options

Submit

Latitude: 50.90899  
Longitude: 6.41161  
Radius: 10000.0  
Year: 2000  
Exact value at location: 295.7273720417102  
1/km<sup>2</sup>

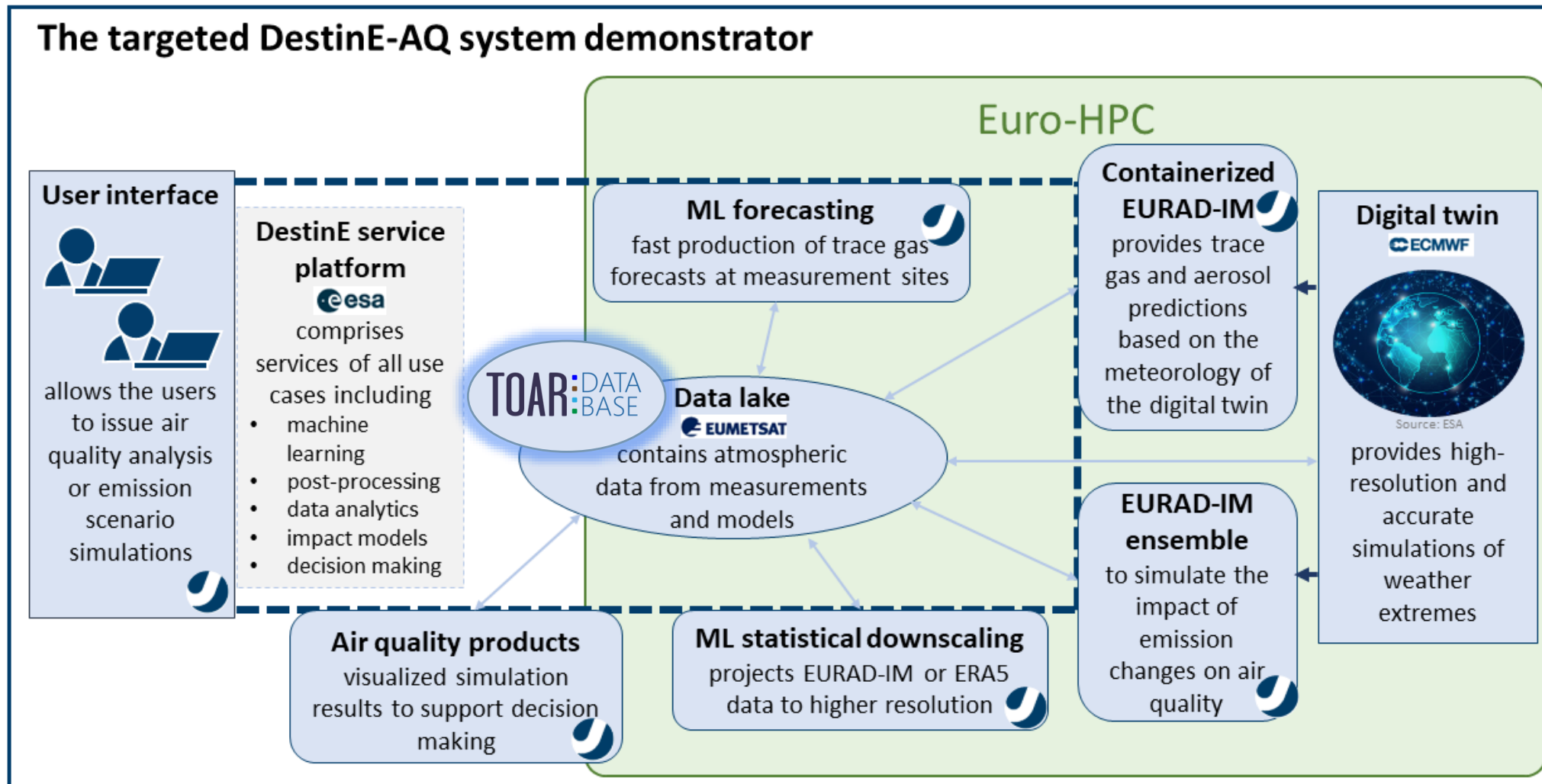
Data origin:

- Datasource: The European Commission, Joint Research Centre
- Copyright: The GHSL has been produced by the EC JRC as open and free data – Reuse is authorised, provided the source is acknowledged. The data are licensed under a CC-BY 4.0 license as documented at [https://ec.europa.eu/info/legal-notice\\_en](https://ec.europa.eu/info/legal-notice_en)
- Data Description: Residential population estimates for target years 1975, 1990, 2000 and 2015 provided by CIESIN GPWv4.10 were disaggregated from census or administrative units to grid cells, informed by the distribution and density of built-up as mapped in the Global Human Settlement Layer (GHSL) global layer per corresponding epoch. Detailed information can be found in [https://ghsl.jrc.ec.europa.eu/documents/GHSL\\_Data\\_Package\\_2019.pdf?t=1478q532234372](https://ghsl.jrc.ec.europa.eu/documents/GHSL_Data_Package_2019.pdf?t=1478q532234372)
- Data URL: [https://ghsl.jrc.ec.europa.eu/ghs\\_pop2019.php](https://ghsl.jrc.ec.europa.eu/ghs_pop2019.php)
- Citation: Schiavina, Marcello; Freire, Sergio; MacManus, Kytt (2019): GHS-POP R2019A - GHS population grid multitemporal (1975-1990-2000-2015). European Commission, Joint Research Centre (JRC) [Dataset] doi:10.2905/0C6B9751-A71F-4062-830B-43C9F432370F PID: <http://data.europa.eu/89h/0c6b9751-a71f-4062-830b-43c9f432370f>



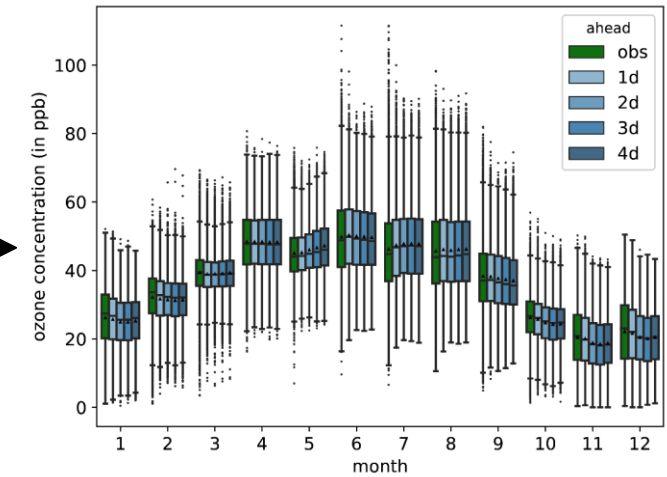
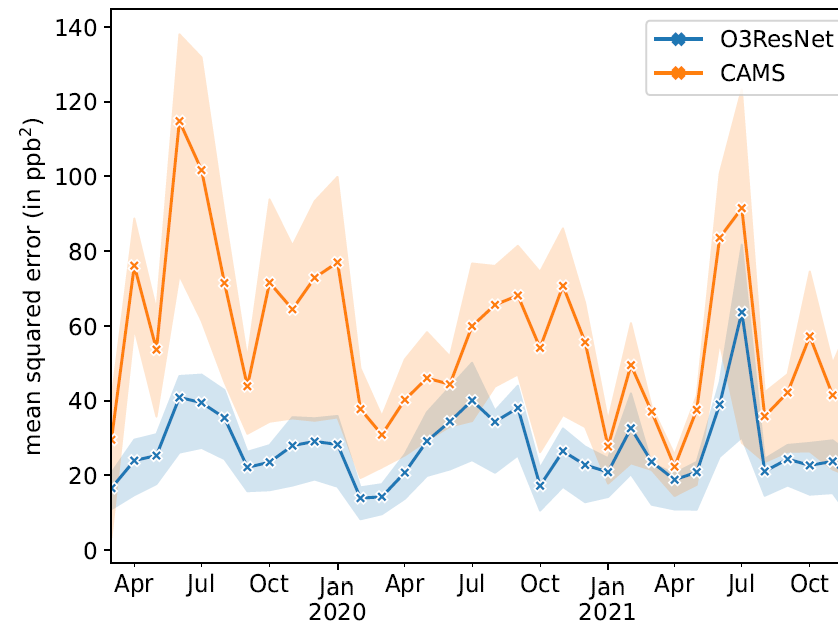
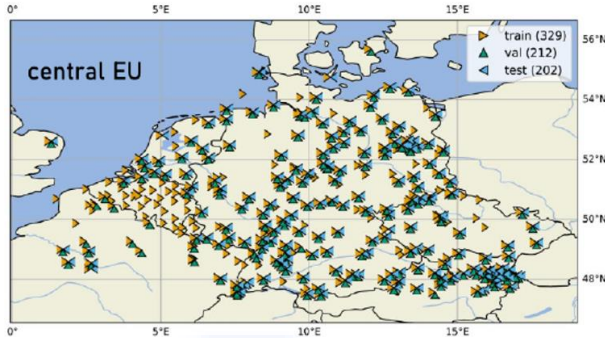
# Use of TOAR data in ML applications

## Destination Earth Air Quality Use Case DE370C



# Use of TOAR data in ML applications

## IntelliAQ ML Time Series Forecasting



ML model to forecast MDA8 ozone at measurement sites

Superior to CAMS regional model ensemble.

(Leufen, L. H., Kleinert, F., and Schultz, M. G.: O3ResNet: A deep learning based forecast system to predict local ground-level daily maximum 8-hour average ozone, submitted to AI4ES, December 2022)





# Attributions

- © Author(s) 2023. This work is distributed under the Creative Commons Attribution 4.0 License.
- Workflow icon created by ghost\_icon from www.flaticon.com: <https://www.flaticon.com/free-icons/workflow>
- Fliessband icon created by smahingstocks from www.flaticon.com: <https://www.flaticon.com/de/kostenlose-icons/fliessband>
- Files and folders icons created by kerismaker from www.flaticon.com: <https://www.flaticon.com/free-icons/files-and-folders>
- Box icons created by kmg design from www.flaticon.com: <https://www.flaticon.com/free-icons/box>
- Checklist icons created by HideMaru from www.flaticon.com: <https://www.flaticon.com/free-icons/checklist>
- Search icons | Deploy icons created by Smashicons from www.flaticon.com: [https://www.flaticon.com/free-icons/\[search|deploy\]](https://www.flaticon.com/free-icons/[search|deploy])
- Graph icons | User icons | Gears icons | Pointing icons | Door icons | Plug icons | Blog icons created by Freepik from www.flaticon.com: [https://www.flaticon.com/free-icons/\[graph|user|gears|pointing|door|plug|blog\]](https://www.flaticon.com/free-icons/[graph|user|gears|pointing|door|plug|blog])
- Space station icons created by IYIKON from www.flaticon.com: <https://www.flaticon.com/free-icons/space-station>
- Magnifying glass icons created by Creative Stall Premium from www.flaticon.com: <https://www.flaticon.com/free-icons/magnifying-glass>
- Recycle icons created by srip from www.flaticon.com: <https://www.flaticon.com/free-icons/recycle>
- Database icons created by Bartama Graphic from www.flaticon.com: <https://www.flaticon.com/free-icons/database>
- Three arrows icons created by alkhalfi design from www.flaticon.com: <https://www.flaticon.com/free-icons/three-arrows>

