TOAR Data User Guide #4

TOAR Data FAQ

toar-data.fz-juelich.de

Version 2.0.0 | September 11, 2025

The TOAR Data Team

Forschungszentrum Jülich GmbH
ESDE | JSC - FSD
52425 Jülich

CONTENTS:

LIST OF FIGURES

LIST OF TABLES

CHAPTER

ONE

GENERAL

1.1 What data services are offered by the TOAR data infrastructure at the Jülich Supercomputing Centre?

The [TOAR data centre team](https://toar-data.org/about-toar-data/#about_our_team) develops and operates

- the [TOAR data portal](https://toar-data.org) - a one stop shop to locate and access tropospheric ozone data from a large variety of measurement platforms and models,
- the [TOAR database](https://esde.pages.jsc.fz-juelich.de/toar-data/toardb_fastapi/docs/toardb_fastapi.html) of global surface observations - one of the world's largest collections of ozone measurements and related precursors and meteorological variables from the 1970s until recently
- the [TOAR data services](https://toar-data.fz-juelich.de/subpages/data_services.html) - a collection of graphical user iterfaces and REST APIs to access, download, analyze, and visualize TOAR data
- a [software repository](https://toar-data.fz-juelich.de/subpages/software.html) featuring easy-to-use Python code to integrate TOAR data in your workflows
- the [TOAR data publication service](https://b2share.fzjuelich.de/communities/TOAR) that allows data providers to publish their datasets including metadata documentation with a citeable doi.

Check out our [home page](https://toar-data.fz-juelich.de) to access all facilities of the TOAR data infrastructure.

1.2 What data are stored in the TOAR data infrastructure at JSC?

The [TOAR database](https://esde.pages.jsc.fz-juelich.de/toar-data/toardb_fastapi/docs/toardb_fastapi.html) only contains global surface station measurements. The [TOAR assessment report](https://igacproject.org/activities/TOAR/TOAR-II) also includes many other data from satellites, vertical sounding, mobile platforms, and models. These are stored at other locations. The [TOAR data portal](https://toar-data.org) provides pointers to all datasets that are used in the assessment.

1.3 Who is working on the TOAR data infrastructure?

Our dedicated [team](https://toar-data.org/about-toar-data/#about_our_team) consists of experienced scientists and engineers as well as young researchers and students full of ideas. We are continuously exploring ways to make the TOAR data infrastructure more user-friendly and efficient. Please [reach out to us](mailto:toar-data@fz-juelich.de[1]) if you would like to get involved in testing new concepts and services or help develop them. All TOAR data related [software](https://gitlab.jsc.fz-juelich.de/esde/toar-data/) is open source.

1.4 How is the TOAR data infrastructure funded?

The [Jülich Supercomputing Centre](https://www.fz-juelich.de/en/ias/jsc) is committed to operate and develop the TOAR data infrastructure as a relevant real-life use case for modern [FAIR](https://www.go-fair.org/) data management and as a platform to develop and test new concepts for web services. Base funding for the TOAR data infrastructure comes from the Helmholtz [research field _Information_](https://www.helmholtz.de/en/about-us/structure-and-governance/program-oriented-funding/scientific-evaluation/research-field-information/). Currently, there is no 3rd party funding available to support TOAR data infrastructure development or data curation.

1.5 What is the TOAR data team doing (and not doing)?

Due to limited funding and person power (see above), the [TOAR data centre team](https://toar-data.org/about-toar-data/#about_our_team) must fully concentrate on all technical aspects of the TOAR data infrastructure, including the processing of new data submissions, the development and operation of dat aservices, and the curation of the documentation and web pages. __We have only very limited capacity to engage in data quality **control**__. This task is under responsibility of the different TOAR assessment paper writing teams.

1.6 What are the use conditions of TOAR data?

All data stored in the [TOAR database](https://esde.pages.jsc.fz-juelich.de/toar-data/toardb_fastapi/docs/toardb_fastapi.html) are licensed with a [Creative Commons CC-BY4.0](https://creativecommons.org/licenses/by/4.0/) deed. This means you are free to download, analyze, reformat, re-use, and distribute all data __provided that you properly acknowledge the data providers and the TOAR data **infrastructure**__. We have developed a specific data service (***LINK***) that assists you in finding the right information in your data citations.

[1] toar-data@fz-juelich.de

1.7 What is the easiest way to access TOAR data?

That depends on the task you want to solve and your data analysis skills. For starters, our [TOAR dashboard](https://toar-data.fz-juelich.de/gui/v2/dashboard/) is clearly the way to go. It offers a powerful and comfortable graphical user interface with rich filtering and visualisation options. To work with larger sets of TOAR data, we recommend to explore the [TOAR REST API services](https://toar-data.fz-juelich.de/subpages/data_services.html), and in particular the [statistics](https://toar-data.fz-juelich.de/api/v2/analysis/#statistics) and [trend analysis](https://toar-data.fz-juelich.de/api/v2/analysis/#trends) service. If you want to include these services in your own programs, check out our [software repository](https://gitlab.jsc.fz-juelich.de/esde/toar-public) with example scripts and Jupyter notebooks.

1.8 How much data can I obtain from the TOAR data infrastructure?

In principle, it is possible to obtain all the data that are stored in the TOAR database. However, to allow a fair and performant access to our services by all users, we are imposing some constraints on our services. As anonymous user, you can browse and view all data, but downloads and analyses are limited. If you become a [registered user](***LINK TO REGISTRATION***) your limits will be increased so that you can perform meaningful analyses on a continental scale. For special cases, we can provide power user privileges for quasi unlimited access. Please get in comntact with us if you need this.

1.9 Am I allowed to copy all TOAR data and build my own database from it?

Technically yes, because of the [CC-BY4.0 license](https://creativecommons.org/licenses/by/4.0/). However, if you plan to do this, we ask you to get in [contact](mailto:toar-data@fz-juelich.de[2]) with us so that we can arrange a meaningful collaboration. We are committed to providing a single access point to ozone surface data that are used in the TOAR assessment, and the construction of database clones will easily lead to inconsistent data versions. We are happy to work with you and distribute work and responsibilities, so let's talk! Should you make a clone of the TOAR database without our explicit consent, please note that you are not allowed to call it "TOAR database" or anything else that woul dgive the impression that your database has been approved by the [TOAR steering committee](https://igacproject.org/activities/TOAR/TOAR-II).

1.10 How can I engage with the TOAR initiative?

Please check out the main [TOAR initiative web page](https://igacproject.org/activities/TOAR/TOAR-II) and [register for the TOAR email list](https://igacproject.us19.list-manage.com/subscribe?u=aeb9ed0dde2c98d3a19eef3b7&id=537c6945c1). The TOAR co-chairs will update you about new developments every 3 months or so. Feel free to reach out to the coordinating lead authors of any TOAR assessment paper if you would like to engage in the scientific analysis of TOAR data and the paper writing.

[2] toar-data@fz-juelich.de

CHAPTER

TWO

TOAR DATA PORTAL

2.1 What is the TOAR data portal?

The TOAR Data Portal is accessible at https://toar-data.org. It's intention is to provide a comprehensive overview about available tropospheric ozone datasets and in particular all datasets that are used in the second TOAR assessment report.

2.2 What data are available in the TOAR Data Portal?

The TOAR data portal[3] provides descriptions and pointers to the following types if data:

- Surface data[4]
- Satellite data[5]
- Aircraft data[6]
- Ozonesonde & LiDAR data[7]
- Model data[8]

The portal does not contain copies of any datasets but points you to the web pages and interfaces of the individual data providers. While we try to maintain the portal in an up-to-date state, we depend on the information that is given to us.

2.3 Why should I use the TOAR Data Portal?

The TOAR data portal[9] provides succinct and standardized information about many tropospheric ozone data sets so that you can decide which data is right for you. It highlights the availability of modern REST APIs, data license conditions, and documentation and thus exposes the level of FAIRness[10] of the different data collections. We see the TOAR data portal[11] as a great starting point for your research.

[3] https://toar-data.org
[4] https://toar-data.org/surface-data/
[5] https://toar-data.org/satellite-data/
[6] https://toar-data.org/aircraft-data/
[7] https://toar-data.org/ozone-sonde-lidar-data/
[8] https://toar-data.org/model-data/
[9] https://toar-data.org
[10] https://www.go-fair.org/
[11] https://toar-data.org

2.4 How can I register my data on the TOAR data portal?

Simply fill the online form[12] or send us an email[13] with the required information. We will be happy to help you structure the information as needed. Please note that we can only list datasets if the repository that is linked has the necessary permissions to provide the data.

2.5 How can I report errors in the TOAR data portal?

Just follow the procedures for registering a new entry and indicate that you want to report or correct an error. Note that you don't have to be the repository owner to send us this information. Anyone can contribute as long as the data resource that shall be listed is properly licensed.

2.6 How can I access data through the TOAR data portal?

This depends on the individual repository. Since the TOAR data portal[14] does not hold copies of the datasets listed, we have no control over how data providers make their data available. You need to consult the individual web pages to find out how you can access the data, how it is formatted, and how it can be used. Our standardized links to REST-API, contact, and license can help you find this information faster.

2.7 What does the REST-API symbol in a repository entry stand for?

A REST API is a web service, where you can issue data queries by sending a URL that contains question marks and ampersands in a specified way. These URLs are interpreted as commands by the web service and it will return a query-specific response or dataset to you. If the repository that is listed has such a REST API service, this icon will lead you directly to it. Usually, you will find the description page how the specific REST API of that dataset can be used.

2.8 How often is the TOAR data portal updated?

Due to limited capacity, we cannot actively monitor the validity of the information andn links provided in the TOAR data portal[15]. If you spot any outdated information or broken links, please contact us[16]. It helps a lot if you can point out clearly what is wrong and provide the corrected information in one email.

[12] https://toar-data.org/contribute/#toar_data_portal_contribute
[13] toar-data@fz-juelich.de
[14] https://toar-data.org
[15] https://toar-data.org
[16] toar-data@fz-juelich.de

2.9 Are you planning to list other data than ozone in the TOAR data portal?

No. This is beyond our capacity and it might make the TOAR data portal[17] more cluttered and more difficult to navigate.

[17] https://toar-data.org

CHAPTER

THREE

TOAR DATABASE USER

3.1 What is the TOAR database about and what are TOAR data services?

The TOAR database is a central data repository for global data from surface ozone and ozone precursor measurements. It supports the [Tropospheric Ozone Assessment Report](https://igacproject.org/activities/TOAR/TOAR-II) by enabling scientists around the world to perform standardized analysis of ozone-related data. The TOAR database is operated by the Jülich Supercomputing Centre at Forschungszentrum Jülich, Germany. Users can not directly access the database itself, but we offer various web services (APIs) which allow everyone to search for data, download data, visualize data, or perform online analysis like data aggregations or trend analysis. The latter uses the quantile regression approach that has been agreed upon by the TOAR science community. All TOAR database-related services and extensive documentation can be accessed through our [home page](https://toar-data.fz-juelich.de).

3.2 From where do you get the TOAR data?

The TOAR database team collected data from 18 large air quality monitoring networks and from many individual data providers. Some of these datasets are available through public data services, others were sent to us by individual data providers. In all cases, we took great care to ensure that the data are properly licensed and that we are allowed to re-distribute the data under a permissive [Creative Commons CC-BY4.0](https://creativecommons.org/licenses/by/4.0/deed.en) license. 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. To enable the analysis of ozone change causes, we also extracted time series from the global [ERA5 reanalysis](https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5) by the European Centre for Medium-Range Weather Forecasts at all station locations in the database.

3.3 What is special about the TOAR database?

The TOAR database is one of the largest collections of ozone-related surface measurements world-wide, it is fully committed to Open Data and FAIR principles, and it is accessible through a variety of modern, user-friendly, and performant web services. Another unique selling point of the TOAR database is its very rich set of metadata, which in particular allows for a globally harmonized characterisation of measurement sites based on a set of pre-processed Earth Observation datasets. For details, please consult the Metadata Reference.

3.4 How much data is in the TOAR database?

The TOAR database contains more than 430,000 time series at almost 24,000 stations. There are about 65 billion observation records ranging from the 1970s to the recent past (most data records end in 2022 or 2023 as this is the common analysis period of the TOAR-II assessment). For German stations, we collect and provide near-realtime data with updates 4-times daily. The total data volume of the TOAR database is dangerously close to 10 Terabytes.

3.5 How can I find data in the TOAR database?

All our services are offered through our [home page](https://toar-data.fz-juelich.de). The easiest way to identify the most suitable data for your purpose is probably our [dashboard](https://toar-data.fz-juelich.de/gui/v2/dashboard/). We also offer a [search endpoint](https://toar-data-dev.fz-juelich.de/api/v2/#search-combined-endpoint-of-stations-and-timeseries) at our REST API, which you can use in your browser or through your own programs. Example Python programs how to use the search endpoint are provided in our [TOAR tools](https://gitlab.jsc.fz-juelich.de/esde/toar-public) repository.

3.6 What do you do in terms of data curation?

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 The TOAR Data Processing Workflow for details). The following automated processing workflow contains several checks and quality tests including statistical tests to detect large discrepancies to expected values. Once a dataset passes these tests it will be inserted into the database where providers and the TOAR data 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*). Finally, data errors are often found when visualizing or analyzing data from the TOAR database. Users can point us to obvious data errors and we will correct them as quickly as possible. All changes to data and metadata are logged and can therefore be re-traced if needed. Major data changes will also induce a new dataset version number.

3.7 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).

3.8 Can I filter data of specific quality levels?

Yes. The data flagging scheme of the TOAR database 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 team. Details on how to specify the desired data quality level can be found in TOAR Near Realtime Data Processing.

3.9 What is this thing about “global metadata”?

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 Point Extraction and Aggregation Service (GEO-PEAS). GEO-PEAS has copies of several Earth Observation data sets with spatial resolution on the order of 100 m. These include, for example, the Human Settlement Layer database from JRC Ispra, the NOAA stable nightlight dataset, and the ESA CCI landcover product. GEO-PEAS then allows to calculate aggregated quantities from this data in a user-defined radius around a location, e.g., a measurement site. A predefined set of such aggregates (for example, maximum population density within 25 km radius) is stored together with the station metadata in the TOAR database and can be used to filter the stations that you want to cinlude in your analysis. A detailed description of the global metadata elements can be found in https://esde.pages.jsc.fz-juelich.de/toar-data/toardb_fastapi/docs/toardb_fastapi.html#stationmetaglobal.

3.10 What is the time resolution of the data in the TOAR database?

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. Note that all TOAR analyses are based on hourly data and hour [statistics and aggregation service](https://toar-data.fz-juelich.de/api/v2/analysis/#statistics) has been specifically developed for this time resolution.

3.11 Does the database also contain data from mobile platforms?

No. This would require a different data model. The TOAR working group on marine and polar ozone has, however, established a collection of measurements from mobile platforms, which is available from https://www.jamstec.go.jp/egcr/e/atmos/observation/toar2oceansdata/index.html. Please make sure to cite the associated publication when using this data collection.

3.12 How do I retrieve data?

We offer different [web services](https://toar-data.fz-juelich.de/subpages/data_services.html) through which you can download either hourly values, aggregated statistics (means, min/max, percentiles, and several ozone-specific metrics), or trend estimates. There is also [software](https://gitlab.jsc.fz-juelich.de/esde/toar-public/toargridding) available to generate gridded data products, e.g. for the evaluation of chemistry transport models. Due to technical limitations, the amount of data that can be retrieved is limited. Registered users (see below) can issue larger download requests than anonymous users, and fo rspecial needs we can increase your quote upon request.

3.13 Who can become a registered user?

The TOAR database is open to everyone with or without registration. Anonymous use of our web services is possible, but the amount o fdata that you can process and retrieve are limited. If you [register](***REGISTRATION LINK***), you will be able to save your preferences, keep a history of your data processing requests, and you will be able to process larger requests and download more data. According to the [CC-BY4.0 license](https://creativecommons.org/licenses/by/4.0/deed.en), there are no use restrictions on our data, but you will be obliged to properly acknowledge the data sources. Please see ***link to citation*** to learn how we support you with this.

3.14 What is the data format of data downloads?

Any query to the TOAR database REST API returns a JSON structure per default. Data queries can also be returned as csv file. Therefore, 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. To speed up processing, some of our services allow you to specify the _fields_ that you want to include in the metadata output. Choosing these wisely can significantly reduce processing time and data volumes. The [TOAR gridding software](https://gitlab.jsc.fz-juelich.de/esde/toar-public/toargridding) generates netCDF files from your data queries.

3.15 Under which conditions may I use the data?

All TOAR-II data are provided without restrictions under a CC-BY 4.0 licence. This licence requires that you acknowledge the data source. 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 (https://toar-data.fz-juelich.de/footer/terms-of-use.html). ***ADD information on the new license endpoint! History of queries as registered user - how to go from there to citations for all the dfata that was requested in an earlier API call?***

3.16 How can I cite the TOAR database?

The TOAR database should be cited as Schröder et al; ***AD paper title!*** (in preparation) ***hopefully not too much longer***

For individual data series or data collections, the original data sources should be cited. A recommended citation is provided with the metadata when data are downloaded and can be obtained by sending a query to the endpoint ********. Please make sure to at least acknowledge our data providers. You will also find a suggested acknowledgement string in the query responses.

3.17 Do you make use of controlled vocabulary?

Yes. You can retrieve the ontology of TOAR data as xml from https://toar-data.fz-juelich.de/api/v2/ontology or you can have see it online as OWL document via https://toar-data.fz-juelich.de/documentation/ontologies/v1.0/. [***Ela: please check if still functional***]

3.18 What is the database schema underlying the TOAR database?

You can check out our [git pages](https://esde.pages.jsc.fz-juelich.de/toar-data/toardb_fastapi/docs/toardb_fastapi.html) documentation for details.

3.19 Where can I get further support?

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

CHAPTER

FOUR

TOAR DATA PROVIDERS

4.1 How can I contribute data to TOAR?

We are always happy to learn about new datasets that we can include in our TOAR database or that we can link in the TOAR data portal. Nevertheless, there are a few conditions before we can accept your data. First, you must make sure that you have the right to distribute the dataset and that we can re-distribute it under the CC-BY4.0 license. If in doubt, please contact us. Furthermore, we can only accept submissions of station data; other types of data can only be linked to via the TOAR data portal. Also, we require at least one full year or two “ozone seasons” of hourly (or higher frequency) measurements and the measured species must be ozone or ozone precursors or basic meteorological variables. And your data must come from well-maintained research-grade instruments, i.e., we do not accept data from low-cost sensor networks - although this may change in the future. Technically, you can either submit data in csv format, send us a link to an API where we can retrieve your data, or send us a database dump. Please adhere to the format specifications dscribed in ** or get in contact with us to find out what needs to be done so that we can accept your data. We are quite flexible and have several reading routines, but we cannot accomodate every possible format. Finally, please ensure that you have sufficiently complete and correct metadata associated with your measurements. Again, we will be happy to help and discuss details with you if you contact us.

4.2 When will my data be inserted into the TOAR database?

At this point, the TOAR database is frozen in preparation of the second TOAR assessment report. All our efforts are now directed to improvingthe data quality and performance of our services. We will still accept new data and inspect it, but the processing will not happen in 2025. The only exeptions will be made for data from Africa, which we will accept until the last possible minute.

4.3 Where do I find the format specification for data submissions?

****LINK****. Please note that, as described above, we may be able to accept data in other formats as well, in particular if you want to submit data from many different sites.

4.4 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.

4.5 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

4.6 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. We don’t do this regularly. Thus, if you have made a new submission to EBAS, please let us know.