demscore

Documentation

Demscore Handbook

User Support



November 2026

Copyright © Demscore
All rights reserved

demscore

DEMSCORE IN NUMBERS

212 DATASETS

PARTNERING RESEARCH PROJECTS	**6**
NUMBER OF OUTPUT UNITS	**57**
DATASET TO OUTPUT UNIT COMBINATIONS	**1350**

26,091 VARIABLES

DIRECT TRANSLATIONS	**195**
INDIRECT TRANSLATIONS	**151**
NUMBER OF VARIABLES AVAILABLE THROUGH OUR GRAPHING TOOLS	**5062**

443,873 VARIABLE FILES

Demscore is a collaboration between the University of Gothenburg, Stockholm-, Uppsala-, and Umeå University, funded by the Swedish Research Council. The data is provided by the partnering research institutes: V-Dem, QoG, UCDP/VIEWS, COMPLAB, REPDEM, and H-DATA.

TABLE OF CONTENTS

1 DEMSCORE in Numbers

A compilation of essential descriptive statistics offering an overview of the DEMSCORE Infrastructure's scope and breadth.

2 Methodology

A summary of the Demscore Methodology, including a description of the general approach, and dataset- and Output Unit combinations.

3 Merge Information

Information on Merge Scores, Differences in Country Identifiers, and Lost Observations.

4 Integrating External Data

How to merge external datasets using DEMSCORE.

5 Download Instructions

A user-friendly guide featuring step-by-step instructions, as well as valuable tips and tricks, for navigating the download interface effectively.

6 Graphing Tools

Graphing tools for exploring and visualizing DEMSCORE data.

7 List of Datasets

A comprehensive list of datasets that are available through DEMSCORE. Each entry includes the dataset's name, year of coverage, analytical level, and the number of variables.

8 List of Output Units

A comprehensive list of DEMSCORE Output Units.

2. DEMSCORE METHODOLOGY

The following section includes a summary of the DEMSCORE Methodology, including a description of the general approach, and dataset- and Output Unit combinations. To read and download the full version of the DEMSCORE Methodology, please visit www.demscore.se

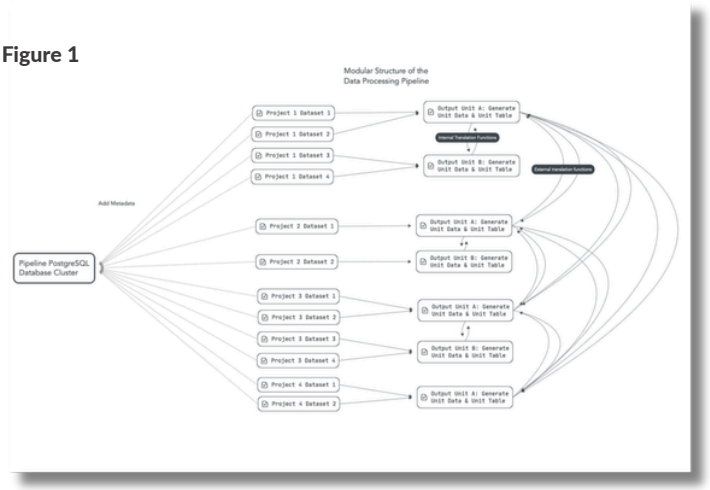
The potential that lies in the interdisciplinarity of DEMSCORE data is accompanied by the challenge of its diversity. On this page, we outline a summary of the methodological considerations, choices, and procedures guiding the development of the DEMSCORE project, stretching across the construction of a common e-infrastructure to fully harmonized data from all sources.

We continually review our methodology—and occasionally adjust it—with the goal of improving the quality of DEMSCORE indicators and indices. We report the challenges of the harmonization process and present the solutions provided by the DEMSCORE Team and its member projects.

In order to present data merged from different sources in a consistent way, we have developed an innovative and systematic data harmonization process. Each dataset in Demscore has its own dataset unit, i.e., unit of analysis. Based on their units of analysis, datasets are grouped into Output Units and finally merged and translated through these Output Units.

Several datasets within the same module have very similar units. That is, if more than one dataset within a module has the combination of columns containing e.g., the country and year as their identifying unit, we take the union of the country-year combinations from each dataset with similar units and define the result as an Output Unit. This is illustrated in figure 1, which also demonstrates that we use these Output Units to translate and merge data from different units and modules together, i.e., we translate between Output Units both within and across projects.

Figure 1



While Dataset A of Project X for example has observations for European countries from 1944 to 2015, Dataset B for Project X has observations for all countries from 1990 to 2021. The Country-Year unit table for Project X will have rows for European countries from 1944 to 1989 as well as rows for all countries from 1990 to 2021. In other words, a unit table always covers all possible combinations of unit values in the identifier columns of datasets with the same dataset unit. This is illustrated using the following example based on a subset of the data above: Member Project X has two datasets available; Dataset A and Dataset B. Dataset A is a table with three columns, called country, year and var_1 with corresponding values in five rows. Dataset A is a table with three columns, called country, year and var_1 with corresponding values in five rows. Dataset A has the Unit Country-Year, meaning each row in the dataset is identified by a unique combination of countries and years.

Table 1: Project X: Dataset A

country	year	var_1
ITA	1988	1
ITA	1989	2
ITA	1990	3
DEU	2013	4
DEU	2014	5
DEU	2015	6

Dataset B is a table with three columns, country_id, year and var_2, with corresponding values in eight rows. It also has Country-Year as its unit, meaning each row in the dataset is identified by a unique combination of countries and years. Please note that Dataset B contains some country and year combinations that can be found in Dataset A as well, but also country and year combinations that are not included in Dataset A.

Table 2: Project X: Dataset B

Country_id	Year	var_2
ITA	1990	a
ITA	1991	b
ITA	1992	c
DEU	2015	d
DEU	2016	e
DEU	2017	f
NPL	2005	g
NPL	2006	h
NPL	2007	i

We first identify a unit grid, i.e., unit table for datasets with common dataset units, in this case the Country-Year unit for Dataset A and Dataset B.

The Country-Year Unit Table for Project X contains all Country-Year combinations from Dataset A as well as all Country-Year combinations from Dataset B without duplicates.

If a user now selects the Output Unit Project X Country-Year Unit, and also selects the variables var_1 from Dataset A and var_2 from Dataset B, or even both at the same time to be expressed in this Output Unit, they receive a new dataset with twelve rows sorted the same as the unit table and containing the selected variables.

Table 3: Unit Table for Project X Country-Year Unit

u_x_cy_country	u_x_cy_year
ITA	1988
ITA	1989
ITA	1990
ITA	1991
ITA	1992
DEU	2013
DEU	2014
DEU	2015
DEU	2016
DEU	2017
NPL	2005
NPL	2006
NPL	2007

As Dataset A does not include all country-years included in Dataset B and vice versa, but the unit table encompasses all combinations from both datasets, var_1 and var_2 have missing values (NA = Not Available) for those Country-Year combinations not included in their original dataset when retrieved through the Country-Year Output Unit of Project X.

Table 4: New Dataset from Selection of Variables from Datasets A and B

u_x_cy_country	u_x_cy_year	var_1	var_2
ITA	1988	1	-11111
ITA	1989	2	-11111
ITA	1990	3	a
ITA	1991	-11111	b
ITA	1992	-11111	c
DEU	2013	4	-11111
DEU	2014	5	-11111
DEU	2015	6	d
DEU	2016	-11111	e
DEU	2017	-11111	f
NPL	2016	-11111	g
NPL	2017	-11111	h
NPL	2018	-11111	i

Note: var_1 and var_2 combined in the Project X Country-Year Unit

At this point we assume that every dataset has a Primary Unit and an attached unit table. Note that many Output Units may have only been constructed using a single dataset. Also note that Output Units are constructed across datasets within the same member project, but never across datasets from different member projects. The Country-Year unit table for Project X will thus not include Country-Year observations from datasets in Project Y if these Country-Years only exist in project Y.

Datasets that belong to the Country-Year Unit of Project X can now be translated to Other Units, e.g., the hypothetical Country-Year Unit of Project Y. Through aggregation and disaggregation, datasets that initially collect information on a Country-Year level also become available in other Output Units such as Cabinet-Date or Agency-Year.

3. MERGE INFORMATION

For more detailed information, we refer users to the explanatory notes of our Methodology Document, available on https://www.demscore.se

Merge Scores

We provide three merge scores in Demscore:

1. The number of non-missing observations in the original Output Unit of the variable.
2. The number of non-missing observations in the chosen end Output Unit for that variable.
3. In direct translations, the number of lost observations between the variable in its original unit and the variable in the chosen end Output Unit.

Here are a few general guidelines on how to read and interpret the merge scores offered in Demscore:

- If the score for a variable is very high in the original Output Unit but very low in the end Output Unit, and at the same time the score for lost observations is very high, this means that overlap in identifier combinations between these two Output Units is low.
- If the merge score is high in the original Output unit, but low in the end Output Unit and at the same time, the number of lost observations is low, this means that you have probably chosen a variable that is available only in very few identifier combinations, compared to the identifier combinations in the end Output Unit. However, the end Output Unit covers a lot of the observations from the original Output Unit of the variable.

The merge scores in Demscore can thus provide the user with a hint of what to expect from the downloaded data. However, we recommend users also investigate their customized dataset and which observations that matched, before deciding on whether or not to use the chosen variable for their analysis.

Differences in Country Identifiers

A commonly used identifier across datasets in DEMSCORE modules is countries, but country definitions and names often vary. This poses a challenge for merging data.

As a general rule, resolving merge conflicts depends on the chosen Output Unit. Merge scripts always prioritize preserving the data quality of the units of that Output Unit. This means that we follow the country definitions of the dataset chosen as the Output Unit. When merging from other datasets we only keep combinations that match the chosen Output Unit.

While this resolves most merge issues, some conflicts remain due to differences in what territories are included in e.g. "France" or "Russia" over time. The result of this process is a set of versions of indicators of democratic institutions and concepts, which allow academics and policymakers alike to understand the different features of a polity.

We list the most notable differences in country definitions that users should pay attention to when using data merged on country identifiers in the Methodology Document and on demscore.se.

For differences in the full country names across data sources and how they are adjusted for merges and translations between Output Units in DEMSCORE, see the original DEMSCORE translation functions in our code which is publicly available on: https://github.com/demscore/

We want to emphasize that DEMSCORE expresses no opinions on sovereign claims to disputed territories neither through the descriptions in the Methodology Document nor through any merge decisions. Nor does DEMSCORE make any judgments concerning which territories qualify as countries.

Missing Data

DEMSCORE indicates different types of missingness for observations in the customized datasets:

Missing in original data = Whenever an observation in the original variable is missing (NA, missing code such as 7777, blank cell), we preserve this missing value. When the original source has special codes for various types of missing, those are preserved.

Missing code: -11111 = DEMSCORE code for observation is missing due to the translation/merge, i.e., missing data due to no data being included for this combination of identifiers in the end Output Unit.

Missing code: -22222 = No observation is merged/translated, but the original data contains information for these identifier combinations elsewhere. For these cases, we use a different code. The user needs to consult the reference documents (Methodology Document Section 5.1. or the Demscore Handbook) to clarify why the translation to the identifier combinations in the end Output Unit was not possible.

Unit Identifier Variables

An Output Unit is defined as an output format in which variables can be retrieved from one or more datasets through a strictly defined output grid. A unit table defining this output grid contains unit identifier columns with u_ prefixes and the table is sorted based on these unit identifier columns and has a fixed number of rows.

Unit columns are based on the columns that constitute the unit of analysis in a dataset. They are added to the original dataset and marked by a unit prefix (consisting of a "u_" and the dataset unit name) before the original variable name.

Unit columns can contain slightly modified data, e.g., missing values are replaced by a default value. Sometimes we add additional columns to the unit table, for instance, if a dataset includes both a country_id column with a numeric country code, we add the variable storing the full country name to the unit table as well for better readability.

Citations

The DEMSCORE project does not have a formal citation of its own. Hence, when using DEMSCORE, we encourage you to cite the respective projects and datasets.

We indicate how every dataset is to be cited in the autogenerated codebook you receive with your data download, both in the dataset description and the codebook entry for each variable.

Most often it is sufficient to cite the dataset a variable originates from, but sometimes there is a variable-specific citation listed in the codebook entry in addition. For these cases, please also add the variable-specific citation to the reference list of your publication. Full references are linked in the codebook entries of the variables and listed in the codebook's bibliography.

We also encourage you to cite the DEMSCORE Methodology Document when using data retrieved through DEMSCORE.

Download ID

The download ID allows the user to share the ID with other users for replication purposes.

A user can type the download ID into the download interface on demscore.se and retrieve the same download selection and files as the original user.

A unique download ID is autogenerated for each download from the DEMSCORE website and will always retrieve the same data, even if the DEMSCORE version was updated in the meantime.

You can find the downloader ID in your autogenerated customized codebook which is included in the zip file retrieved from the download interface.

4. INTEGRATING EXTERNAL DATA IN DEMSCORE

DEMSCORE is a platform that supports detailed cross-country analysis by providing tools to harmonize and merge diverse datasets. One of its strengths is its ability to integrate external datasets with its internal data, even when the datasets use different sets of identifiers or structures.

Using DEMSCORE to Merge External Data

In this example, Table 1 and Table 2 represent existing DEMSCORE datasets, while Table 3 is an external dataset that we aim to integrate.

Table 1: Project X, Dataset A provides a historical dataset spanning from 1900 to 2024 for countries such as Canada, Argentina, and Japan. This dataset uses Country and Year as identifiers to organize the data.

Table 1: Project X: Dataset A

Country	Year	var_1
CAN	1932	1
CAN	1999	2
ARG	1800	3
ARG	1801	4
JPN	1994	5
JPN	1997	6

This dataset serves as a baseline for integrating additional sources, using Country and Year to identify each observation.

Table 2: Project Y, Dataset A includes an additional identifier, Country_ID, alongside Country and Year.

Table 2: Project Y: Dataset A

Country	Country_ID	Year	var_2
CAN	123	1932	A
CAN	123	1999	B
ARG	456	1805	C
ARG	456	1806	D
JPN	789	1994	E
JPN	789	1997	F

At this stage, Country and Year are the main identifiers connecting Table 1 and Table 2. The inclusion of Country_ID becomes important when introducing our external data.

Integrating Data from External Sources

To illustrate how DEMSCORE integrates data from external sources, we use Table 3, which relies on Country_ID and Year as the primary keys. Since Table 2 already contains Country_ID, it acts as the bridge to merge the external data with DEMSCOREs internal datasets:

Table 3: External: Dataset A

Country_ID	Year	var
123	1932	x
123	1999	y
456	1800	z
456	1801	x
789	1994	y
789	1997	z

Here, Country_ID becomes the critical key that allows DEMSCORE to align data from Table 3 with its existing datasets, ensuring that the integration is accurate even when country names differ. The presence of Country in Table 1 and Table 2 facilitates initial alignment, while Country_ID enables a seamless merge with the external data.

Merged Dataset Example

By using Country, Country_ID and Year, DEMSCORE creates a unified dataset that brings together information from internal and external sources:

Table 4: Merged Dataset From Project X, Y and External Dataset

Country	Country_ID	Year	var_1	var_2	var
CAN	123	1932	1	A	x
CAN	123	1999	2	B	y
ARG	456	1800	3	-11111	z
ARG	456	1801	4	-11111	x
ARG	456	1805	-11111	C	-11111
ARG	456	1806	-11111	D	-11111
JPN	789	1994	5	E	y
JPN	789	1997	6	F	z

In this merged dataset:

- Country is the initial connector between Table 1 and Table 2.
- Country_ID allows for the inclusion of external data from Table 3.
- Year ensures temporal alignment, synchronizing observations across all three tables
- Any missing data for a specific Country-Year combination is marked as -11111.

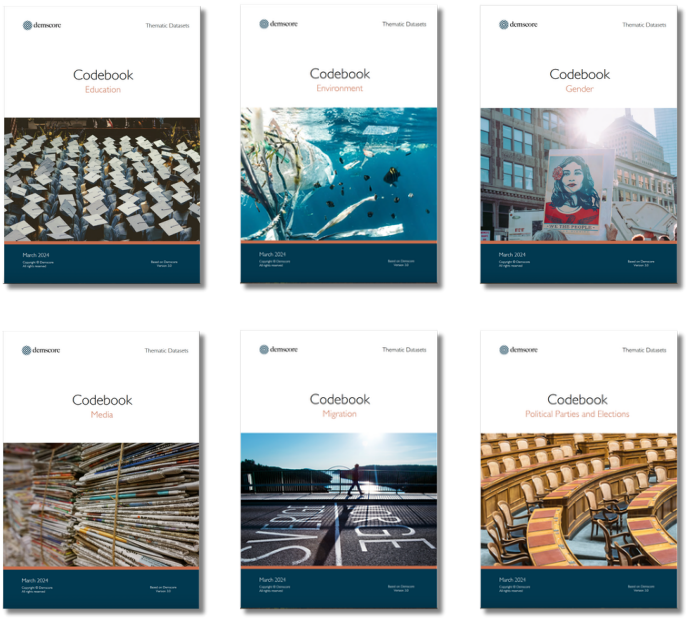
How DEMSCORE Handles Different Identifiers

DEMSCORE is a valuable tool for researchers who want to merge datasets, even if neither is a part of DEMSCORE's existing data. For instance, if you have two external datasets using different country identifiers - like Gleditsch and Ward (GW) codes in one and ISO codes in another - DEMSCORE can create a country merge table. In the DEMSCORE interface, you can select both the ISO and GW codes within any County-Year Output Unit. DEMSCORE then creates a merge table that aligns these identifiers, letting you combine variables from both datasets without manually handling discrepancies.



8. Additional tips

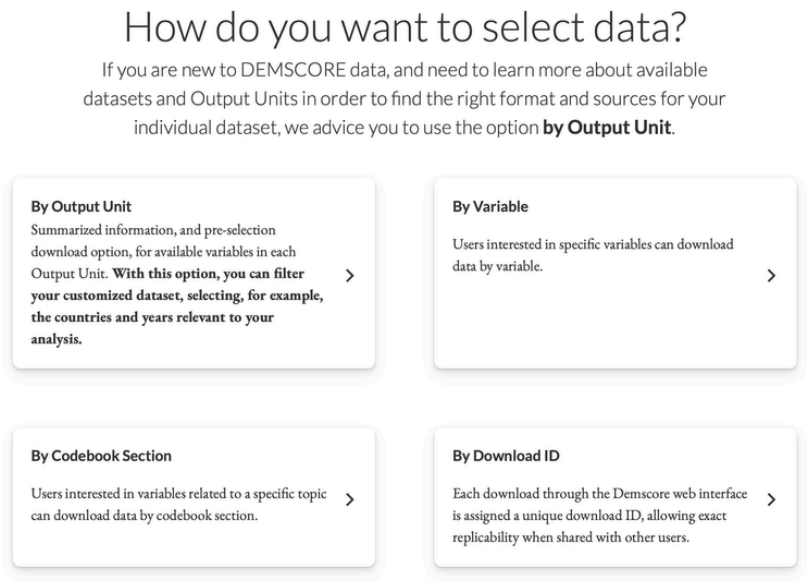
If you need inspiration for your variable selection, you can either browse the partnering projects' codebooks containing all their variables or our thematic datasets that group variables across all sources into topics.



More thematic datasets and additional user-friendly resources are available on our website: https://www.demscore.se

5. HOW TO USE THE DOWNLOAD INTERFACE

Demscore data can be retrieved from the download interface either by variable, by codebook section, by downloader ID, or by Output Unit. We recommend users to thoroughly read the Explanatory Notes section of the Methodology document before downloading data through Demscore.



Select how to retrieve data

Users interested in specific variables can download data by variable, and users interested in variables related to a specific topic can download data by codebook section. All variables from all datasets are thematically grouped into codebook sections in a PostgreSQL database. This organization enables the user to easily select all variables across all modules related to their topic of interest that are available in their Output Unit of Interest.

If the purpose of the download is to replicate data, users can download data by downloader ID. Each download through the DEMSCORE web interface is assigned a unique downloader ID, allowing exact replicability when shared with other users.

If you are new to DEMSCORE data and need to learn more about available datasets and Output Units to find the right format and sources for your customized dataset, we advise you to consult the list of Output Units here: https://www.demscore.se/output-unit-selection/.
The page includes information about Output Units available in DEMSCORE, and links to the download interface with the selected Output Unit filled in advance.

With the current setup, a user can generate a customized dataset and accompanying codebook in a matter of seconds. For example, creating a customized dataset with a tailored codebook encompassing 20 variables from ten datasets takes approximately 25 seconds, which is enormously time efficient compared to merging ten datasets by hand.

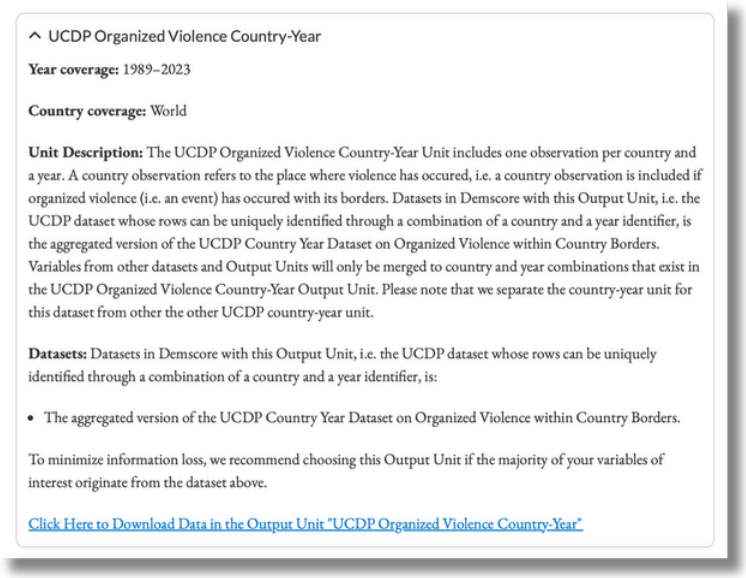
How to Select Data by Output Unit and Generate a Dataset

1. Select an Output Unit

Navigate to the Output Unit Selection Page. Here, you can select from various Output Unit options, such as:

- Country-Year
- Country-Region
- Cabinet and Party
- Dyad and Conflict
- Date and Event
- Predictions
- etc.

The Output Unit you select determines how your chosen variables will be merged, setting the key identifiers used to align and integrate the data accurately. Each Output unit includes information on year coverage, country coverage, and a description of the unit. For example, the Country-Year unit is ideal for examining country-level data over time. Let us take the UCDP Organized Violence Country-Year Output Unit as an example:



^ UCDP Organized Violence Country-Year

Year coverage: 1989–2023

Country coverage: World

Unit Description: The UCDP Organized Violence Country-Year Unit includes one observation per country and a year. A country observation refers to the place where violence has occured, i.e. a country observation is included if organized violence (i.e. an event) has occured with its borders. Datasets in Demscore with this Output Unit, i.e. the UCDP dataset whose rows can be uniquely identified through a combination of a country and a year identifier, is the aggregated version of the UCDP Country Year Dataset on Organized Violence within Country Borders. Variables from other datasets and Output Units will only be merged to country and year combinations that exist in the UCDP Organized Violence Country-Year Output Unit. Please note that we separate the country-year unit for this dataset from other the other UCDP country-year unit.

Datasets: Datasets in Demscore with this Output Unit, i.e. the UCDP dataset whose rows can be uniquely identified through a combination of a country and a year identifier, is:

- The aggregated version of the UCDP Country Year Dataset on Organized Violence within Country Borders.

To minimize information loss, we recommend choosing this Output Unit if the majority of your variables of interest originate from the dataset above.

Click Here to Download Data in the Output Unit "UCDP Organized Violence Country-Year"

If you are interested in studying organized violence and its relationship to political changes, for instance, you could select this Output Unit and merge it with democracy-related datasets for the same country-year combinations.

Once you have selected this Output Unit, click the "Click Here to Download Data in the Output Unit 'UCDP Organized Violence Country Year'" button.

Output Unit

An Output Unit, e.g., QoG Country-Year, is defined as an output format in which variables can be retrieved from one or more datasets through a strictly defined output grid. A unit table defining this output grid contains unit identifier columns with u_ prefixes and the table is sorted based on these unit identifier columns and has a fixed number of rows. An Output Unit has specific definitions for the level at which observations are presented, e.g., country definitions. For example, variables from a QoG dataset may have been collected under QoG country definitions, but in Demscore can also be retrieved through a V-Dem Output Unit which follows V-Dem country definitions.

2. Select a File Format

The next step is to choose a file format for your dataset. You can download the data in one of the following formats:

- R (.rds)
- STATA (.dta)
- CSV (.csv)

3. Customize your Dataset: Unit Columns, Empty Rows, and Countries

Now, you can customize your dataset by adjusting the following options:

Include Unit Columns

We recommend including Unit Columns, as they provide unique identifiers for each row in the dataset. These are helpful when merging or comparing datasets in the future.

Exclude Empty Rows

By default, all rows in the dataset are included, even those with missing data. However, you have the option to exclude empty rows, which can be useful if the variables you have selected have very few observations in your chosen Output Unit.

If you opt to exclude empty rows, any row that consists entirely of missing values—such as Demscore's default placeholder -11111 or true missing values like NA—will be removed. This can help you create a more streamlined dataset, focusing only on rows with meaningful data, which is especially useful when dealing with sparse variables.

However, keep in mind that excluding empty rows can affect your ability to easily merge or compare this dataset with others that use the same Output Unit. If the number of rows differs between datasets, you will need to rely on the unit identifiers (such as country or year) to perform a more complex merge later on.

Tip: Use this option when working with variables that have limited data points, but be cautious if you plan to merge with other datasets later.

Include All Countries

You can choose to include data for all countries or select specific ones. If you wish to focus on certain countries, uncheck the "Include all countries" box, and manually search for the countries you want to include. In this example, we will select all countries.



4. Select Variables

Next, select the variables for your dataset. Each variable is displayed with a label and its corresponding Demscore internal long tag (in parentheses), which indicates the dataset from which the variable originates.

For example, variables with the tag "ucdp_orgv_cy" come from the UCDP Country-Year Dataset on Organized Violence within Country Borders. Choose variables that align with your research focus.

In our example, we will select variables related to organized violence and democracy. Additionally, we will include control variables, such as GDP and population size, to account for socioeconomic factors.

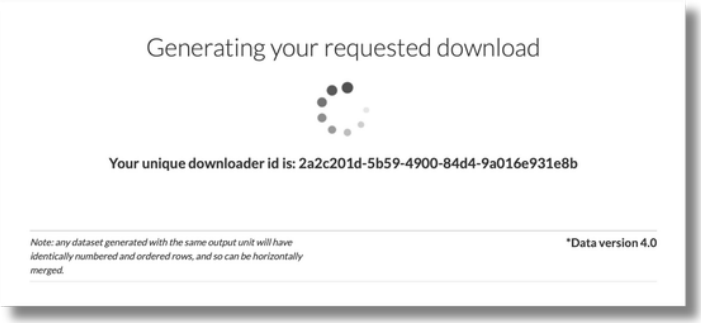
5. Select Year Range

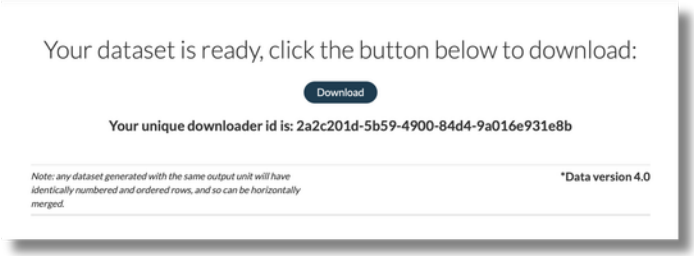
Now, choose the year range for your dataset. You can either select a specific time period or include the entire range of years available for your chosen Output Unit. In our example, we will select the year range from 1990 to 2022, as we want to investigate the post-Cold War period up to the present.



6. Generate Dataset

Once you have finalized your selections, click the 'Generate Dataset' button. Your dataset will be processed and made available for download, along with a customized codebook that provides details on the variables you selected.





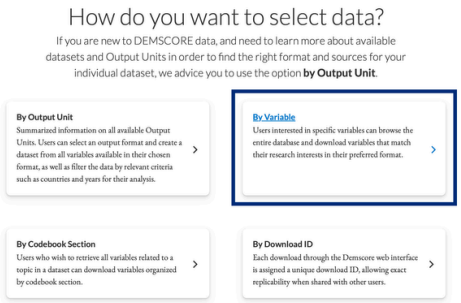


How to Select Data by Variable and Generate a Dataset

The “Download by Variable” option allows you to search for keywords across all variables and their metadata available in the DEMSCORE database. This feature is especially useful for finding variables that match your specific research focus without needing to browse each dataset or codebook individually.

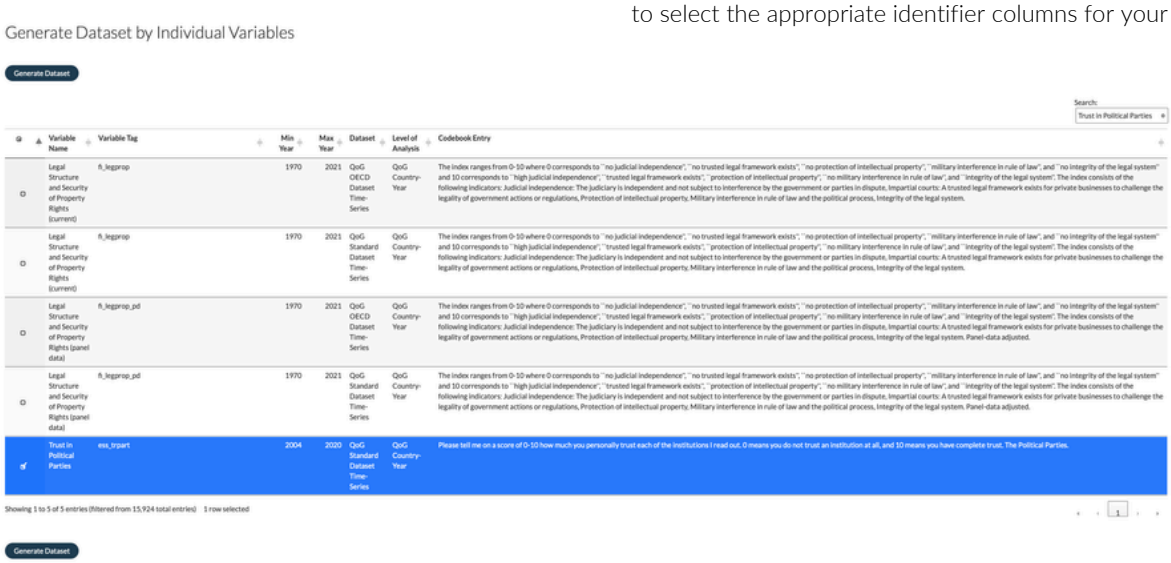
1. Select the “By Variable” Option in the Download Interface

To begin, select the “By Variable” option in the Download Interface.



2. Search for Keywords and Select Variables

When on the 'Generate Dataset by Individual Variables' page, search for keywords related to your chosen topic. In this example, we want to examine the relationship between democracy and trust in political parties, so we search for these terms and select relevant variables.



The table displays details about each variable, including the minimum and maximum years, dataset origin, the original level of analysis for the variable, and a codebook entry with more detailed information about what the variable measures.

After selecting all your variables, click the 'Generate Dataset' button located at either the top left corner or the bottom of the page.

3. Customize Your Dataset- Output Unit, File Format, Unit Columns, and Empty Rows

Output Unit

Choose an Output Unit based on the selected variables. The DEMSCORE Output Units determine the format in which you will retrieve your data. For example, 'COMPLAB Country-Year' means you will retrieve a dataset with one row per country and year, using COMPLAB country definitions and available year identifiers. The system will suggest only the output units in which all selected variables are available.

For this example, we will choose 'COMPLAB Country-Year.'

File Format

Select a file format for your dataset. Available options include:

- R (.rds)
- STATA (.dta)
- **CSV (.csv)**

Include Unit Columns

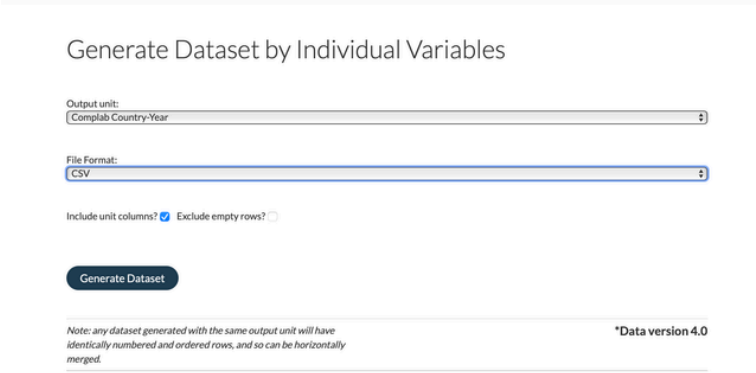
We recommend including Unit Columns to add unique identifiers for each row, which are useful when merging or comparing datasets later. If you choose not to include unit columns, be sure to select the appropriate identifier columns for your analysis.

Exclude Empty Rows

By default, all rows in the dataset are included, even those with missing data. However, you have the option to exclude empty rows, i.e. rows in which none of your selected variables matches the chosen output unit. If you choose to exclude empty rows, any row that consists entirely of missing values- such as Demscores default placeholder -11111 or true missing values like NA - will be removed. This can help you create a more streamlined dataset, focusing only on rows with meaningful data, which is especially useful when dealing with sparse variables.

However, keep in mind that excluding empty rows can affect your ability to easily merge or compare this dataset with others that use the same output unit. If the number of rows differ between datasets, you will need to rely on the unit identifiers (such as country or year) to perform a more complex merge later on.

Tips: Use this option when working with variables that have limited data points, but be cautious if you plan to merge with other datasets later.



6. Generate Dataset

Once you have finalized your selections, click the 'Generate Dataset' button. Your dataset will be processed and made available for download, along with a customized codebook containing details on the variables you selected.

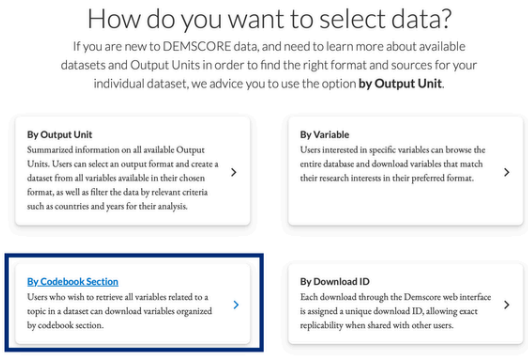


How to Select Data by Codebook Section and Generate a Dataset

The 'Download by Codebook Section' option allows users to browse and download variables grouped by thematic sections, making it easy to access data on specific topics across multiple datasets. This option is particularly useful for users interested in downloading all relevant variables related to a specific theme without searching each dataset or codebook individually.

1. Select the 'By Codebook Section' Option in the Download Interface

To begin, select the “By Codebook Section” option in the Download Interface.

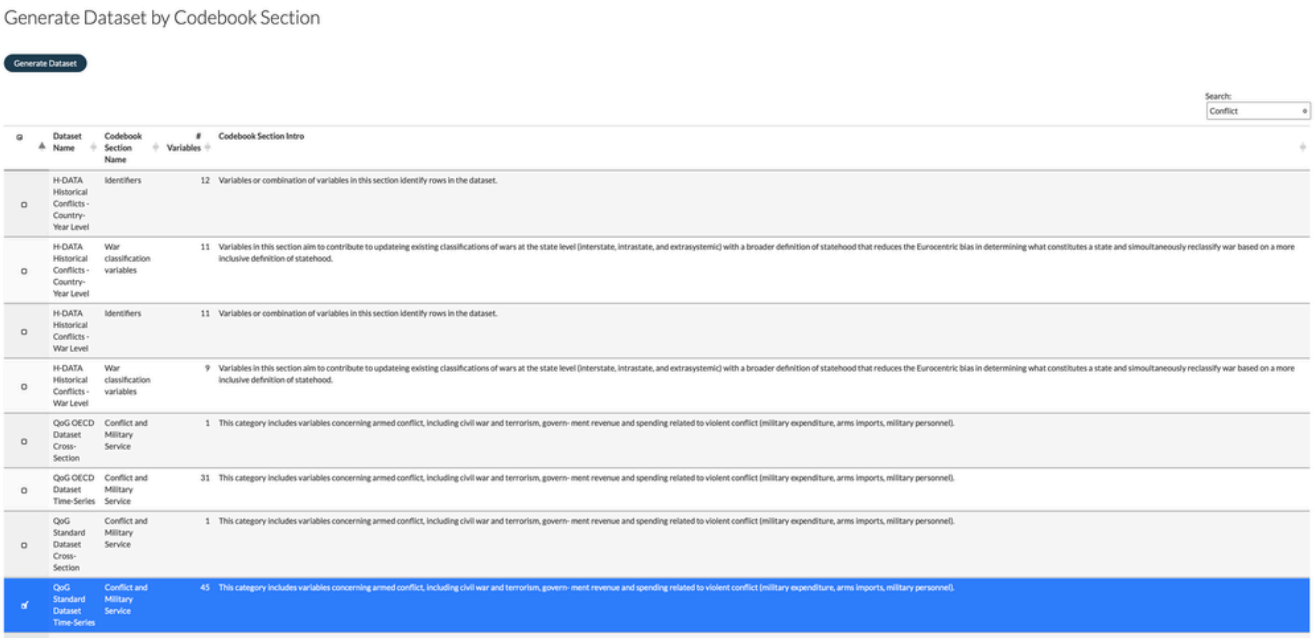


2. Search for Keywords and Select Codebook Section

You are now on the 'Generate Dataset by Codebook Section' page, where all variables are organized into thematic sections. This setup allows you to quickly explore topics of interest, drawing from all available datasets.

You can search for keywords related to your topic of interest within the codebook section intros. For example, if your research focuses on conflict, you can type in that keyword. In the table, you can view details about each codebook section, including the dataset name, codebook section name, the number of variables included in that section, and a codebook intro further describing the variables.

Once you have selected all the codebook sections of interest, click the 'Generate Dataset' button at the top left corner or at the bottom of the page.



3. Customize Your Dataset- Output Unit, File Format, Unit Columns, and Empty Rows

Output Unit

Choose an Output Unit based on the selected codebook section(s). The system will suggest only those output units in which all selected variables are available. For example, "QoG Country-Year" means that you will retrieve a dataset with one row per country and year, using QoG country definitions and available year identifiers.

For this example, we will select 'QoG Country-Year'.

File Format

Select a file format for your dataset. Available options include:

- R (.rds)
- STATA (.dta)
- CSV (.csv)

Include Unit Columns

We recommend including Unit Columns to add unique identifiers for each row, which are helpful when merging or comparing datasets later.

Exclude Empty Rows

By default, all rows in the dataset are included, even those with missing data. However, you have the option to exclude empty rows, i.e. rows in which none of your selected variables matches the chosen output unit.

If you choose to exclude empty rows, any row that consists entirely of missing values- such as Demscores default placeholder -11111 or true missing values like NA - will be removed. This can help you create a more streamlined dataset, focusing only on rows with meaningful data, which is especially useful when dealing with sparse variables.

However, keep in mind that excluding empty rows can affect your ability to easily merge or compare this dataset with others that use the same output unit. If the number of rows differs between datasets, you will need to rely on the unit identifiers (such as country or year) to perform a more complex merge later on.

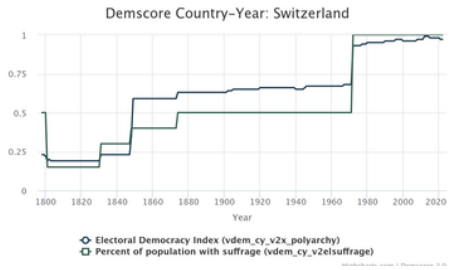
Tips: Use this option when working with variables that have limited data points, but be cautious if you plan to merge with other datasets later.

6. Generate Dataset

Once you have finalized your selections, click the 'Generate Dataset' button. Your dataset will be processed and made available for download, along with a customized codebook containing details on the variables you selected.

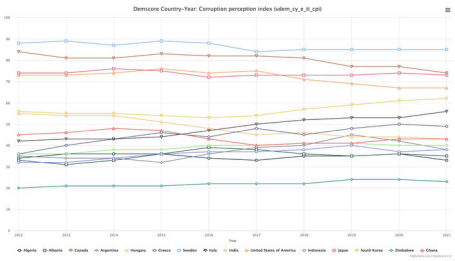


6. GRAPHING TOOLS



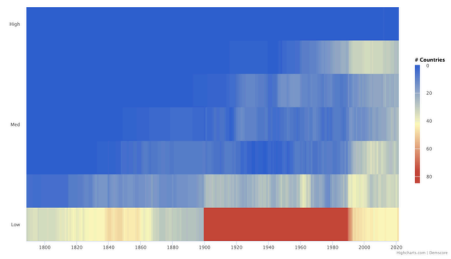
Country Graph

The Country Graph is a visual tool for comparing and analyzing multiple indicators or indices within a specific country.



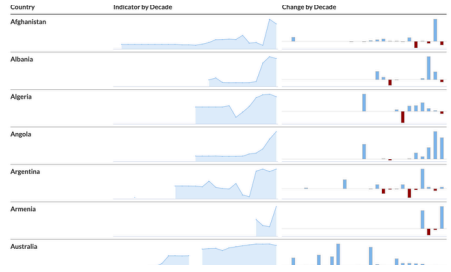
Variable Graph

The Variable Graph is a visual tool for analyzing and comparing trends in a specific indicator or index, across different countries.



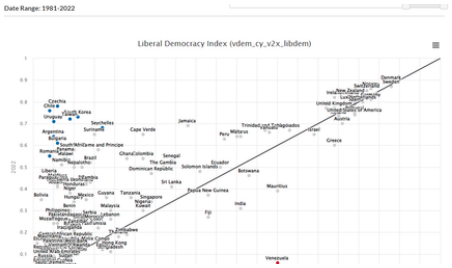
Heatmap

The Heatmap tool displays one indicator/index on a heatmap - a graphical representation of data where colors represent values.



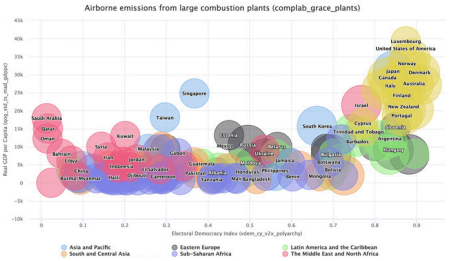
Sparkline Charts

A sparkline chart is a line chart used to depict the trend of a chosen indicator, such as a variable or index, over selected units of measurement, such as countries or regions.



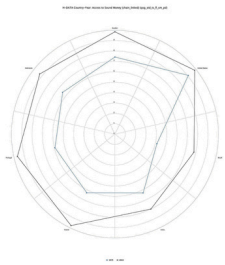
Comparison Graph

Creates individual line charts for each chosen indicator within a country. Each line on the chart corresponds to a distinct indicator, showcasing its trend or pattern over a given period.



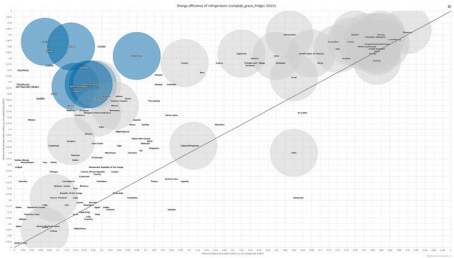
Bubble Chart

Displays data on a two-dimensional chart with radial lines. Each line represents a unit of measurement, and the data points are connected to form a shape, offering a quick visual comparison of multiple countries.



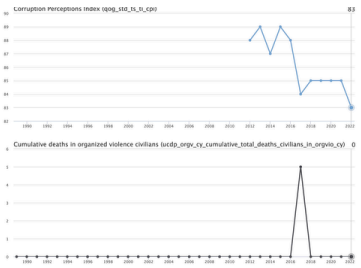
Radar Graph

Contrasts the current distribution of units of measurement over a specific indicator with the distributions from previous units of time. The size of the bubbles represents an additional variable.



Comparison Bubble Graph

A visual representation of three-dimensional data on a two-dimensional plane, using bubbles to depict data points. The X and Y axes represent two indicators, and the size of the bubbles represents a third indicator.



Multiple Variable Graph

Creates individual line charts for each chosen indicator within a country. Each line on the chart corresponds to a distinct indicator, showcasing its trend or pattern over a given period.

7. LIST OF DATASETS

All datasets are available for download through DEMSCORE

Name	Year Coverage	Analytical Level	Variables (N)	Rows (N)
COMPLAB GRACE - Governing the Anthropocene	1970-2022	Countries	33	1937
COMPLAB GRACE 3 - Governing the Anthropocene	1990-2020	Countries	26	1333
COMPLAB MIGPOL Antidiscrimination Dataset	2000-2018	Countries	11	608
COMPLAB MIGPOL Antidiscrimination Raw Data	2000-2018	Countries	68	608
COMPLAB MIGPOL DEMIG QuantMig	1721-2014	Changes in Migration Policies	14	9051
COMPLAB MIGPOL GLOBALCIT Country-Year	1960-2023	Countries	107	12600
COMPLAB MIGPOL Historical Immigration Policies Database	1789-2010	Countries	19	6236
COMPLAB MIGPOL IMISEM	2018-2024	Countries	828	61
COMPLAB MIGPOL IMPIC 2026	1980-2018	Countries	114	1279

Name	Year Coverage	Analytical Level	Variables (N)	Rows (N)
COMPLAB MIGPOL IMPIC Political Rights	1980-2010	Countries	15	1085
COMPLAB MIGPOL IMPIC Raw Data	1980-2010	Migration Tracks	278	13485
COMPLAB MIGPOL MIPEX	2007-2024	Countries	96	846
COMPLAB MIGPOL Migrant Social Policy Dataset	1980-2018	Countries	44	1521
COMPLAB SPIN The Child Benefit Dataset (CBD)	1960-2015	Countries	33	322
COMPLAB SPIN The Housing Benefit Dataset (HBEN)	2001-2020	Countries	123	718
COMPLAB SPIN The Out-of-Work Benefits Dataset (OUTWB)	2001-2022	Countries	76	785
COMPLAB SPIN The Parental Leave Benefit Dataset (PLB)	1950-2024	Countries	185	1094
COMPLAB SPIN The Social Assistance and Minimum Income Protection Interim Dataset (SAMIP)	1990-2023	Countries	44	1359
COMPLAB SPIN The Social Citizenship Indicator Program (SCIP)	1930-2005	Countries	196	288
COMPLAB SPIN The Social Insurance Entitlements Dataset (SIED)	2000-2024	Countries	196	925

Name	Year Coverage	Analytical Level	Variables (N)	Rows (N)
COMPLAB SPIN The Student Support and Fees Dataset (SSFD)	2005-2022	Countries	36	165
H-DATA Diplomatic Representation	1817-1914	Country Dyads	8	119438
H-DATA Foreign Minister Dataset	1777-2017	Foreign Ministers	62	1328
H-DATA Historical Cabinets	1919-1958	Cabinets	48	407
H-DATA Historical Conflicts - Country-Year Level	1816-1946	Countries	23	18983
H-DATA Historical Conflicts - War Level	1816-1946	Historical Wars	20	793
H-DATA Information Capacity Dataset	1750-2015	Countries	17	17350
H-DATA Leader Survival Dataset (PLT post-1789)	1789-2022	other	55	10622
QoG EQI Regional Level 2021 (with all NUTS2 regions)	2021	European Regions	15	238
QoG EU Regional Dataset Long Data	1960-2019	European Regions	337	22297
QoG EU Regional Dataset Wide Data (NUTS 1)	1960-2019	European Regions	626	7871
QoG EU Regional Dataset Wide Data (NUTS 2)	1960-2019	European Regions	632	16528
QoG Environmental Indicators Dataset	1946-2020	Countries	414	11722
QoG European Quality of Government Index CATI - Country Level (2010, 2013, 2017, 2021 & 2024)	2010-2024	Countries	33	114
QoG European Quality of Government Index Individual Level (2010 & 2013)	2010-2013	Individual Respondents	59	118707

Name	Year Coverage	Analytical Level	Variables (N)	Rows (N)
QoG European Quality of Government Index Individual Level (2017)	2017	Individual Respondents	67	77966
QoG European Quality of Government Index Individual Level (2021)	2021	Individual Respondents	67	129991
QoG European Quality of Government Index Individual Level (2024)	2024	other	73	135242
QoG European Quality of Government Index Regional Level (2010, 2013, 2017, 2021 & 2024)	2010-2024	European Regions	21	1050
QoG Expert Survey 2020	2020	Countries	29	117
QoG OECD Dataset Cross-Section		Countries	31	32
QoG OECD Dataset Time-Series	1946-2025	Countries	908	2272
QoG PERCEIVE Survey Dataset	2017	Individual Respondents	84	17147
QoG European Quality of Government Index Individual Level (2017)	2017	Individual Respondents	67	77966
QoG Politics, Institutions and Services in Swedish Municipalities	1980-2015	Municipalities	375	10440
QoG Standard Dataset Cross-Section		Countries	84	194
QoG Standard Dataset Time-Series	1946-2025	Countries	1810	12585
QoG Swedish Agency Database Budget Data	1971-2014	Swedish Agencies	17	7102
QoG Swedish Agency Database Formal Instruction Data	1960-2014	Swedish Agencies	43	1923
REPDEM Basic	1943-2026	Cabinets	407	1153
REPDEM Basic (Monthly)	1943-2026	Cabinets	409	24165

Name	Year Coverage	Analytical Level	Variables (N)	Rows (N)
REPDEM Basic (Quarterly)	1943-2026	Cabinets	409	8754
REPDEM Basic (Yearly)	1943-2026	Cabinets	408	3122
REPDEM Election Dataset	1942-2026	Elections	21	5084
REPDEM Election Dataset - Alliances	1945-2026	Elections	16	523
REPDEM PAGED Party Dataset (WE+CEE)	1944-2021	Cabinets	28	6379
REPDEM PAGED Potential Governments Dataset (WE+CEE)	1944-2021	Cabinets	62	1564356
REPDEM PAGED Western, Central and Eastern Europe	1944-2021	Cabinets	704	772
REPDEM PAGED Western, Central and Eastern Europe (Monthly)	1944-2021	Cabinets	705	16955
REPDEM PAGED Western, Central and Eastern Europe (Quarterly)	1944-2021	Cabinets	705	6394
REPDEM PAGED Western, Central and Eastern Europe (Yearly)	1944-2021	Cabinets	705	2168
REPDEM Party Dataset (Basic)	1944-2026	Cabinets	45	9095
REPDEM Potential Governments Dataset (Basic)	1944-2026	Cabinets	66	7952140
UCDP Actor Dataset Version 26.1	1989-2025	Actors in Violent Conflict	35	1928
UCDP Battle-Related Deaths Dataset, Conflict Level Version 26.1	1989-2025	Violent Conflicts	25	1650
UCDP Battle-Related Deaths Dataset, Dyadic Level Version 26.1	1989-2025	Conflict Dyads	25	2081

Name	Year Coverage	Analytical Level	Variables (N)	Rows (N)
UCDP Cities and Armed Conflict Events (CACE)	1989-2017	other	42	142902
UCDP Conflict Issues Dataset Version 23.2 (Dyad-Issue-Year)	1989-2017	Conflict Issues	35	14711
UCDP Conflict Issues Dataset Version 23.2 (Dyad-Year)	1989-2017	Conflict Dyads	143	1786
UCDP Conflict Termination Dataset, Conflict Level Version 4-2024	1946-2023	Violent Conflicts	36	2752
UCDP Conflict Termination Dataset, Dyadic Level Version 4-2024	1946-2023	Conflict Dyads	38	3432
UCDP Country Year Dataset on Organized Violence within Country Borders version 26.1	1989-2025	Countries	74	7132
UCDP Deadly Electoral Conflict dataset (DECO)	1989-2017	other	41	4233
UCDP Dyadic Dataset Version 26.1	1946-2025	Conflict Dyads	74	7132
UCDP External Support Dataset - Actor Year	1975-2017	External Support in Violent Conflict	97	4781
UCDP External Support Dataset - Dyad Year	1975-2017	Conflict Dyads	96	2272
UCDP External Support Dataset - Triad Year	1975-2017	External Support in Violent Conflict	38	10852
UCDP External Support in Non-state Conflict Dataset	1989-2011	Conflict Dyads	17	468
UCDP Georeferenced Event Dataset (GED) Version 26.1	1989-2025	Violent Events	49	417968
UCDP Interstate Conflict Level Onset Dataset v.25.1	1946-2024	Intrastate Conflicts	13	12558
UCDP Interstate Country Level Multiple Onset Dataset v.25.1	1946-2024	Intrastate Conflicts	13	12558

Name	Year Coverage	Analytical Level	Variables (N)	Rows (N)
UCDP Interstate Country Level Onset Dataset v.25.1	1946-2024	Countries	13	12553
UCDP Intrastate Conflict Level Onset Dataset v.25.1	1946-2024	Intrastate Conflicts	13	12606
UCDP Intrastate Country Level Multiple Onset Dataset v.25.1	1946-2024	Intrastate Conflicts	13	12606
UCDP Intrastate Country Level Onset Dataset v.25.1	1946-2024	Countries	13	12553
UCDP Non-State Conflict Dataset Version 26.1	1989-2025	Violent Conflicts	32	1821
UCDP Non-state Conflict Issues and Actors Dataset	1989-2011	Conflict Dyads	27	468
UCDP One-sided Violence Dataset Version 26.1	1989-2025	Conflict Dyads	17	1408
UCDP Peace Agreement Dataset Version 22.1	1975-2021	Peace Agreements	67	374
UCDP Peacemakers at Risk (PAR) Dataset version 1.0-2016	1989-2009	other	41	935
UCDP The Ethnic One-Sided Violence (EOSV) Dataset	1989-2013	other	51	831
UCDP Violent Political Protest Dataset Version 26.1	1989-2025	Conflict Dyads	15	93
UCDP/PRIO Armed Conflict Dataset Version 26.1	1946-2025	Violent Conflicts	28	2752
V-Dem Coder-Level v16	1789-2025	Country Expert Coders	1314	824711
V-Dem Country-Date v16	1789-2025	Countries	3953	78878
V-Dem Country-Year: V-Dem Full+Others v16	1789-2025	Countries	4593	28092
V-Dem Episodes of Regime Transformation Dataset	1900-2025	Countries	41	19857

Name	Year Coverage	Analytical Level	Variables (N)	Rows (N)
V-Dem V-Party Coder Level v2	1965-2019	Country Expert Coders	155	30636
V-Dem V-Party v2	1900-2019	Political Parties	384	11898
VIEWS Country-Month Conflict Predictions (Input Data: January 2022 - March 2023)	2022-2025	Countries	9	6876
VIEWS Country-Month Conflict Predictions (Input Data: April 2023 - August 2026)	2023-2029	Countries	10	6876
VIEWS PRIO-GRID-Month Conflict Predictions (Input Data: January 2022 - March 2023)	2022-2025	PRIO-GRID Cells	4	471960
VIEWS PRIO-GRID-Month Conflict Predictions (Input Data: April 2023 - August 2026)	2023-2029	PRIO-GRID Cells	5	471960

8. LIST OF OUTPUT UNITS

The Demscore approach allows users to select an Output Unit and access variables from different datasets, ensuring variables are transformed to match the chosen unit, either directly or through aggregation/disaggregation while maintaining information from the original unit identifiers, such as country names or government, and matching them as well as possible.

Download data by Output Units here: www.demscore.se/data/output-unit-selection/

Country-year format

DEMSCORE Country-Year

Year coverage: 1750-2025
Country coverage: World
Unit description: This Output Unit includes one observation per country and year and covers the full country-year range in Demscore. Between 1789 2023, it uses V-Dem Country names and definitions; the period prior to 1789 is covered by H-DATA country names. Countries that do not exist in V-Dem are filled in with Gleditsch and Ward country names. This is the case for the Bahamas, Belize, Brunei, East Timor, Gambia, Macedonia, Surinam, and Tibet.

The Demscore country-year unit offers a convenient way to explore the data, available observations, and variable translations, as it covers the full country-year range available in Demscore. Users should be aware that the Demscore unit may result in some data quality loss compared to the Module-specific units due to country merges. Therefore, we recommend users use the most suitable Module-specific unit for their analyses and use the Demscore country-year unit predominantly to explore the data.

V-Dem Country-Year

Year coverage: 1789–2025
Country coverage: World
Unit description: This Output Unit includes one observation per country and year. Choosing this Output Unit thus means that selected variables get merged based on country and year identifiers used in V-Dem. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss. Variables from other datasets and Output Units will only be merged to country and year combinations that exist in the V-Dem Country-Year Output Unit.

QoG Country-Year

Year coverage: 1946–2025
Country coverage: World
Unit description: This Output Unit includes one observation per country and year. Choosing this Output Unit thus means that selected variables get merged based on country and year identifiers used in QoG. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss. Variables from other datasets and Output Units will only be merged to country and year combinations that exist in the QoG Country-Year Output Unit.

UCDP Organized Violence Country-Year

Year coverage: 1989–2024
Country coverage: World
Unit description: The UCDP Organized Violence Country-Year Unit includes one observation per country and a year. A country observation refers to the place where violence has occurred, i.e. a country observation is included if organized violence (i.e. an event) has occurred within its borders. Datasets in Demscore with this Output Unit, i.e. the UCDP dataset whose rows can be uniquely identified through a combination of a country and a year identifier, is the aggregated version of the UCDP Country Year Dataset on Organized Violence within Country Borders. Variables from other datasets and Output Units will only be merged to country and year combinations that exist in the UCDP Organized Violence Country-Year Output Unit. Please note that we separate the country-year unit for this dataset from other the other UCDP country-year unit.

COMPLAB Country-Year
Year coverage: 1783–2024
Country coverage: World
Unit description: This Output Unit includes one observation per country and year. Choosing this Output Unit thus means that selected variables get merged based on country and year identifiers used in COMPLAB. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss. Variables from other datasets and Output Units will only be merged to country and year combinations that exist in the Complab Country-Year Output Unit.

REPDEM Country-Year
Year coverage: 1946–2025
Country coverage: Europe
Unit description: The REPDEM Country-Year Unit includes one observation per country and year. This unit is not an original unit of any REPDEM dataset but was created from an aggregation of the REPDEM PAGED datasets available on a Cabinet-Date level. The unit thus includes one observation per country and year covered by REPDEM data. Variables from other datasets and Output Units will only be merged to country and year combinations that exist in this Output Unit.

H-DATA Country-Year
Year coverage: 1750–2017
Country coverage: World
Unit description: This Output Unit includes one observation per country and year. Choosing this Output Unit thus means that selected variables get merged based on country and year identifiers used in H-DATA. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss. Variables from other datasets and Output Units will only be merged to country and year combinations that exist in the H-DATA Country-Year Output Unit.

Country/Regional format

QoG Agency-Fiscal Year
Year coverage: 1971–2014
Country coverage: Sweden
Unit description: This Output Unit includes one observation per Swedish agency and fiscal year. Choosing this Output Unit thus means that selected variables get merged based on agency and year identifiers used in QoG. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss. Variables from other datasets and Output Units will only be merged to agency and year combinations that exist in the QoG Agency-Fiscal Year Output Unit.

QoG Country
Year coverage: Data from around 2020 is included.
Country coverage: World
Unit description: This Output Unit includes one observation per country. Choosing this Output Unit thus means that selected variables get merged based on country identifiers used in QoG. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss. Variables from other datasets and Output Units will only be merged to countries that exist in the QoG Country Output Unit.

QoG Municipality-Year
Year coverage: 1980–2015
Country coverage: Sweden
Unit description: This Output Unit includes one observation per Swedish municipality and year. Choosing this Output Unit thus means that selected variables get merged based on municipality and year identifiers used in QoG. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss. Variables from other datasets and Output Units will only be merged to municipality and year combinations that exist in the QoG Municipality-Year Output Unit.

QoG NUTS Region

Year coverage: –
Country coverage: EU Member States
Unit description: This Output Unit includes one observation per region. Choosing this Output Unit thus means that selected variables get merged based on region identifiers used in QoG. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss. Variables from other datasets and Output Units will only be merged to regions that exist in the QoG Region Output Unit.

QoG NUTS Region-Year

Year coverage: 1960–2024
Country coverage: EU Member States
Unit description: This Output Unit includes one observation per region and year. Choosing this Output Unit thus means that selected variables get merged based on region and year identifiers used in QoG. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss. Variables from other datasets and Output Units will only be merged to region and year combinations that exist in the QoG NUTS Region-Year Output Unit.

Cabinet and Party format

V-Dem Party-Country-Year

Year coverage: 1900–2019
Country coverage: World
Unit description: This Output Unit includes one observation per party, country and year. Choosing this Output Unit thus means that selected variables get merged based on party, country, and year identifiers used in V-Dem. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss. Variables from other datasets and Output Units will only be merged to party, country, and year combinations that exist in the V-Dem Party-Country-Year Output Unit.

REPDEM Cabinet-Date

Year coverage: 1944–2025
Country coverage: Europe
Unit description: This Output Unit includes one observation per cabinet and its start date. Choosing this Output Unit thus means that selected variables get merged to a cabinet and a date. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss.

REPDEM Cabinet-Month

Year coverage: 1943–2025
Country coverage: Europe
Unit description: The REPDEM Cabinet-Month Unit includes one observation per cabinet and month this cabinet was in power. Variables from other datasets and Output Units will only be merged to country and year combinations that exist in this Output Unit.

REPDEM Cabinet-Quarter

Year coverage: 1943–2025
Country coverage: Europe
Unit description: REPDEM Cabinet-Quarter Unit includes one observation per cabinet and quarter of a year. Variables from other datasets and Output Units will only be merged to country and year combinations that exist in this Output Unit.

REPDEM Cabinet-Party

Year coverage: 1942–2025
Country coverage: Europe
Unit description: This Output Unit includes one observation per cabinet and party. Variables from other datasets and Output Units will only be merged to country and year combinations that exist in this Output Unit.

REPDEM Cabinet-Potential Governments

Year coverage: 1943–2025
Country coverage: Europe
Unit description: This Output Unit includes one observation per cabinet and potential coalition partners. Variables from other datasets and Output Units will only be merged to country and year combinations that exist in this Output Unit.

REPDEM Cabinet-Year

Year coverage: 1943–2025
Country coverage: Europe
Unit description: This Output Unit includes one observation per cabinet and year. Variables from other datasets and Output Units will only be merged to country and year combinations that exist in this Output Unit.

H-Data Cabinet-Date

Year coverage: 1919–1953
Country coverage: Europe
Unit description: This Output Unit includes observations per cabinet and the date they got into office. Choosing this Output Unit thus means that selected variables get merged to a cabinet and a date. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss.

Dyad and Conflict format

UCDP Conflict-Year

Year coverage: 1946–2024
Country coverage: World
Unit description: This Output Unit includes one observation per conflict and year. Choosing this Output Unit thus means that selected variables get merged based on conflict and year identifiers used in UCDP. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss. Variables from other datasets and Output Units will only be merged to conflict and year combinations that exist in the UCDP Conflict-Year Output Unit.

UCDP Conflict-Location-Year

Year coverage: 1946–2023
Country coverage: World
Unit description: This Output Unit includes one observation per conflict, location and year. Choosing this Output Unit thus means that selected variables get merged based on conflict, location (Gleditsch and Ward) and year identifiers used in UCDP. To create this unit, we stretch the dataset using the comma-separated observations in the gwno column, i.e. creating one row per location, year and conflict. Variables from other datasets and Output Units will only be merged to location, conflict, and year combinations that exist in the UCDP Dyad-Location-Year Output Unit. This unit allows users to combine variables collected on both conflict-year and e.g., country-year level in one dataset.

UCDP Dyad-Year

Year coverage: 1946–2024
Country coverage: World
Unit description: This Output Unit includes one observation per dyad and year. Choosing this Output Unit thus means that selected variables get merged based on dyad and year identifiers used in UCDP. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss. Variables from other datasets and Output Units will only be merged to conflict and year combinations that exist in the UCDP Dyad-Year Output Unit.

UCDP Dyad-Issue-Year

Year coverage: 1989–2023
Country coverage: World
Unit description: This Output Unit includes one observation per dyad, conflict issue, and year. Choosing this Output Unit thus means that selected variables get merged based on dyad, issue, and year identifiers used in UCDP. Variables from other datasets and Output Units will only be merged to location, conflict, and year combinations that exist in the UCDP Dyad-Issue-Year Output Unit.

UCDP Dyad-Location-Year

Year coverage: 1946–2024
Country coverage: World
Unit description: This Output Unit includes one observation per dyad, location and year. Choosing this Output Unit thus means that selected variables get merged based on dyad, location (Gleditsch and Ward) and year identifiers used in UCDP. To create this unit we stretch the dataset using the comma-separated observations in the gwno column, i.e. creating one row per location, year and dyad. Variables from other datasets and Output Units will only be merged to location, dyad, and year combinations that exist in the UCDP Dyad-Location-Year Output Unit.

H-DATA Dyad-Year

Year coverage: 1817–1914
Country coverage: World
Unit description: The unit for this dataset is a dyad and a year. The dyad consist of two country variables, with country 2 being the country that is diplomatically represented by country 1 in a given year. Choosing this Output Unit thus means that selected variables get merged based on country and year identifiers used in H-DATA. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss. Variables from other datasets and Output Units will only be merged to country dyad and year combinations that exist in the H-DATA Dyad-Year Output Unit.

H-DATA Country-Year-War

Year coverage: 1816–1945
Country coverage: World
Unit description: This Output Unit includes one observation per country, year and war. Choosing this Output Unit thus means that selected variables get merged based on country, war, and year identifiers used in H-DATA. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss. Variables from other datasets and Output Units will only be merged to country, war, and year combinations that exist in the H-DATA Country-Year-War Output Unit.

Respondent-Level Format

V-Dem Country-Date-Coder

Year coverage: 1789–2025
Country coverage: World
Unit description: This Output Unit includes one observation per coder, country and date. Choosing this Output Unit thus means that selected variables get merged based on coder, country, and year identifiers used in V-Dem. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss. Variables from other datasets and Output Units will only be merged to coder, country, and date combinations that exist in the V-Dem Country-Date Output Unit.

V-Dem Party-Date-Coder

Year coverage: 1900–2019
Country coverage: World
Unit description: This Output Unit includes one observation per party, date and coder. Choosing this Output Unit thus means that selected variables get merged based on party, date, and coder identifiers used in V-Dem. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss. Variables from other datasets and Output Units will only be merged to party, date, and coder combinations that exist in the V-Dem Party-Date-Coder Output Unit.

QoG EQI Respondent ID 2010–2013
Year coverage: 2010–2013
Country coverage: Europe
Unit description: This Output Unit includes one observation per respondent in the 2010 to 2013 round of the EQI survey data. Choosing this Output Unit thus means that selected variables get merged based on coder identifiers used in QoG. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss. Variables from other datasets and Output Units will only be merged coder identifiers that exist in this Output Unit.

QoG EQI Respondent ID 2017
Year coverage: 2017
Country coverage: Europe
Unit description: This Output Unit includes one observation per respondent in the 2017 round of the EQI survey data.

QoG EQI Perceptions Respondent ID 2021
Year coverage: 2021
Country coverage: Europe
Unit description: This Output Unit includes one observation per respondent in the 2021 round of the EQI survey data.

QoG EQI Perceptions Respondent ID 2024
Year coverage: 2024
Country coverage: Europe
Unit description: This Output Unit includes one observation per respondent in the 2024 round of the EQI survey data.

Date and Event Format

H-DATA Leader-Date
Year coverage: 1789–2022
Country coverage: World
Unit description: This Output Unit includes observations per leader and the date they got into office. Variables from other datasets and Output Units will only be merged to leader and date combinations that exist in the H-DATA Leader-Date Output Unit.

H-DATA Minister-Date
Year coverage: 1789–2017
Country coverage: 13 great powers: Austria (the Habsburg Empire/Austria-Hungary), Britain, China (Qing Empire/Republic/People's Republic of China), France, Italy, Japan, the Netherlands, Prussia
Unit description: This Output Unit includes observations per foreign minister and the date they got into office. Variables from other datasets and Output Units will only be merged to minister and date combinations that exist in the H-DATA Minister-Date Output Unit.

UCDP Event-ID
Year coverage: 1989–2022
Country coverage: World
Unit description: The UCDP Event ID Unit includes one observation per violent event. Variables from other datasets and Output Units will only be merged to events that exist in the UCDP Event ID Output Unit.

UCDP Event-ID (Separated IDs)
Year coverage: 1989–2024
Country coverage: World
Unit description: The UCDP Event ID (Separated IDs) Unit includes one observation per violent event in the CACE and DECO datasets.

UCDP Peacemakers-at-Risk Event ID
Year coverage: –
Country coverage: –
Unit description: This Output Unit includes one observation per event. The Event IDs are not similar to the IDs in UCDP GED, hence, they are kept separately in their own unit.

V-Dem Country-Date
Year coverage: 1789–2025
Country coverage: World
Unit description: This Output Unit includes one observation per country and date. Choosing this Output Unit thus means that selected variables get merged based on country and year identifiers used in V-Dem. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss. Variables from other datasets and Output Units will only be merged to country and date combinations that exist in the V-Dem Country-Date Output Unit.

Predictions

VIEWS Country-Month
Year coverage: 2022–2029
Country coverage: World
Unit description: This Output Unit includes one prediction per PRIO-GRID cell and month. Choosing this Output Unit thus means that selected variables get merged based on PRIO-GRID cell and month identifiers used in VIEWS. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss. Variables from other datasets and Output Units will only be merged to PRIO-GRID cell and month combinations that exist in the VIEWS Country-Month Output Unit.

VIEWS PRIO-GRID-Month
Year coverage: 2022–2029
Country coverage: World
Unit description: This Output Unit includes one prediction per PRIO-GRID cell and month. Choosing this Output Unit thus means that selected variables get merged based on PRIO-GRID cell and month identifiers used in VIEWS. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss. Variables from other datasets and Output Units will only be merged to PRIO-GRID cell and month combinations that exist in the VIEWS Country-Month Output Unit.

Other

QoG Agency-Agency Instruction
Year coverage: 1960–2014
Country coverage: World
Unit description: This Output Unit includes observations per agency and agency instruction number. Choosing this Output Unit thus means that selected variables get merged based on agency and agency instruction identifiers used in QoG. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss. Variables from other datasets and Output Units will only be merged to agency and agency instruction combinations that exist in the QoG Agency-Agency Instruction Output Unit.

UCDP Actor
Year coverage: –
Country coverage: World
Unit description: This Output Unit includes one observation per actor. Choosing this Output Unit thus means that selected variables get merged based on actor identifiers used in UCDP. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss. Variables from other datasets and Output Units will only be merged actor identifiers that exist in this Output Unit.

UCDP Actor-Dyad-Year
Year coverage: 1975–2017
Country coverage: World
Unit description: This Output Unit includes one observation per external supporter, actor, dyad and year. No data is translated from or to this unit, hence variables from this dataset are only available in their original form in Demscore. Another version of this dataset, UCDP ESD DY is available in the UCDP Dyad-Year Output Unit. Please choose this unit and dataset if you want to retrieve UCDP External Support Data in other Output Units.

UCDP Peace Agreement

Year coverage: 1975–2021
Country coverage: World
Unit description: Choosing this Output Unit thus means that selected variables get merged based on peace agreement identifiers used in UCDP. Variables from other datasets and Output Units will only be merged to peace agreements that exist in the UCDP Peace Agreement Output Unit.

UCDP Peace Agreement-Conflict-Year

Year coverage: 1975–2021
Country coverage: World
Unit description: This Output Unit includes one observation per peace agreement, conflict and year. Choosing this Output Unit thus means that selected variables get merged based on peace agreement, conflict and year identifiers used in UCDP. To create this unit we stretch the dataset using the comma-separated observations in the conflict_id column, i.e. creating one row per conflict_id, year and peace agreement. Variables from other datasets and Output Units will only be merged to peace agreement, conflict, and year combinations that exist in the UCDP Peace Agreement-Conflict-Year Output Unit.

UCDP Peace Agreement-Country-Year

Year coverage: 1975–2021
Country coverage: World
Unit description: This Output Unit includes one observation per peace agreement, country and year. Choosing this Output Unit thus means that selected variables get merged based on peace agreement, country (Gleditsch and Ward) and year identifiers used in UCDP. To create this unit we stretch the dataset using the comma-separated observations in the gwno column, i.e. creating one row per country, year and peace agreement. Variables from other datasets and Output Units will only be merged to peace agreement, country, and year combinations that exist in the UCDP Peace Agreement-Country-Year Output Unit.

UCDP Peace Agreement-Dyad-Year

Year coverage: 1975–2021
Country coverage: World
Unit description: This Output Unit includes one observation per peace agreement, dyad and year. To create this unit we stretch the dataset using the comma-separated observations in the dyad_id column, i.e. creating one row per dyad_id.

UCDP Triad-Year

Year coverage: 1975–2017
Country coverage: World
Unit description: This Output Unit includes one observation per external supporter, actor, dyad and year. No data is translated from or to this unit, hence variables from this dataset are only available in their original form in Demscore. Another version of this dataset, UCDP ESD DY is available in the UCDP Dyad-Year Output Unit. Please choose this unit and dataset if you want to retrieve UCDP External Support Data in other Output Units.

COMPLAB Country-Year-Change

Year coverage: 1721-2020
Country coverage: World
Unit description: This Output Unit includes one observation per country, year and migration policy. Choosing this Output Unit thus means that selected variables get merged based on migration policy, country and year identifiers used in COMPLAB. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss. Variables from other datasets and Output Units will only be merged to migration policy, country and year combinations that exist in the Complab Country-Year-Change Output Unit.

COMPLAB Country-Year-Track

Year coverage: 1980-2010
Country coverage: World
Unit description: This Output Unit includes one observation per migration track, country and year. Choosing this Output Unit thus means that selected variables get merged based on migration track country and year identifiers used in COMPLAB. Variables from datasets with this Output Unit can be retrieved through Demscore in their original form and without any information loss. Variables from other datasets and Output Units will only be merged to migration track, country and year combinations that exist in the Complab Country-Year-Track Output Unit.

demscore

50+
OUTPUT UNITS

145+
DATASETS

1,000+
MERGE OPTIONS

26,000+
VARIABLES

Demscore, acronym for research Infrastructure for Democracy, Environment, Migration, Social policy, Conflict, and Representation, offers harmonized world-leading Social Science Data. The data is open source and free to use worldwide.

A fully normalized, joint PostgreSQL database, sophisticated programming, and a web-based interface enable the selection of variables from all six partnering data infrastructures. This results in a custom-designed dataset and codebook generated automatically in a matter of seconds, making Demscore an enormously time-efficient resource compared to merging several datasets manually.

Explore our Thematic Datasets

Demscore's Thematic Datasets assemble a wide range of variables from datasets across all partnering Modules related to a selected topic.

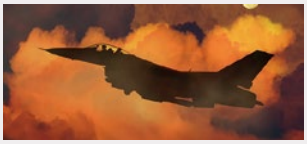


Environment | Political Parties and Election | Corruption | Security and Violence | Education

For additional user-friendly resources, including more thematic datasets, online visualization tools, static datasets, and a quick-start handbook, please visit www.demscore.se



Varieties of Democracy (V-Dem)



Uppsala Conflict Data Program and Violence Early-Warning System (UCDP/VIEWS)



Quality of Government (QoG)



Comparative Policy Laboratory (COMPLAB)



Representative Democracy Data Archive (REPDEM)



Historical Data Archive (H-DATA)

DEMSCORE IS FUNDED BY

Swedish Research Council
UNIVERSITY OF GOTHENBURG
UPPSALA UNIVERSITET
Stockholms universitet
UMEÅ UNIVERSITY