



TOAR Data FAQ

Release 1.0

The TOAR Data Infrastructure Team

Jul 04, 2022

CONTENTS:

1	Frequently asked questions about general aspects	1
1.1	What is the difference between the TOAR data portal and the TOAR database?	1
1.2	How can I publish the data I have used for a journal publication?	1
1.3	Under which licence will the data be published?	1
1.4	Can I make sure that the data is not published before the journal paper has been published?	1
1.5	What is the scope of the data collection in TOAR-II, both with respect to the portal and the database?	2
1.6	What are the design principles for a good REST API? Can you support us in developing a REST API?	2
2	Frequently asked questions about the TOAR Data Portal	3
2.1	What functions does the TOAR Data Portal offer?	3
2.2	Which data sources are linked from the TOAR Data Portal?	3
2.3	How come my data are not listed in the TOAR data portal?	4
2.4	What is the link to the registration of datasets for including them in the TOAR data portal?	4
2.5	How can ship-based measurements be contributed to TOAR?	4
2.6	Concerning aircraft field campaigns: will there be one entry in the data portal for each field campaign? They reside on separate web pages.	4
2.7	I have datasets which include several species. Do I have to reformat those so that they can be included in the TOAR data portal?	4
2.8	Do you have a preferred format and metadata profile for data that is listed in the data portal?	5
2.9	How do we inform you about an update to an existing entry in the TOAR data portal?	5
2.10	What is hidden behind the links to REST API in the TOAR data portal entries?	5
2.11	Are you actively monitoring all the pointers in the TOAR data portal in case there are changes on the web pages or services hosting the original datasets?	5
2.12	Since we are considering precursor information in TOAR-II, will we have links for that?	5
3	Frequently asked questions about the TOAR Database (user)	7
3.1	Which data are available in TOAR DB?	7
3.2	Where does the data come from?	7
3.3	Which metadata are available in TOAR DB?	7
3.4	Where do I find the metadata catalogue?	8
3.5	How do I retrieve data?	8
3.6	What can I search for?	8
3.7	What is the result of a search?	8
3.8	How to interpret the retrieved data?	8
3.9	What is the resolution of the data?	8
3.10	How long will TOAR data remain available?	9
3.11	Is the data historic or near-real time?	9
3.12	Has the data been curated?	9
3.13	What is the quality of the data?	9
3.14	How has the data been processed?	9

3.15	Can I search for a certain quality level?	10
3.16	Under which conditions may I access the data?	10
3.17	Under which conditions may I use the data?	10
3.18	How do I get access to TOAR DB?	10
3.19	How is the data delivered, instantly or later on as a batch?	10
3.20	Which interfaces are available?	10
3.21	How can I cite the TOAR DB?	11
3.22	What about acknowledging the original data provider?	11
3.23	Are the FAIR principles applied to the TOAR data infrastructure?	11
3.24	Is there a GUI available for accessing the TOAR database?	11
3.25	Where can I get further support?	11
4	Frequently asked questions about the TOAR Database (data provider)	13
4.1	How flexible are you with respect to the time format? Your recommendation is not exactly ISO and doesn't include seconds.	13
4.2	Do you accept data in resolutions other than hourly?	13
4.3	If we already submitted data to the GAW data centre. Do we have to resubmit this to TOAR?	13
4.4	What happens if we submit a corrected version of our data to the GAW data centre after you harvested the data from there. Do we have to inform you about this?	13
4.5	Is there a place to upload measurement intercomparison data?	14
4.6	Some data in TOAR-v1 did not comply with the CC-BY 4 license. Will data providers be contacted to obtain a new agreement?	14

FREQUENTLY ASKED QUESTIONS ABOUT GENERAL ASPECTS

1.1 What is the difference between the TOAR data portal and the TOAR database?

The portal (<https://toar-data.org>) only provides pointers to datasets, primarily those, which we cannot include in the TOAR database of surface observations. For example, satellite, aircraft, model, or ozone sonde and lidar data. The TOAR database (<https://toar-data.fz-juelich.de>) contains as many time series of stationary surface observations as we can get and process.

1.2 How can I publish the data I have used for a journal publication?

In case the journal publication is contributing to the Tropospheric Ozone Assessment Report the related datasets can be published as part of the TOAR community. Get in contact with the TOAR data centre team, who will work with the authors to create the appropriate metadata record(s) and deposit the data with EUDAT's B2SHARE service (<https://b2share.fz-juelich.de/communities/TOAR>).

1.3 Under which licence will the data be published?

CC-BY 4.0

1.4 Can I make sure that the data is not published before the journal paper has been published?

It is possible to put an embargo on the data until the journal paper is out but no longer than 12 months.

1.5 What is the scope of the data collection in TOAR-II, both with respect to the portal and the database?

We aim to provide as much information about TOAR-related data as possible. Specifically, we really want to have links to all datasets referenced in any of the TOAR-II papers included, or, if these are ground-based time series, the data to be included on the TOAR database. While the portal focuses on ozone data only, the database also includes precursor and meteorological observations.

1.6 What are the design principles for a good REST API? Can you support us in developing a REST API?

We are certainly willing to share our experiences with you, but we cannot provide much support for API development. Please write to support@toar-data.org if you want to get in touch with us.

FREQUENTLY ASKED QUESTIONS ABOUT THE TOAR DATA PORTAL

The TOAR Data Portal is accessible at <https://toar-data.org>. It is the home of tropospheric ozone data.

2.1 What functions does the TOAR Data Portal offer?

toar-data.org provides the links to several sources of satellite, surface, aircraft, and ozone sonde & LiDAR data. In addition, it gives information on how to contribute data to the TOAR data base and on how to contribute to TOAR science.

2.2 Which data sources are linked from the TOAR Data Portal?

Currently the following sources are linked (click on the links below to see its contents):

- [Satellite data](#)¹
 - [Satellite Ozone Data](#)²
 - [Satellite Ozone Precursor Data \(soon\)](#)³
- [Surface data](#)⁴
- [Aircraft data](#)⁵
- [Ozonesonde & LiDAR data](#)⁶

¹ <https://toar-data.org/satellite-data/>

² <https://toar-data.org/satellite-data/>

³ <https://toar-data.org/satellite-data/>

⁴ <https://toar-data.org/surface-data/>

⁵ <https://toar-data.org/aircraft-data/>

⁶ <https://toar-data.org/ozone-sonde-lidar-data/>

2.3 How come my data are not listed in the TOAR data portal?

The TOAR data portal is maintained as a voluntary effort on a best effort basis. We rely on information from the scientific community to identify and describe the tropospheric ozone datasets which you can see at <https://toar-data.org>. If you spot a dataset which is not listed there, please drop us that information via https://toar-data.org/contribute/#toar_data_portal_contribute, and we will see if we can add it to the portal. Please note that we can only link to datasets if it is clear that the repository, which is linked, has the necessary permissions to provide the data.

2.4 What is the link to the registration of datasets for including them in the TOAR data portal?

https://toar-data.org/contribute/#toar_data_portal_contribute

2.5 How can ship-based measurements be contributed to TOAR?

Unfortunately, the database model of the TOAR database only allows for stationary observations. Our suggestion for ship-based data (generally: moving platform data) is to establish a collection of files in a homogenous format and publish those on the TOAR data publication server⁷. Please fill out the form https://toar-data.org/contribute/#toar_data_portal_other_contribute. We will then work on this and will generate a doi for citations. A pointer to this collection can then be added to the TOAR data portal⁸.

2.6 Concerning aircraft field campaigns: will there be one entry in the data portal for each field campaign? They reside on separate web pages.

The TOAR data portal allows for different levels of granularity and in the future we plan to make this granularity visible. It is up to the providers to define how they want data listed, be it as a collection or as separate entries.

2.7 I have datasets which include several species. Do I have to reformat those so that they can be included in the TOAR data portal?

No. As long as the dataset contains (also) ozone data, it is eligible to be included in the portal.

⁷ <https://b2share.fz-juelich.de/communities/TOAR>

⁸ <https://toar-data.org>

2.8 Do you have a preferred format and metadata profile for data that is listed in the data portal?

No. As we don't process your data, any format is accepted. It would be great if the TOAR-II working groups can contribute to the harmonisation of data formats and metadata, but this is out of scope for our data gathering efforts.

2.9 How do we inform you about an update to an existing entry in the TOAR data portal?

You can either write an email to support@toar-data.org or fill in the registration form https://toar-data.org/contribute/#toar_data_portal_contribute again and make a comment that this is a revision and not a new entry.

2.10 What is hidden behind the links to REST API in the TOAR data portal entries?

A REST API is a web service, where you can format the URL with question marks and ampersands and issue commands to a web page. These will then be processed and the result returned to you. The pointer we provide in the data portal will usually direct you to a description page how the specific REST API of that dataset can be used.

2.11 Are you actively monitoring all the pointers in the TOAR data portal in case there are changes on the web pages or services hosting the original datasets?

No, this would be too much work for us. We will occasionally check the links, but we have to rely on you, the community, to provide us with updates or point us to broken links or other errors if you discover them.

2.12 Since we are considering precursor information in TOAR-II, will we have links for that?

Given the limited resources that we have for maintaining the data portal, we would prefer to focus on ozone datasets at this point. Adding other species may make the portal less clear and less complete for any given species. But we already added a link for ozone precursor data from satellites, which will be filled soon by the community.

FREQUENTLY ASKED QUESTIONS ABOUT THE TOAR DATABASE (USER)

The landing page of the TOAR Database Infrastructure is accessible at <https://toar-data.fz-juelich.de>.

3.1 Which data are available in TOAR DB?

The TOAR database primarily collects ground-level ozone observations from around the world. In TOAR-II there are also increased efforts to include ozone precursor observations and meteorological data to allow for trend attributions and other scientific analyses. All data in the TOAR database are harmonized and quality controlled, but the level of quality control differs (see questions on *data curation* and *data quality*). Only data from research-grade instruments is accepted. The current status of data in the TOAR database can be queried with the following URL: https://toar-data.fz-juelich.de/api/v2/database_statistics/.

3.2 Where does the data come from?

TOAR-II collects data from 18 large air quality monitoring networks and from many individual data providers. These datasets are augmented with extra metadata which is obtained through processing of various earth observation datasets (see chapter 4.2.4 (Station characterisation through geospatial data) of guide https://toar-data-dev.fz-juelich.de/documentation/TOAR_UG_Vol03_Database.pdf). Furthermore, meteorological information is added to the air pollutant time series through extraction from gridded ECMWF reanalysis version 5 (ERA5). Detailed information on the data sources can be found in chapter 2 (TOAR Data Sources) of guide https://toar-data-dev.fz-juelich.de/documentation/TOAR_TG_Vol02_Data_Processing.pdf.

3.3 Which metadata are available in TOAR DB?

The TOAR database version 2 has an extensive metadata schema which includes a lot of information about the measurement location (station) and the measurement itself (timeseries). In addition, we generate globally uniform metadata through a set of queries to our geospatial data set point extraction and aggregation service (GEO-PEAS). A detailed description of all metadata can be found in https://esde.pages.jsc.fz-juelich.de/toar-data/toar_db_fastapi/docs/toar_db_fastapi.html#stationmetaglobal.

3.4 Where do I find the metadata catalogue?

The API for this is under development. Currently all metadata is listed at https://esde.pages.jsc.fz-juelich.de/toar-data/toardb_fastapi/docs/toardb_fastapi.html.

You can retrieve the ontology of TOAR data as xml from <https://toar-data-dev.fz-juelich.de/api/v2/ontology> or you can have see it online as OWL document via <https://toar-data-dev.fz-juelich.de/documentation/ontologies/v1.0/>.

3.5 How do I retrieve data?

Currently, the only way to obtain data from the TOAR database is to use the REST API <https://toar-data.fz-juelich.de/api/v2>. This web page contains the documentation of the REST API. We are developing a graphical user interface and hope to make a first version of this available by October 2022. Please note that the REST API is being developed further so that more powerful and comfortable queries should become available over time.

3.6 What can I search for?

Currently, you can list data such as stationmeta, timeseries, or variables via the API.

You can search for all attributes being equal to a given value.

Example:

<https://toar-data.fz-juelich.de/api/v2/stationmeta/?country=DE,NL>
finds all stations in Germany and The Netherlands.

3.7 What is the result of a search?

Any query to the TOAR database REST API returns a JSON structure.

Data queries can also be returned as csv file.

3.8 How to interpret the retrieved data?

Retrieved data can easily be processed with a Python script. Since JSON is a standardized human readable format, you can of course interpret the data with any tool of your choice.

csv files can be processed with your favourite tool.

3.9 What is the resolution of the data?

The TOAR database stores timeseries data in hourly resolution or finer. Most data have been collected as hourly data, but the database can also work with half-hourly, 15-minute or 10-minute data.

3.10 How long will TOAR data remain available?

The data is stored permanently and will be available for at least 10 years.

3.11 Is the data historic or near-real time?

The focus of TOAR is on the analysis of long-term trends and the TOAR database therefore primarily contains “historic” data. We are beginning to implement near realtime data streams for at least some of the major air quality data providers, but this has low priority at present.

3.12 Has the data been curated?

Yes. Data curation consists of a first manual inspection to check if the data can be processed at all and adheres to our metadata definitions and data format description (see https://toar-data.fz-juelich.de/documentation/TOAR_TG_Vol02_Data_Processing.pdf for details). The following automated processing workflow contains several checks and quality tests including statistical tests to detect large discrepancies to expected values and frequency distributions. Once a dataset passes these tests it will be inserted into the database where providers and the TOAR data centre team can visualize and inspect the data again. Other curation steps include the harmonization of metadata information and the augmentation of metadata through processed earth observation information (see *question on metadata*).

3.13 What is the quality of the data?

TOAR only accepts data from research grade instruments and relies on quality control exerted by the data provider (monitoring agency, scientific institution or other). Nevertheless, data processing errors and other factors can lead to errors in the data that is stored in the database. We try to identify such errors through an automated quality control tool and, in some cases, through manual inspection. Furthermore, preliminary analyses of TOAR data for the scientific papers produced in TOAR-II will identify data errors and we implemented a feedback function so that users can alert us to obvious or likely data errors. It is impossible to guarantee the correctness of all TOAR data, but we take data quality serious and do our best to achieve the maximum possible quality of the data in the TOAR database. Since the focus of TOAR is on tropospheric ozone, the data quality of ozone is likely better than that of ozone precursor species. The quality of the meteorological data from ERA5 can be assessed through the ERA5 validation report (<https://confluence.ecmwf.int/display/CKB/ERA5>).

3.14 How has the data been processed?

This process is detailed in https://toar-data.fz-juelich.de/documentation/TOAR_TG_Vol02_Data_Processing.pdf

3.15 Can I search for a certain quality level?

Yes. The data flagging scheme of the TOAR database version 2 allows to distinguish between quality flags set by the data provider and data quality flags assigned from our automated quality control tool or visual inspection of the TOAR data centre team. Details on how to specify the desired data quality level can be found in https://toar-data.fz-juelich.de/documentation/TOAR_UG_Vol03_Database.pdf, section 5.

3.16 Under which conditions may I access the data?

Access to TOAR-II data is open and free. Please see *question on data use* for the data use conditions.

3.17 Under which conditions may I use the data?

All TOAR-II data are provided without restrictions under a CC-BY 4.0 license. This license requires that you acknowledge the data source. To facilitate this, each response to a TOAR data query via the REST API contains a citation and acknowledgement metadata element. It is your responsibility as data user to make sure that proper acknowledgements are given. See also data use policy.

3.18 How do I get access to TOAR DB?

Currently, the TOAR database version 2 can only be accessed via the REST API at <https://toar-data.fz-juelich.de/api/v2>. There is no registration or login required. For a documentation on the REST API, see https://toar-data.fz-juelich.de/documentation/TOAR_UG_Vol03_Database.pdf

3.19 How is the data delivered, instantly or later on as a batch?

Until now, all processing of TOAR data happens instantly and the results will be returned immediately (more complex queries may take several seconds). As we add more sophisticated analyses options, we may have to implement a data warehouse scheme where your query will be stored and queued and you will be informed by email when the results are available.

3.20 Which interfaces are available?

Currently, the TOAR database version 2 only provides a Representational State Transfer (REST) Application Programming Interface (API). Detail on the interfaces are provided at <https://toar-data.fz-juelich.de/api/v2> and https://toar-data.fz-juelich.de/documentation/TOAR_UG_Vol03_Database.pdf. A graphical user interface is in preparation.

3.21 How can I cite the TOAR DB?

The TOAR database should be cited as Schröder et al; TOAR Data Infrastructure; <https://doi.org/10.34730/4d9a287dec0b42f1aa6d244de8f19eb3>

For individual data series and a small set of data series the original data sources should be cited. A recommended citation is provided with the metadata when data are downloaded.

3.22 What about acknowledging the original data provider?

Each response to a TOAR data query via the REST API contains a citation and acknowledgement metadata element. It is your responsibility as data user to make sure that proper acknowledgements are given.

3.23 Are the FAIR principles applied to the TOAR data infrastructure?

Yes. We are proud to have built one of the FAIRest data infrastructures for atmospheric data in the world.

3.24 Is there a GUI available for accessing the TOAR database?

A GUI is under development. A first version is planned to be available in October 2022.

3.25 Where can I get further support?

For further questions please send an email to support@toar-data.org.

FREQUENTLY ASKED QUESTIONS ABOUT THE TOAR DATABASE (DATA PROVIDER)

4.1 How flexible are you with respect to the time format? Your recommendation is not exactly ISO and doesn't include seconds.

The closer you come to our recommended format, the easier it is for us to insert your data. We can work with the ISO format (including the „T“ between date and time) and you can include seconds in the format. However, seconds will be ignored.

4.2 Do you accept data in resolutions other than hourly?

Hourly is the minimum resolution your data should have. We also accept data in half-hour, 15 minutes, or 10 minutes resolution. They will be automatically aggregated when querying the database. If you have data with a time resolution not listed here, please contact us at toar-data@fz-juelich.de

4.3 If we already submitted data to the GAW data centre. Do we have to resubmit this to TOAR?

GAW data will be harvested from EBAS (provided that license issues can be sorted out quickly).

4.4 What happens if we submit a corrected version of our data to the GAW data centre after you harvested the data from there. Do we have to inform you about this?

Yes, please! It would be great to have automatic tracking of changes, but we are not there yet.

4.5 Is there a place to upload measurement intercomparison data?

Such data can be posted on the TOAR data publication server (<https://b2share.fz-juelich.de/communities/TOAR>). However, a TOAR WG should try to harmonize such information before sending the data to us.

4.6 Some data in TOAR-v1 did not comply with the CC-BY 4 license. Will data providers be contacted to obtain a new agreement?

Yes. The TOAR-II database is a separate new database which is filled from scratch. We prefer to re-insert original data as there may have been changes applied in the meantime and our new metadata is more extensive. But if need be, „old“ data can be copied, if the data provider agrees.