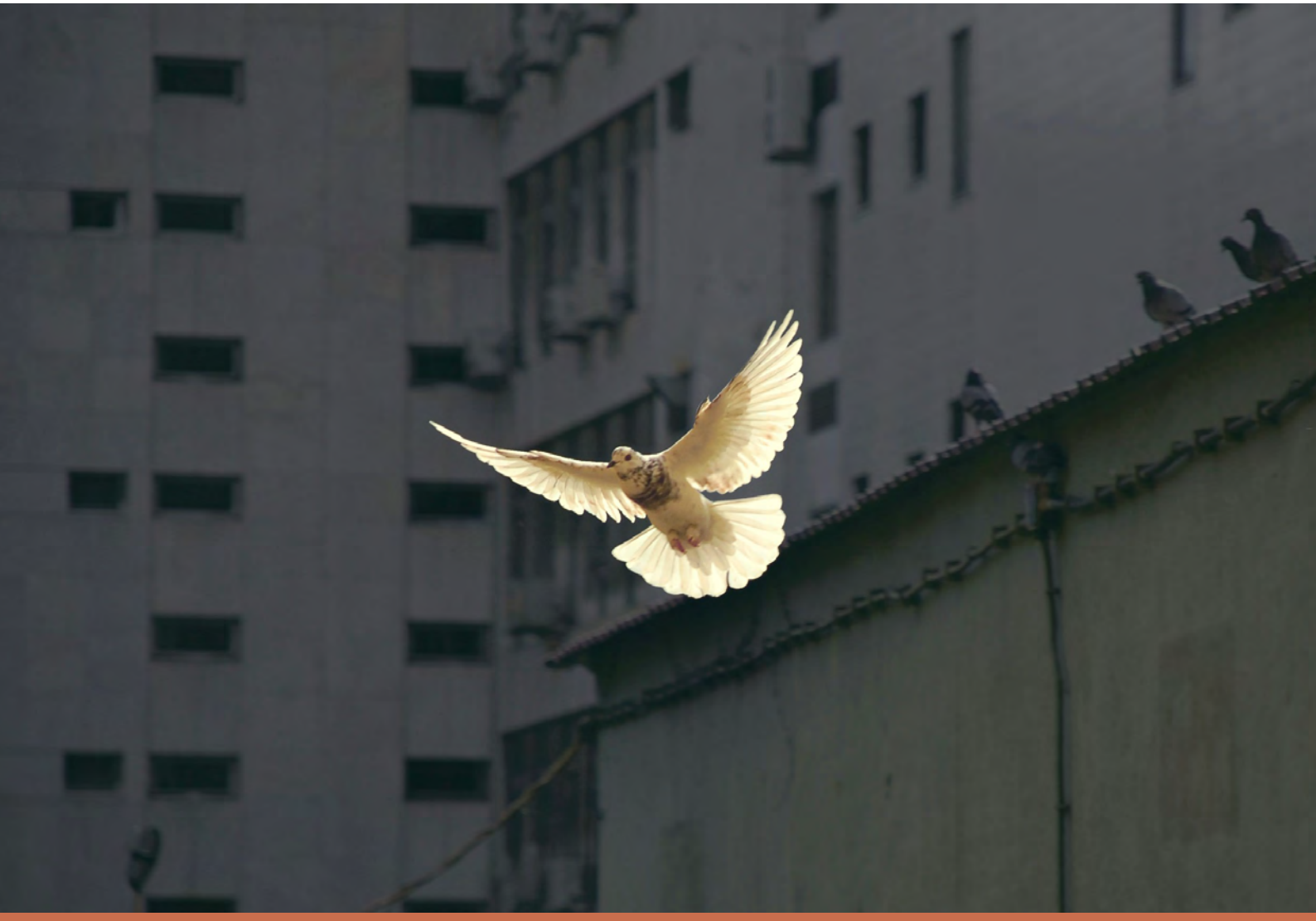
demscore

Thematic Datasets

Codebook

Peace



March 2025

Based on Demscore
Version 5.0

Funders: Demscore is funded by the Swedish Research Council, University of Gothenburg, Stockholm University, Uppsala University, and Umeå University. For more information, please visit: `https://demscore.se`

For data enquiries: contact@demscore.se

Summary of Table of Contents

Table of Contents

1 Explanatory Notes

1.1 Release Notes v5

Demscore provides worldwide free access to harmonized data on Democracy, Environment, Migration, Social Policy, Conflict and Representation from several of the world's most prominent social science research institutes. The interdisciplinary nature of Demscore data facilitates large-scale comparative analyses. This is essential to advance adequate policy responses to complex societal challenges associated with the Sustainable Development Goals (SDGs) and beyond, facing Sweden, Europe, and the world today.

With a firm commitment to transparency and openness, Demscore v5 enables users to gain comprehensive insights into various topics across the social sciences. The joint infrastructure ensures data integrity and quality at the highest international standards and maximizes usability in the measurement of contextual data with 25.000 variables across nearly all countries in the world, from 1750 to the present.

This creates critical time- and cost saving advantages in data collection, management, distribution, and not the least for end-users in the scientific community. Demscore's unique approach to translating and merging data scales up to more than 410.000 variable versions available in the infrastructure, storing more than 10 billion non-missing observations.

This collaborative effort between leading Swedish universities pushes the scale of social science data to a new level and offers unprecedented possibilities for interdisciplinary research and knowledge advancement.

These are the key features of Demscore:

1. **Customized Download:** A fully normalized, joint PostgreSQL database, sophisticated programming, and a user-friendly web-based interface for users to generate custom-designed datasets and codebooks for download.
2. **Translations and Data Merges:** Demscore currently offers more than 1000 merge options between datasets.
3. **Metadata:** Demscore takes information on and organization of metadata to new heights with the inclusion of customized codebooks, a detailed methodology document, and a comprehensive handbook.
4. **Handling of Missing Data:** Demscore pioneers in developing an innovative approach to tackle missing data. Researchers can now account for missing values with increased precision, leading to more robust and reliable analyses.
5. **Merge Scores:** Demscore introduces a unique merge mechanism. This powerful tool enables researchers to combine datasets effortlessly, uncovering connections and patterns that were previously hidden in isolated data silos.
6. **Thematic Datasets:** Demscore provides researchers with curated thematic datasets, each focused on a specific topic. These datasets bring together relevant variables from across the Demscore partners, facilitating in-depth investigations and comprehensive analyses of specific domains.
7. **Interactive Web Portal:** In addition to all the above, Demscore's web portal offers interactive visualization tools, user support and additional information on all partners and data sources.

For more information, please visit `https://www.demscore.se/` or contact contact@demscore.se.

1.2 New in Demscore version 5

A detailed description of changes and additions made for version 5 compared to version 4 can be found in the Methodology Document.

1.3 The Demscore Codebook

The autogenerated Demscore Codebook lists variable entries for those variables chosen by the user along with citation guidelines and licenses per variable.

The meta data is extracted from the codebooks per dataset stored in a table in the Demscore PostgreSQL database with one row per variable for all datasets. This table includes codebook entries, variable tags, labels, and other variable information in LaTeX format used to generate an automated codebook.
Demscore maintains a single set of standard entries for metadata across all datasets, to which all project members contribute their information. Additionally, variables within different datasets may have varying sets of additional information requirements specific to each dataset. These dataset-specific entries are also included, but they are presented as variable-specific metadata beneath the standard entries.
At the outset of the harmonization process, Demscore underwent a thorough variable name cleanup. This involved tasks such as replacing spaces or dots in variable names with underscores and converting all letters to lowercase. Notably, the original tags remain preserved and stored in the PostgreSQL table. Each variable in Demscore is accessible in both short and long forms. The short form comprises the cleaned version of the original variable tag, while the long form starts with the dataset name from which it originates, followed by the cleaned variable name.

For instance, the original name of the variable *MinisterPersonalID* from the H-DATA Foreign Minister Dataset is included as *ministerpersonalid* (short form) and *hdata_fomin_ministerpersonalid* (long form) in Demscore.

In addition, each dataset includes Demscore unit-identifier variables which are named according to the following naming scheme: Beginning with u_, followed by the name of the primary unit and finally the variable tag. The *year-* variable from the COMPLAB SPIN The Out-of-Work Benefits Dataset (OUTWB), which is part of the primary unit u_complab_country_year has the Demscore unit identifier name *u_complab_country_year_year*.

1.4 Methodology

For details on our methodology please see the Demscore Methodology document available for download on the Demscore website.

1.5 Citations

The Demscore project does not have a formal citation of its own. Hence, when using Demscore, we suggest that you cite the respective projects and datasets. We indicate how every dataset is to be cited in the autogenerated codebook you retreive 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 to that. 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 suggest you to also cite the Demscore Methodology Document when using data retrieved through Demscore.

1.6 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 a 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.

Please note that an observation that is missing in its original output unit does no take the value -11111, but appears as NA/blank cell in the customized dataset.

1.7 Download ID

The download ID can be shared with other users for replication purposes. A user can type the download ID into the Demscore website and retrieve the same download selection and files as the original user. This 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.

Download ID:

1.8 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.

1.9 Thematic Dataset

This thematic dataset offers variables related to peace and conflict resolution, covering topics such as peace agreements, ceasefires, disarmament, mediation efforts, and peacekeeping. It also includes data on political stability, reconciliation, and measures aimed at fostering harmony and reducing violence.

With a diverse range of structured variables, this dataset is an essential resource for those interested in analyzing the dynamics of peacebuilding, conflict prevention, and the mechanisms that sustain societal harmony.

1.10 Output Unit Identifier Variables in the Chosen Unit

u_demscore_country_year_country: The column is created based on V-Dem, H-DATA AND GW. It is based on the following datasets: H-DATA Information Capacity Dataset H-DATA Foreign Minister Dataset V-Dem Episodes of Regime Transformation Dataset V-Dem Country-Year: V-Dem Full+Others

u_demscore_country_year_code: NA

u_demscore_country_year_year: The column is created based on V-Dem, H-DATA AND GW. It is based on the following datasets: H-DATA Information Capacity Dataset H-DATA Foreign Minister Dataset V-Dem Episodes of Regime Transformation Dataset V-Dem Country-Year: V-Dem Full+Others

2 COMPLAB

Based at Stockholm University, the **Comparative Policy Laboratory (COMPLAB)**, provides vital policy data across three areas: environmental, social, and migration policy. The **Social Policy Indicators (SPIN)** database provides the foundations for new comparative and longitudinal research on causes and consequences of welfare states. Building on T.H. Marshall's ideas about social citizenship, SPIN makes available comparative data on social rights and duties of citizens, thereby moving research beyond analyses of welfare state expenditures. The SPIN database is instead oriented towards analyses of institutions as manifested in social policy legislation. Data are carefully collected in a coherent and consistent methodological manner to facilitate quantitative research of social policy across time and space. To date, SPIN covers 36 countries, of which several have data on core social policy programs from 1930 to 2019. More information is available on the project's website: `https://www.su.se/comparative-policy-laboratory/data/spin-1.644259` **GRACE, Governing the Anthropocene – Environmental Policy and Outcomes in a Comparative Perspective**, is a longitudinal and comparative study on environmental governance has created a dataset of national policy responses for environmental management and protection in 37 countries for the period 1970-2022. `https://www.su.se/comparative-policy-laboratory/data/grace-1.645779` **The Migration Policy Database (MIGPOL)** consists of a range of indicators compiled on behalf of leading data projects in the field of comparative migration policy research. It also contains original data on the rights of irregular migrants which will soon be added to Demscore. `https://www.su.se/comparative-policy-laboratory/data/migpol-1.645783` Read more about COMPLAB here: `https://www.su.se/comparative-policy-laboratory/`

2.1 COMPLAB MIGPOL Historical Immigration Policies Database

Dataset tag: complab_migpol_impic_antidisc

Output Unit: COMPLAB Country-Year, i.e., data is collected per country and year. That means each row in the dataset can be identified by one country in combination with a year, using the columns country_code (ISO 3-letter-code) and year or country_nr (ISO numeric code) and year. If necessary, an additional country column storing the countries' full names is created as a unit identifier. Please note that we synchronize Complab country variable names in Demscore to country_full_name, country_nr and country_id.

Description: The Historical Immigration Policies Database (HIP) covers 31 countries from either 1789 or their independence until the 2010s. These countries include: Argentina, Australia, Austria, Belgium, Botswana, Brazil, Canada, Chile, Denmark, Finland, France, Germany, Hong Kong, Ireland, Italy, Japan, Kuwait, the Netherlands, New Zealand, Norway, Saudi Arabia, Singapore, South Africa, South Korea, Spain, Sweden, Switzerland, Taiwan, the United Kingdom, the United States, and Venezuela. HIP enables scholars to reassess long-established views on the historical development of immigration policies, test new arguments with longitudinal data, and explore the relationship between immigration policies and slow-changing domestic and international variables. The dataset spans a wide range of variables important to political scientists, such as regime type, wealth (including natural resource wealth), and economic structure. Immigration policy dimensions—such as entry rules, rights, and enforcement—are coded separately, allowing researchers to analyze their long-term co-evolution. HIP is relevant to a broad audience in international relations and can be used to investigate immigration policy's connection to topics like North-South relations, democratization and autocratization trends, and the rise of far-right ideologies and populism. It holds particular promise for the growing subfield of historical international relations, with its focus on the evolution of states, state systems, and international ties. More information is available on: `https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/F7V8YL#`

Dataset citation: PETERS, MARGARET and Borang, Frida and Kalm, Sara; Lindvall, Johannes and Shin, Adrian, 2024, *Historical Immigration Policy dataset (HIP)*, https://doi.org/10.7910/DVN/F7V8YL, Harvard Dataverse, V2

Link to original codebook
`https://dataverse.harvard.edu/file.xhtml?fileId=10143944&version=2.0`

More detailed information on the dataset can be found at the following web page: `https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/F7V8YL#`

2.1.1 Historical Policy

Variables in this section provides coding of immigration policies of 31 states from 1789/ independence to 2010.

2.1.1.1 National Restrictions (natcode)

Long tag: complab_migpol_hip_natcode

Original tag: hip_natcode

Dataset citation: PETERS et al. (2024)

Merge scores:

Non-missing observations in original unit: Sum: 4747, Percent: 27.33

Non-missing observations in chosen unit: Sum: 4531, Percent: 15.12

Lost observations in chosen unit: Sum: 216 Percent: 4.55

Description:

DESCRIPTION: Does the law pertain to specic nationalities? Are exclusions based on national origin? Did the state sign a bilateral labor migration treaty or another international agreement on immigration?
VALUES:
1 = Only descedents of natives allowed in.
2 = A few nationalities allowed entrance but not many.
3 = Many nationalities allowed in but not all or migrants from some regions excluded.
4 = Almost all nationalities allowed in.
5 = No exclusions based on nationality.
MISSINGS:
Empty cell
COVERAGE:
1783-2010

3 H-DATA

The Historical Data Archive (H-DATA) is a hub of historical country-level data running as far back as the French revolution (1789) and offers unparalleled depth of data and temporality, enabling researchers to answer critical questions about the past but to also understand the origins of, and find historical parallels to, present-day problems. H-DATA works to collect, integrate, and curate historical data from Demscore's other modules. By adding this long-term historical dimension, H-DATA makes it possible for researchers to study the path dependency of political institutions where changes are incremental or rare thus making long time-series essential to understanding their causes and consequences. By extending data back into time, H-DATA helps deepen and further our understanding of the conditions of the complex global challenges that we face today. More information is available on the project's website: `https://www.su.se/english/research/research-projects/h-data`

3.1 H-DATA Leader Survival Dataset (PLT post-1789)

Dataset tag: hdata_plt

Output Unit: H-DATA Leader-Date, i.e., data is collected per leader and the day they got into power.

Description:
Timespan: 1789-2022
Coverage: 10,662 leader spells in 186 countries
The Leader Survival Dataset is the post-1789 part of the "Political Leaders through Time" (PLT) dataset, and provides data on entry and exit dates, type of leader position held, biographical background information, as well as appointment and exit reasons for 10,662 individual leaders in 186 countries (or territories) from 1789-2022. The dataset was used by Per Andersson Jan Teorell in their article "The Double-Edged Sword: How State Capacity Prolongs Autocratic Tenure but Hastens Democratization".

Dataset citation:
Gerring John, Nong Xin, Chatterton Ben, Cojocaru Lee, Dalli Cem Mert, Knutsen Carl Henrik, Kokkonen Andrej, Smith Daniel Steven, Teorell Jan, Selsky Sam, Ward Daisy, Jeon Ji Yeon. "Leader Tenure through the Ages: The Growth of Constraints." Unpublished manuscript, University of Texas at Austin, 2024.

Link to original codebook:
`https://www.su.se/polopoly_fs/1.803891.1740574267!/menu/standard/file/Codebook%20Leader%20Survival%20Data_last%20update%2020250225.pdf`

License: CC-BY-SA 4.0 International
`https://creativecommons.org/licenses/by-sa/4.0/legalcode`

More detailed information on the dataset can be found at the following web page: `https://www.su.se/english/research/research-projects/h-data/datasets-1.610144`

3.1.1 Territory Variables

In this section, we introduce the set of variables that describe each territory – its name (and alternate names), numeric code, GIS polygon, type (local unit, regional unit, country, colony, empire, confederation, dynastic conglomerate, stateless, tribe/chiefdom), relationship to other territories (if it is part of a superordinate entity), and region (Europe, Africa,...). Before beginning, several general clarifications are in order. All sorts of territories are contained in the PLT. Evidently, we are more concerned with larger, more enduring, and more powerful units. But there was no point in establishing a lower threshold for inclusion, as this would be difficult to define and to enforce. Instead, we differentiate among different types of territories with the Territory type variable (below). These territorial designations evidently overlap: cities lie within regions, regions lie within countries, and

so forth. However, territories of the same type should be discrete (non-overlapping). For example, a latitude/longitude point ought to lie within one and only one country at a particular point in time. A basic principle is that political entities with different territorial jurisdictions (or effective control) receive different territorial ID codes. Jurisdictions may overlap, or they may be layered (as cities lie within regions and regions lie within countries), but they are nonetheless unique. We do not assign the same ID to differently sized territories. This means that in situations of civil war, camps controlling different territory must receive separate territory IDs, if they are coded at all. For example, if the US Confederacy is coded it must have a unique ID, separate from the United States. We recognize that there are plenty of disputed territories and poorly mapped territories, especially as one moves backward in time. Our project does not attempt to sort out these controversies; we simply follow conventional wisdom among historians. Within/without relationships are established with GIS polygons (for sizeable entities) and points (for small entities). Of course, we do not have comprehensive GIS codes for every territory. But we do have them for large entities and we hope to add to our collection so that coverage is comprehensive at some point in the future. Superordinate/subordinate relationships can be ascertained by indicating, for each territory, all superordinate territories that it is a part of. For a typical city, this might be (a) region, (b) country, and (c) world region. For a colony, it might be (a) empire, (b) world region. (Because subordinate relationships can be inferred from superordinate relationships, the former do not need to be noted separately.)

3.1.1.1 Territory code, continuous (territorycode)

Long tag: hdata_plt_territorycode

Original tag: territorycode

Dataset citation: Gerring et al. (2024)

Description:

Question: What is the continuous code of this territory?
Clarification: This primary territorial ID code emphasizes continuity. This is consistent with country units as defined by V-Dem (Coppedge et al. 2022), a protocol that we follow for the modern era.
More specifically, wherever there is a fair degree of territorial continuity, or where at least the capital or core of the territory remains the same, the territory retains the same code. The Ottoman Empire and Turkey, for example, receive the same primary code.
In cases where a state is vanquished, occupied, or collapsed for a relatively short period of time – e.g., some European states during the Napoleonic wars and the great wars of the twentieth century – we assume continuity as long as statehood is restored after the disruption.
In cases where multiple states are absorbed by a dominant state, we assign the same code to both units. For example, Prussia receives the same code as Germany.
In cases where multiple units are brought together peacefully into a larger unit and where there is a dominant previous unit that forms the core of the new state, we assign the same code to the previously dominant unit. For example, New South Wales receives the same code as Australia.
Likewise, where a state fragments into several states we assign the same code to a successor state if it is the dominant unit of the previous agglomeration. For example, Serbia receives the same code as Yugoslavia.
In cases where there is an interruption of state control, perhaps due to the dissolution of the state or foreign takeover, after which a state is reestablished over the same (or similar) territory, we assign the same code to both units if the interruption is less than 50 years.
However, where there are fundamental changes in governance and territory, we introduce separate territory codes. For example, we do not assume continuity between the Byzantine and Ottoman Empires, or Islamic Andalusia and Spain. Likewise, if there are a multitude states, none of which appears to be dominant or is in some way a continuation of the predecessor or successor state, we assign a new code. Colonies of New England each receive their own code, separate from the code assigned to the United States, for example.
[numeric]

4 QOG

The **Quality of Government (QoG)** Institute was founded in 2004 by Professor Bo Rothstein and Professor Sören Holmberg. It is an independent research institute within the Department of Political Science at the University of Gothenburg. QoG is comprised of about 30 researchers who conduct and promote research on the causes, consequences and nature of Good Governance and the Quality of Government (QoG) - that is, trustworthy, reliable, impartial, uncorrupted and competent government institutions. QoG's award-winning datasets focus on concepts related to quality of government, transparency, and public administration. The main objective of QoG's research is to address the theoretical and empirical problem of how political institutions of high quality can be created and maintained. A second objective is to study the effects of Quality of Government on a number of policy areas, such as health, the environment, social policy, and poverty. The QoG datasets draw on a number of freely available datasources. More information on how the variables are complied for different QoG datasets can be found in the respective QoG codebooks available on their website. More information is available on the project's website: `https://www.gu.se/en/quality-government`

4.1 QoG Standard Dataset Time-Series

Dataset tag: qog_std_ts

Output Unit: QoG Country-Year, i.e., data is collected per country and year. That means there is one row for each combination of country and year in the dataset. This unit is identified using the cname column and the year column.

Description: The QoG Standard dataset is our largest dataset. It consists of approximately 2100 variables from more than 100 data sources related to Quality of Government. In the QoG Standard TS dataset, data from 1946 to 2024 is included and the unit of analysis is country-year (e.g., Sweden-1946, Sweden-1947, etc.).

Dataset citation: Teorell, Jan, Aksel Sundström, Sören Holmberg, Bo Rothstein, Natalia Alvarado Pachon, Cem Mert Dalli, Rafael Lopez Valverde, Victor Saidi Phiri Lauren Gerber. 2025. The Quality of Government Standard Dataset, version Jan25. University of Gothenburg: The Quality of Government Institute, https://www.gu.se/en/quality-government doi:10.18157/qogstdjan25. University of Gothenburg: The Quality of Government Institute, `https://www.gu.se/en/quality-government doi:10.18157/qogstdjan24`

Link to original codebook
`https://www.qogdata.pol.gu.se/data/codebook_std_jan25.pdf`

License: The QoG datasets are open and available, free of charge and without a need to register your data. You can use them for your analysis, graphs, teaching, and other academic-related and non-commercial purposes. We ask our users to cite always the original source(s) of the data and our datasets.

We do not allow other uses of these data including but not limited to redistribution, commercialization and other for-profit usage. If a user is interested in such use or has doubts about the license, they will have to refer to the original source and check with them if this is allowed and what requirements they need to fulfill.

Be mindful that the original data sources are the only owners of their data and they can adjust their license without previous warning.

More detailed information on the dataset can be found at the following web page: `https://www.gu.se/en/quality-government/qog-data/data-downloads/standard-dataset`

4.1.1 Civil Society, Population and Culture

This category includes variables that relate to social capital, personal beliefs, size and distribution of the population as well as ethnic and linguistic fractionalization.

4.1.1.1 Global Peace Index (gpi_gpi)

Long tag: qog_std_ts_gpi_gpi

Original tag: gpi_gpi

Dataset citation: Teorell et al. (2025)

Variable citation: Institute for Economics and Peace (2022)

Merge scores:

Non-missing observations in original unit: Sum: 2540, Percent: 16.61

Non-missing observations in chosen unit: Sum: 2472, Percent: 8.25

Lost observations in chosen unit: Sum: 68 Percent: 2.68

Description:

The GPI (scaled from 1 to 5, 5 being least peaceful) measures a country's level of Negative Peace using three domains of peacefulness. The first domain, Ongoing Domestic and International Conflict, investigates the extent to which countries are involved in internal and external conflicts, as well as their role and duration of involvement in conflicts.

The second domain evaluates the level of harmony or discord within a nation; ten indicators broadly assess what might be described as Societal Safety and Security. The assertion is that low crime rates, minimal terrorist activity and violent demonstrations, harmonious relations with neighbouring countries, a stable political scene and a small proportion of the population being internally displaced or made refugees can be equated with peacefulness.

Seven further indicators are related to a country's Militarisation-reflecting the link between a country's level of military build-up and access to weapons and its level of peacefulness, both domestically and internationally. Comparable data on military expenditure as a percentage of GDP and the number of armed service officers per head are gauged, as are financial contributions to UN peacekeeping missions.

5 Bibliography

Gerring, J., Nong, X., Chatterton, B., Cojocaru, L., Dalli, C. M., Knutsen, C. H., Kokkonen, A., Smith, D. S., Teorell, J., Selsky, S., Ward, D. & Jeon, J. Y. (2024), Leader tenure through the ages: The growth of constraints. Unpublished manuscript, University of Texas at Austin.

Institute for Economics and Peace (2022), 'Global peace index 2022: Measuring peace in a complex world'. Accessed 01-09-2022.
URL: *http://visionofhumanity.org/resources*

PETERS, M., Borang, F., Kalm, S., Lindvall, J. & Shin, A. (2024), 'Historical Immigration Policy dataset (HIP)'.
URL: *https://doi.org/10.7910/DVN/F7V8YL*

Teorell, J., Sundström, A., Holmberg, S., Rothstein, B., Alvarado Pachon, N., Dalli, C. M., Lopez Valverde, R., Phiri, V. S. & Gerber, L. (2025), *The Quality of Government Standard Dataset, version Jan25.*
URL: *https://www.gu.se/en/quality-government*