EITI SOE database: user guide

Queries, filters, downloads, and API

Table of contents

Overview of the EITI SOE database

EITI's state-owned enterprises (SOE) database is an open access tool that provides regularly updated data on state-owned enterprises. The data covers all SOEs' transactions reported through the EITI since 2017, drawing on disclosures submitted from implementing countries. The data includes payments to governments, disaggregated per revenue stream; links to published audited financial statements; links to stock exchange listings; and commodities sold by SOEs.

The SOE database allows users to:

- explore the data using an interactive website
- query the data using SQL
- share/access the data via an integrated API.

SOE database web interface

The SOE database provides a user-friendly web interface that allows users to navigate tables, filter and search data, and perform other operations. The key pages are described below.

About/home page

The About/home page includes information about the database, a quick link to the list of SOEs from the EITI summary data files and a link to the full SOE database.
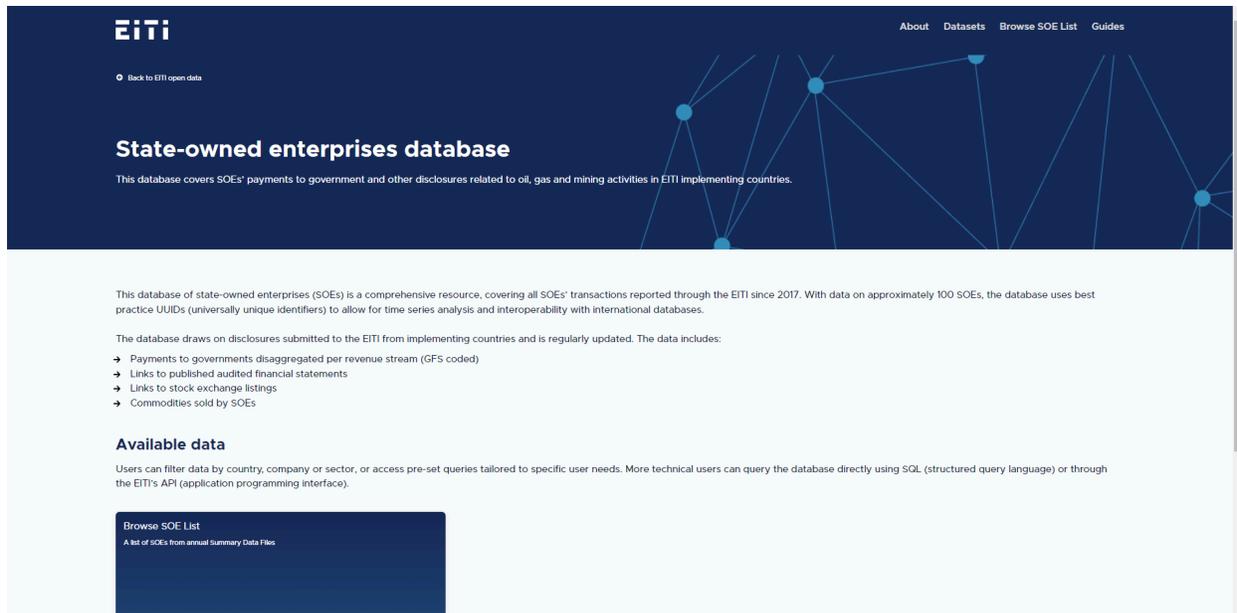
Figure 1. EITI SOE database about/ home page

Datasets page:

The datasets page provides access to the tables, views and pre-set queries that are available:

Firstly, users can explore the SOE data:

- The list of SOEs from EITI's summary data files
- Annual SOE payments by governments
- SOE annual revenue generated: data on payments to governments annually from SOEs operating in EITI implementing countries

Users can explore the full database:

- About: information from the 'about' tab on summary data files and EITI reports
- Agencies: data specific to declared government entities
- Companies: data related to declared company payments
- Projects: data specific to declared projects

A custom SQL query is also available.

Explore the SOE data

Focus on...	Focus on...	Focus on...
SOE List	**Declared Revenue by Year**	**Annual Revenues from SOEs**
→	→	→

Explore the full database

ABOUT

The About section of the summary data files, with one row per declaration

73 rows

→

AGENCIES

The data specific to declared government entities

2,319 rows

→

COMPANIES

The data related to declared company payments

Many rows

→

PROJECTS

The data specific to declared projects

4,977 rows

→
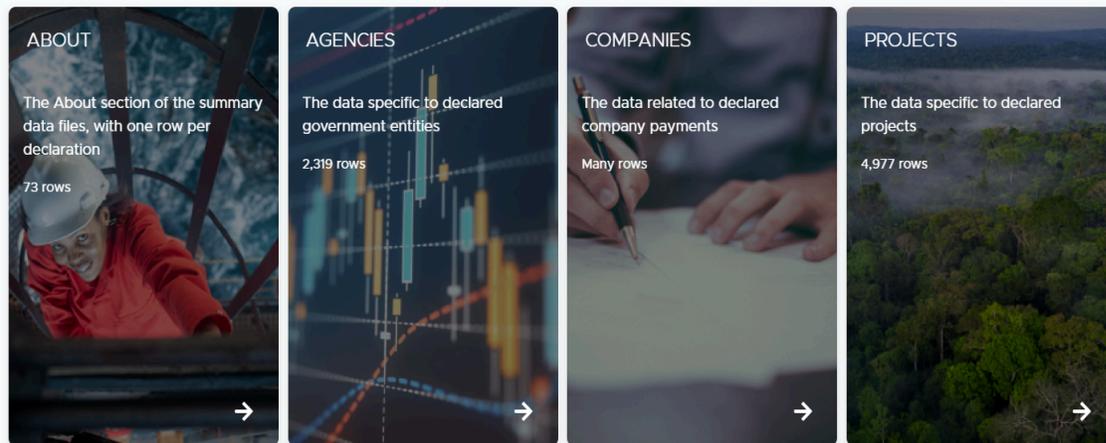
Custom SQL query

Figure 2. datasets page

Browse the SOE list page:

This page provides easy access to the SOE data. Users can search by country, sector, commodity, etc and easily access audited financial statements where available.

Figure 3. SOE list page

Table page/s, views, and pre-set queries

SOE specific views and the full database tables allow users to:

- explore the contents of a database by applying sorts, filters, searching for texts, and facets
- import the table/data as JSON or CSV
- create SQL queries using the View and edit SQL button

Figure 4. agencies table page

Aside from full tables, the SOE database also has pages for views and saved queries. These work the same way as table pages.

Applying filters

Filters allow you to narrow down your search to the most relevant records. When you visit any table or view page in the SOE database, you will see a filter interface and predefined filters for the data.

Figure 5. Notice the filter interface and predefined searches (e.g.country, year)

To apply a filter, select a **column**, an **operation** (e.g. =, contains), and a **value.** Then click **Apply**.

You can apply multiple filters at the same time but you must **apply each filter separately**. For example, to filter the Declared Revenue by year table above for the **year 2018** and for the countries of **Ghana, Tanzania,** and **Chad**, you will first:

1. **Apply** a filter for the **year = 2018**

Figure 6. Year filter applied

2. Then **apply** a second filter for **country in Ghana, Tanzania, Chad.** After pressing **Apply** you will see the resulting filtered data:

Figure 7. Multiple filters applied

IMPORTANT: Remember to click **Apply** to apply your filters to the data.

Removing a filter

1. Select **– remove filter –** in the **column dropdown** of the filter interface for the filter you want to remove. Note that you can **remove multiple filters at one time**. Click **Apply**.

Figure 8. Removing filters

Accessing the data as CSV or JSON

The SOE database allows you to download your query results in formats like CSV or JSON making it easy to use the data in other tools or for reporting.

Any table, view, or query result can be accessed directly from the SOE database interface either as JSON or CSV by:

1. Selecting an option from the **This data as** menu. By default, this will display the JSON or CSV representation of the table you are looking at. (**NOTE:** Just the current rows you are viewing on the page and NOT the entire table on the browser.)

> This data as json, CSV (advanced) or download the CSV

2. Using the **Advanced export** form **that provides more control over the resulting file**.

> ### Advanced export
> JSON shape: default, array, newline-delimited
> CSV options: ☐ **download file** ☐ **stream all rows** [Export CSV]

The advanced export options are:

- **download file** – instead of displaying CSV in your browser, this forces your browser to download the CSV to your downloads directory.

- **expand labels** – if your table has any foreign key references this option will cause the CSV to gain additional `COLUMN_NAME_label` columns with a label for each foreign key derived from the linked table. In this example the `city_id` column is accompanied by a `city_id_label` column.

- **stream all rows** – by default CSV files only contain the first max_returned_rows records. This option will cause the SOE database to loop through every matching record and return them as a single CSV file.

> **TO LEARN MORE ABOUT CSV EXPORTS:**
> You can visit https://docs.datasette.io/en/latest/csv_export.html

Queries are just URLs and vice versa

In the SOE database, a query can be as simple as typing a URL in your browser. Just like navigating to a website, you can navigate to a specific dataset and apply filters directly through the URL. You can then save the URLs, which you can reuse to retrieve/update your search results at a later date.

For example:

1. You can access the **eiti_database** and see its tables, views, and canned queries by going to: https://soe-database.eiti.org/
2. You can access the **declaration_companies** table inside of **eiti_database** by going to eiti_database: companies: 32,909 rows
3. You can access the **soe_companies_revenue_annual** view inside of the **eiti_database** by going to eiti_database: Declared Revenue by Year: 162 rows
4. You can query **the agencies_rev enue_from_soe_annual** view for records from **Norway** during the **year 2018** sorted by **percentage_country_usd** by going to eiti_database: Annual Revenues from SOEs: 178 rows
5. You can apply a complex query getting the list of companies from the declarations_company table that pay to specific government entities from the declaration agencies table (see below)

```sql
SELECT
    a.government_entity,
    a.eiti_id_government,
    GROUP_CONCAT(DISTINCT c.company_name) AS company_names,
    GROUP_CONCAT(DISTINCT c.eiti_id_company) AS eiti_id_companies,
    a.country AS government_country
FROM
    declaration_companies c
JOIN
    declaration_agencies a ON c.eiti_id_government = a.eiti_id_government
GROUP BY
    a.government_entity,
    a.eiti_id_government,
    a.country;
```

Figure 9. A complex query applied to the database

by going to
eiti_database: select rowid, company_name, eiti_id_company, project_name, eiti_id_project, company_commodities, [company_audited_financial statement_or_equivalent], company_public_listing_or_website, company_sector, company_id, company_type, revenue_value, original_company_name, opencorporates_id, company_relation, government_entity, eiti_id_government,

> **revenue_stream_name, levied_on_project, reported_by_project, reporting_currency, payment_in_kind, in_kind_volume, in_kind_unit, comments, country, iso_alpha3_code, eiti_id_declaration, iso_alpha2_code, year, start_date, end_date from companies order by rowid limit 26**

Similarly, if you apply any filters, facets or queries on the SOE database user interface, you will notice that the page URL changes to reflect the filter, facet, or query that you've applied to the data.

Querying the data using SQL

There are several ways to query the data using SQL commands.

From the datasets page

You can query directly from the datasets page using the **Custom SQL query** dialog.
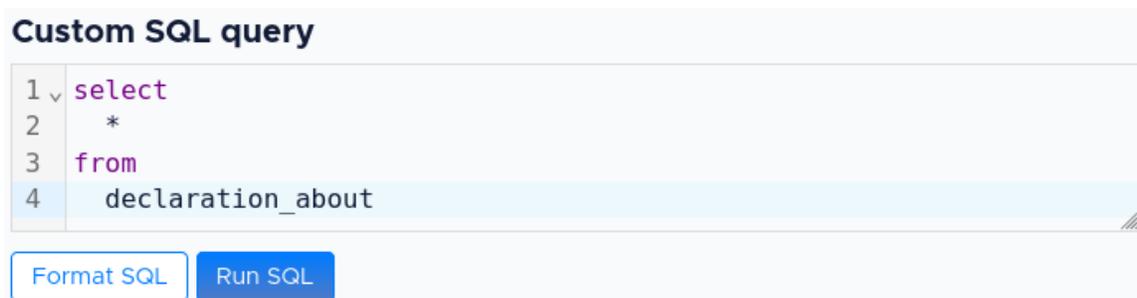
Figure 10. The Custom SQL query in the datasets page

To run your Custom SQL query, click **Run SQL.**

From the table and view pages

You can also apply a query on the database from any table or view page by clicking

This will open the Custom SQL query dialog

Figure 11. The Custom SQL query in the declaration_agencies table

Once the SQL commands are entered and applied, the data can be downloaded or used as an API for continued use.

IMPORTANT
When running custom SQL queries from any page (e.g. table, view, etc.) on the SOE database, you are not limited to just using that table or view. You can use any table or view on the database that you are allowed to access in your query.

Using the SOE database as an API

The SOE database can serve as an API allowing you to integrate data queries into your applications. This enables automated data retrieval and integration with other systems.

Accessing CSVs

By default, showing or exporting the data on the SOE database as a CSV will return a URL that can be used to access the CSV representation of the data. You can then use this URL to import the CSV, download the CSV, or access the CSV programmatically.

The following options can be used to customise the CSVs returned in the SOE database (from Datasette's documentation).

- `?_header=off`

- - this removes the first row of the CSV file specifying the headings (i.e. only the row data will be returned)
- `?_stream=on`
  - stream all matching records, not just the first page of results.
- `?_dl=on`
  - causes Datasette to return a `content-disposition: attachment; filename="filename.csv"` header.

Accessing JSONs

Similarly, the SOE database provides a JSON API for its associated databases. Anything you can do through the SOE database user interface can also be accessed as JSON via the API. This means that the database tables, views, as well as the results of queries and filters can easily be accessed as JSON by using the appropriate URL.

To access the API for a page, you can click the **This data as json** link or add **.json** extension to it.

For example, to access the JSON representation of the **rows in the [eiti_database: Annual Revenues from SOEs: 178 rows](#)**l page, you can simply add .json at the end of the URL like so: [soe-database.eiti.org/eiti_database/Annual~20Revenues~20from~20SOEs.json](#)

This will open the JSON for the first 50 rows (default) in the view. You can change the number of rows exported by adding the `?_size=` parameter in the URL.

```
1  {
2      "ok": true,
3      "next": "25",
4      "rows": [
5          {
6              "Year": "2018",
7              "Country": "Philippines",
8              "Government Entity": "PHILIPPINE PORTS AUTHORITY (PPA)",
9              "Revenue Generated": 0,
10             "Currency": "PHP",
11             "USD Equivalent": 0,
12             "Share of the Country Total": 0,
13             "EITI ID": "0130300a-538c-4ba7-8491-a3df4c2fc45e"
14         },
15         {
16             "Year": "2018",
17             "Country": "Ukraine",
18             "Government Entity": "THE STATE SERVICE OF GEOLOGY AND SUBSOIL OF UKRAINE (DERZHGEONADRA)",
19             "Revenue Generated": 3493044444,
20             "Currency": "UAH",
21             "USD Equivalent": 95010808876.8,
22             "Share of the Country Total": 0.05,
23             "EITI ID": "01874f0f-737d-4d4f-89f7-c2d07e7e60a5"
24         },
25         {
26             "Year": "2018",
27             "Country": "Tanzania",
28             "Government Entity": "MINISTRY OF MINERALS",
29             "Revenue Generated": 1915346760,
30             "Currency": "TZS",
31             "USD Equivalent": 4318015196146.8,
32             "Share of the Country Total": 0.29,
33             "EITI ID": "0984e695-ace7-454a-a822-6ea22fd71226"
34         },
35         {
36             "Year": "2018",
37             "Country": "Albania",
38             "Government Entity": "MUNICIPALITY OF RROSKOVEC",
39             "Revenue Generated": 3076584,
40             "Currency": "ALL",
41             "USD Equivalent": 332301837.84,
42             "Share of the Country Total": 0,
43             "EITI ID": "0a25998b-4072-44ed-8bbe-28a0884c3529"
44         },
45         {
46             "Year": "2018",
47             "Country": "Mongolia",
48             "Government Entity": "GENERAL AGENCY FOR LABOR WELFARE SERVICES",
49             "Revenue Generated": 3042048000,
50             "Currency": "MNT",
51             "USD Equivalent": 8039889500160,
52             "Share of the Country Total": 0.03,
53             "EITI ID": "0bc48680-3685-422d-9927-e23573b7825d"
54         },
55         {
56             "Year": "2018",
57             "Country": "Cote d'Ivoire",
58             "Government Entity": "DIRECTION GÉNÉRALE DES MINES ET DE LA GÉOLOGIE (DGMG)",
59             "Revenue Generated": 1350136200,
60             "Currency": "XOF",
61             "USD Equivalent": 749659074641.4,
62             "Share of the Country Total": 0.06,
63             "EITI ID": "0d8c7312-4b6e-4fb7-8437-c22ff925551d"
64         },
65         {
66             "Year": "2018",
67             "Country": "Togo",
68             "Government Entity": "DIRECTION GÉNÉRALE DES MINES ET DE LA GÉOLOGIE (DGMG)",
69             "Revenue Generated": 578629800,
70             "Currency": "XOF",
71             "USD Equivalent": 321282460560.6,
72             "Share of the Country Total": 1.15,
73             "EITI ID": "0d8c7312-4b6e-4fb7-8437-c22ff925551d"
74         },
75         {
76             "Year": "2018",
77             "Country": "Nigeria",
78             "Government Entity": "NIGERIAN CONTENT DEVELOPMENT AND MONITORING BOARD (NCDMB)",
79             "Revenue Generated": 45200000,
80             "Currency": "USD",
81             "USD Equivalent": 45200000,
82             "Share of the Country Total": 0.04,
83             "EITI ID": "142e5955-89ff-43e1-aaaa-af7ce193689a"
84         },
85         {
86             "Year": "2018",
87             "Country": "Philippines",
88             "Government Entity": "DEPARTMENT OF ENERGY (DOE)",
89             "Revenue Generated": 0,
90             "Currency": "PHP",
91             "USD Equivalent": 0,
```

```
JSON of query or filter results
```

When accessing the JSON representation of queries or filter results, you will follow the same approach. Put **.json** after the page url (e.g. table or view) and put the query after that.

For example, in order to get the JSON representation of our filter in the **soe_companies_revenue_annual** view for the **year 2018** and countries **Ghana, Tanzania,** and **Chad**, we use the following URL:

soe-database.eiti.org/eiti_database/Annual+Revenues+from+SOEs.json?Year__contains=2018

Table or URL arguments

The SOE database has several URL options for fine-tuning how the CSV or JSON values are exported. For example:

- `?_size=1000` or `?_size=max`
  - Sets a custom page size. This can be used to get a subset of the rows or all the rows being queried as long as the number of rows in the do not exceed the max_returned_rows setting (1,000,000 by default). Use `max` to get `max_returned_rows`.
- `?_sort=COLUMN`
  - Sorts the results by the specified column.
- `?_col=COLUMN1&_col=COLUMN2`
  - List specific columns to display. These will be shown along with any primary keys.
- `?_search=keywords`
  - For SQLite tables that have been configured for full-text search executes a search with the provided keywords.
- `?_facet=column`
  - Facet by column. Can be applied multiple times, see Facets. Only works on the default JSON output, not on any of the custom shapes.
- `?column__exact=value` or `?_column=value`
  - Returns rows where the specified column exactly matches the value.
- `?column__contains=value`
  - Rows where the string column contains the specified value (`column like "%value%"` in SQL).

To learn more, you can refer to Datasette's documentation on its JSON API and Pages and API endpoints.

Other links and resources

- ○ Website: https://datasette.io/
- ○ Documentation: https://docs.datasette.io/en/latest/
- ○ GitHub repository: https://github.com/simonw/datasette

Technical information

The EITI SOE database is a **Datasette** product. Datasette is an open source multi-tool for publishing and exploring data. Datasette is built using Python and leverages SQLite as its underlying database engine. It provides a simple yet powerful interface for interacting with data, making it accessible to users with varying levels of technical expertise.

The version used for the EITI SOE database instance is: 1.0a13 (latest

Version and changelog

Date	Person	Changelog
2024-05-08	CLi	Initial version prepared by Civic Literacy Initiative
2024-05-18	CLi	Added initial content
2024-07-29	EITI Data Team	Customised to EITI SOE database