

Climate Trace Data on Google BigQuery

Climate TRACE monthly releases and other source level datasets are publicly available on Google BigQuery. This guide provides instructions on how to access, query and export this data.

Data Location

The Climate TRACE dataset is hosted in the following BigQuery project: [trace-data-383422](#). The project contains a public dataset named `climate_trace`.

Direct link to the `climate_trace` dataset:

https://console.cloud.google.com/bigquery?ws=!1m4!1m3!3m2!1strace-data-383422!2sclimate_trace

💡 Click the "star" icon next to the `trace-data-383422` project or `climate_trace` dataset in BigQuery to have it pinned to your console.

Data Catalog

The dataset contains the following tables:

Emissions tables

- `emissions_sources` : Individual emissions sources data with location, emissions quantity, activity and emissions factors
- `country_emissions` : Aggregated emissions data by country and subsector
- `gadm_emissions` : Emissions aggregated to GADM administrative boundaries
- `city_emissions` : Emissions aggregated to city boundaries
- `emissions_sources_location` : Location and boundary reference data for emissions sources

Metadata Tables

- `emissions_sources_confidence` : Confidence score for different data elements (capacity, activity, emissions factors etc)
- `emissions_sources_uncertainty` : Uncertainty ranges for emissions data
- `emissions_sources_ownership` : Parent company and ownership structure information
- `ownership_data_source` : Source documentation for ownership data

- **geometries** : Geographic boundaries for GADM administrative regions and cities referenced by the **geometry_ref** column

Emissions Reduction Solutions (ERS) tables

- **strategy** : Emissions reduction strategies with emissions factor changes and activity impacts
- **strategy_crosswalk** : Links emission sources to applicable emissions reduction strategies
- **gadm_spatially_uncertain_strategy** : Strategies for spatially uncertain sources at GADM level
- **gadm_spatially_uncertain_strategy_crosswalk** : Links GADM regions to strategies
- **city_spatially_uncertain_strategy** : Strategies for spatially uncertain sources at city level
- **city_spatially_uncertain_strategy_crosswalk** : Links cities to strategies
- **reductions** : Emissions reduction calculations for individual sources
- **gadm_spatially_uncertain_reductions** : Reductions for spatially uncertain sources at GADM level
- **city_spatially_uncertain_reductions** : Reductions for spatially uncertain sources at city level

Sector specific datasets

More granular datasets for specific transportation subsectors - **road-transportation** , **domestic-shipping** and **international-shipping** :

- **road_segments** : Road transportation data at road segment level
- **shipping_voyages** : Individual ship voyage data including vessel identifiers (IMO/MMSI), voyage emissions, distances and port information for both domestic and international shipping
- **road_segments_reductions** : Emissions reductions from applying ERS to the road segments data
- **shipping_vessel_reductions** : Emissions reductions from applying ERS to the shipping vessel data

Data Updates and Versioning

Climate TRACE tables are updated monthly as part of our regular release schedule. Each table is labeled with its release version, allowing you to track which version of the data you're working with. To check the release version of a table:

- Click on the table name in the BigQuery Console to view its details
- Look for the "Labels" section in the table details - the **release** label shows the version (e.g., v5_1_0)

Updates are applied to all tables as we receive updated data from sector teams.

Access Requirements

1. **Google Cloud Account:** Users need a Google Cloud account to access the data.
2. **BigQuery Access:** The **climate_trace** dataset is public with Read-Only access.
3. **Query Permissions:** Users must have the required permissions on their own Google Cloud project to be able to run queries against BigQuery tables.

Required Permissions

To run queries, users need the **bigquery.jobs.create** permission on their Google Cloud project. This permission should be included if you have one of the following roles on your project:

1. **Project Owner:** Full control over the project, including all BigQuery resources.
2. **BigQuery Job User:** Allows running queries and creating jobs in BigQuery.
3. **BigQuery User:** Includes more comprehensive permissions, including job creation.
4. **BigQuery Admin:** Provides full control over BigQuery resources.

Getting Started

1. Sign in to your Google Cloud Console.
2. Navigate to BigQuery.
3. In the Explorer pane, add the **trace-data-383422** project. Alternatively, navigate to the **climate_trace** dataset here:
https://console.cloud.google.com/bigquery?ws=!1m4!1m3!3m2!1strace-data-383422!2scclimate_trace.
4. Expand the **climate_trace** dataset to view available tables.
5. Write and run your queries against these tables.

Sample Query

Here's a simple query to get you started:

SQL

```
SELECT iso3_country, start_time, SUM(emissions_quantity) AS  
co2e_100yr  
FROM `trace-data-383422.climate_trace.country_emissions`  
WHERE subsector='electricity-generation'  
AND gas='co2e_100yr'  
GROUP BY iso3_country, start_time  
ORDER BY iso3_country, start_time
```

Exporting Data From BigQuery

BigQuery supports multiple export methods depending on the data size and format needed:

Method 1: Download query results (small to medium datasets)

1. Run your query in the BigQuery console
2. Click “Save Results” in the query results panel
3. Choose your format: csv, json, parquet

Method 2: Export to Google Cloud Storage (large datasets)

1. Create a Google Cloud storage bucket in your project
2. Make sure the active project (top left dropdown) is set to your own project and you have **BigQuery Job User** (or higher) permissions.
3. You can export the data to **Parquet** as follows:

SQL

```
# As an example if you wanted to export shipping voyages data for  
2025  
# This query exports to a folder named shipping_voyages_2025 in  
your cloud  
# storage bucket  
# Replace `your-bucket-name` with your own GCS bucket  
EXPORT DATA OPTIONS(  
  uri='gs://your-bucket-name/shipping_voyages_2025/*.parquet',  
  format='PARQUET',
```

```

        overwrite=true
    ) AS
SELECT *
FROM `trace-data-383422.climate_trace.shipping_voyages`
WHERE EXTRACT(YEAR FROM start_date) = 2025;

```

- Alternatively, you could use the **bq** command line tool to export data.

 **Prerequisites:** Install the Google Cloud SDK (includes the bq command-line tool). See installation instructions at <https://cloud.google.com/sdk/docs/install>

- Use the **bq** command line tool to export to parquet format (recommended for large datasets):

Shell

```

bq extract --destination_format=PARQUET \\\
'trace-data-383422:climate_trace.shipping_voyages' \\\
'gs://your-bucket-name/shipping_voyages_*.parquet'

```

Method 3: Filtered Export with Query Results

To export only specific years or filtered data, or to create your own copy of the climate trace tables:

- Create a temporary table with your filtered query:

SQL

```

CREATE OR REPLACE TABLE
your-project.your-dataset.shipping_2020_2025 AS
SELECT *
FROM trace-data-383422.climate_trace.shipping_voyages
WHERE EXTRACT(YEAR FROM start_time) >= 2020
AND EXTRACT(YEAR FROM start_time) <= 2025;

```

- Analyze the data within BigQuery or export the table using Method 1 or Method 2 above

Cost Considerations

Please note that BigQuery decouples storage and query costs. For public datasets like `climate_trace`:

- Storage costs are covered by the dataset owner i.e. Climate TRACE.
- Query costs are incurred on the querying user's project and account.
- All queries and exports run as BigQuery jobs and will be billed to your Google Cloud project and the associated billing account. Ensure that your own project is set as the active project to avoid permissions issues and to properly track costs to your billing account.

BigQuery charges based on the amount of data processed by your queries. To optimize costs:

- Use WHERE clauses to filter data and reduce processing
- Select only the columns you need instead of using SELECT *
- Use the query validator in the BigQuery console which shows data processing amount for the query before running it.

Additional Resources

- [BigQuery Documentation](#)
- <https://cloud.google.com/bigquery/docs/best-practices-performance-compute>
- [Climate TRACE Website](#)
- [BigQuery Pricing](#)
- Climate TRACE [Terms of Use](#)

For further assistance or questions about the Climate TRACE dataset, please contact @Ishan Saraswat [ishan@watttime.org].