The Annotated Database Bibliography

Codebook

v.1, February 2013

Nadine Ansorg, Matthias Basedau, Felix Haaß, Julia Strasheim

Institutions for Sustainable Peace (ISP)

GIGA German Institute of Global and Area Studies

---

The authors are grateful to the participants of the First ISP Network Conference “Institutions for Sustainable Peace. From Research Gaps to New Frontiers” and Nina von Uexkueell for their valuable comments on a first draft of the ADBB. They furthermore thank Nadine Schroeder and Sven Weimert for research assistance.
# Table of Contents

Introduction .......................................................................................................................... 4  
Coding Process ......................................................................................................................... 5  
General Information .................................................................................................................. 5  
  Database ISP ID ..................................................................................................................... 5  
  Name ...................................................................................................................................... 5  
  Author(s) ............................................................................................................................... 6  
  Institution .............................................................................................................................. 6  
  Funded by .............................................................................................................................. 6  
  Publication date ....................................................................................................................... 6  
  Latest version .......................................................................................................................... 6  
Topic ......................................................................................................................................... 6  
  Content of variables – original or data collection/construction ............................................. 6  
  Content of variables – other data .......................................................................................... 6  
  Cluster Institutions ............................................................................................................... 6  
  Cluster Conflict/Peace ........................................................................................................... 8  
  Cluster Divisions .................................................................................................................. 9  
  Cluster Surrounding Conditions ............................................................................................ 10  
Codebook ................................................................................................................................. 10  
Data Feature Article .................................................................................................................. 10  
Source of Data .......................................................................................................................... 10  
Type of Database ..................................................................................................................... 10  
Methodological Type of Data .................................................................................................... 11  
Scope ....................................................................................................................................... 11  
  Level of Aggregation ............................................................................................................. 11  
  Time Span ............................................................................................................................. 11  
  Geographical Scope ............................................................................................................. 11  
  Availability ............................................................................................................................ 11  
Methodological Remarks ......................................................................................................... 11  
  Methodological Challenge 1: Scope/Time ........................................................................... 11  
  Methodological Challenge 2: Scope/Geography ................................................................. 11  
  Methodological Challenge 3: Level of Aggregation/Time .................................................... 11  
  Methodological Challenge 4: Level of Aggregation/Groups ................................................ 12
Methodological Challenge 5: Level of Aggregation/Geography ........................................ 12
Methodological Challenge 6: Transparency ................................................................ 12
Methodological Challenge 6: Availability ..................................................................... 12
Innovation .................................................................................................................. 12
Missing data .............................................................................................................. 12
Comments .................................................................................................................. 12
Introduction

The Annotated Database Bibliography (ADBB) provides a systematic overview and critical assessment of the existing datasets on institutions, conflict and peace, and divisions and diversity. It identifies data gaps and assesses a number of pertinent methodological criteria. It includes information on four different categories of a database: general information, topic, scope, and methodological remarks. The categories each contain several subcategories that include more detailed information. The codebook presents all these categories including the subcategories in the same order.

To identify the universe of cases we relied on a variety of sources to select relevant datasets. First, we consulted existing data overview articles. Such articles are widely available in the study of conflict and – albeit to a lesser extent – in the study of institutions.\(^2\) Second, we reviewed leading journals in the field of peace and conflict research that primarily publish original quantitative work and specific data feature articles.\(^3\) Third, we examined metadatabases,\(^4\) such as The Quality of Government Institute’s data collection.\(^5\) Furthermore, we relied on web compilations of datasets, inter alia Paul Hensel’s collection of international relations and comparative politics datasets.\(^6\) We also applied snowball sampling, by adding datasets that we found in the course of researching other datasets.\(^7\)

By applying these techniques we collected a body of 609 databases (see ISP database collection). In a subsequent step we settled on specific selection criteria and coding rules to categorize datasets in a single file. Starting point for our selection criteria is the definition of a “database” as a structured collection of data that contains systematic quantitative or qualitative information across different units of analysis on the topical clusters institutions, conflict/peace, and divisions/diversity.

A dataset had to meet a further set of criteria to be included in our final analysis. First, it had to contain originally collected or coded data on at least one of the three main topics of institutions, conflict and peace, or divisions and diversity. We excluded replication datasets that are used in novel analyses, since our analytical focus was on innovative data, rather than methods or theory. We included, however, indices that are constructed on the basis of existing data such as, to name only one example, the Failed State Index published by the Fund for Peace.\(^8\) Second, the dataset had to


\(^3\) Such journals included particularly, but were not limited to, the *Journal of Peace Research, the Journal of Conflict Resolution, International Interactions, or Conflict Management and Peace Science*.

\(^4\) “Metadatabases” are database collections including a variety of independent databases to analyze specific topics, usually without collecting their own data.

\(^5\) Online at <http://www.qog.pol.gu.se/>.

\(^6\) Online at <http://www.paulhensel.org/data.html>. See also the collection of Markus Eberhardt online at <https://sites.google.com/site/medecon/development-economics/devecondata>.

\(^7\) We are aware that we most likely were not able to include all existing datasets using these methods, especially since the availability of datasets has grown tremendously over the last decade. Our selection criteria also may have led to a selection bias, including only those datasets that are cited in any of our sources. However, we are trying for an approximation of completeness and we are very happy for readers to email us their datasets to isp@giga-hamburg.de in case they fit our criteria (see below).

\(^8\) Online at <http://www.fundforpeace.org/global/?q=fsi>.
cover information within the period 1946 to 2012. Third, the dataset had to include information on a range of units of analysis (i.e., more than two). The included datasets typically had more than 20 units of analysis, indicating that a certain number of analyzed units is necessary for generalized conclusions on the causal relationship between our three topical clusters. This meant we excluded qualitative single-case studies and noncomparable case studies, but included subnational quantitative data (e.g., survey data or data on subnational units within countries such as Indonesia or India). Fourth, the dataset must have been published in or before December 2012 and at least one version of the dataset (e.g., an updated version) must have been published later than 1990. We excluded datasets that were still in the pilot phase and had only published preliminary results such as the Varieties of Democracy Project. Ninth, the dataset had to be published in English, the lingua franca in modern science. Sixth, the dataset was not allowed to exclusively cover World Bank high-income countries, due to our interest in developing and conflict-affected countries. Seventh, since we were interested in intrastate phenomena, such as civil war, we did not include datasets that exclusively cover international relations data (i.e., data that mainly captures inter state relations).

Coding Process
Each database passed through several coding and checking steps before entering the ADBB: In a first coding process a database was coded according to the coding rules that can be found in this codebook and in the master document by either the team members or student assistants. The coding results where then checked in a second step by one of the team members. Continuous communication on critical cases or questions in database meetings accompanied the whole coding and checking process.

General Information
Includes all general information on the database.

Database ISP ID
A unique identifier of every database coded for the Annotated Database Bibliography.

Name
The official name of the database plus acronym. The name explicitly stated by the database authors is used here.

---


10 Several languages are at the disposal of the authors, but we decided for universal accessibility. In fact, only few data bases such as the war archive (Kriege-Archiv) of the Arbeitsgemeinschaft Kriegsursachenforschung were excluded on the basis of this criterion.

11 The countries in question were mainly identified by using the World Bank category and respective datasets were excluded from the sample. The World Bank’s list of high-income countries includes: Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Luxembourg, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, South Korea, Spain, Sweden, Switzerland, United Kingdom, United States (list taken from <http://data.worldbank.org/about/country-classifications/country-and-lending-groups>) We included datasets on the Israel–Palestine conflict, however, due to its geographic location in a conflict-prone region and its political relevance.
Author(s)
First and last names of the authors of database. If there is an official recommendation to cite the database, these names were used. If no names are explicitly mentioned, the names probably closest related to the database are used.

Institution
Institutional affiliation of the author(s).

Funded by
Includes information on the funding of the database.

Publication date
Year of first publication.

Latest version
Year of latest publication.

Topic

Content of variables – original or data collection/construction
Provides short description of the kind of variables included. Only include descriptions of originally coded or constructed (e.g. new indices) variables are included.

Content of variables – other data
Provides short description of the kind of variables additionally included; e.g. what variables did the author directly take from other datasets such as UCDP/PRIO, World Bank, Freedom House. Shows “N.a.” if no other data is provided.

General rule for coding clusters and subclusters: only code “1” if data was originally collected or constructed originally by the author(s).

Cluster Institutions
Provides information if database contains information on the cluster “institutions”.

(0) Database does not contain information on the cluster “institutions”.

(1) Database contains information on the cluster “institutions”.

Sub cluster I – state, political
Provides information if database contains information on the sub cluster “state – political institutions”, i.e. formal political state institutions such as administrative territorial state structure, electoral system, party system, form of government, state capacity, regime type, human rights (also violations).

(0) Database does not contain information on the sub cluster “state – political institutions”.

(1) Database contains information on the sub cluster “state – political institutions”.
(1) Database contains information on the sub cluster “state – political institutions”.

Sub cluster I – state, socio-economic
Provides information if database contains information on the sub cluster “state – socio-economic institutions”, i.e. formal economic state institutions such as land reform, wealth sharing, welfare system.

(0) Database does not contain information on the sub cluster “state – socio-economic institutions”.
(1) Database contains information on the sub cluster “state – socio-economic institutions”.

Sub cluster I – state, security
Provides information if database contains information on the sub cluster “state – security institutions”, i.e. formal institutions of a state’s (broader) security sector such as military, police, customs, judiciary, peacekeeping missions (post-conflict institutions).

(0) Database does not contain information on the sub cluster “state – security institutions”.
(1) Database contains information on the sub cluster “state – security institutions”.

Sub cluster I – non-state, political
Provides information if database contains information on the sub cluster “non-state – political institutions”, i.e. informal political institutions that are not part of the state structure such as informal ethnic coalitions.

(0) Database does not contain information on the sub cluster “non-state – political institutions”.
(1) Database contains information on the sub cluster “non-state – political institutions”.

Sub cluster I – non-state, socio-economic
Provides information if database contains information on the sub cluster “non-state – socio-economic institutions”, i.e. informal socio-economic institutions that are not part of the state structure such as NGOs or unions.

(0) Database does not contain information on the sub cluster “non-state – socio-economic institutions”.
(1) Database contains information on the sub cluster “non-state – socio-economic institutions”.

Sub cluster I – non-state, security
Provides information if database contains information on the sub cluster “non-state – security institutions”, i.e. informal security institutions that are not part of the state structure such as private security providers, local defense groups.

(0) Database does not contain information on the sub cluster “non-state – security institutions”.
(1) Database contains information on the sub cluster “non-state – security institutions”.

Sub cluster I – peace agreements and negotiations/mediation
Provides information if database contains information on the sub cluster “peace agreements and negotiations/mediation”, i.e. special peace related institutions such as the content of peace agreements, negotiations, or mediation.
(0) Database does not contain information on the sub cluster “peace agreements and negotiations/mediation”.
(1) Database contains information on the sub cluster “peace agreements and negotiations/mediation”.

**Cluster Conflict/Peace**
Provides information if database contains information on the cluster “conflict/peace”.

(0) Database does not contain information on the cluster “conflict/peace”.
(1) Database contains information on the cluster “conflict/peace”.

**Sub cluster C – state, low-level**
Provides information if database contains information on the sub cluster “state, low-level conflict”, i.e. violence resulting in less than 25 BRD/year, where at least one state is one side to the conflict such as “social conflict”, police raids, human rights violations.

(0) Database does not contain information on the sub cluster “state, low-level conflict”.
(1) Database contains information on the sub cluster “state, low-level conflict”.

**Sub cluster C – state, medium-level**
Provides information if database contains information on the sub cluster “state, medium-level conflict”, i.e. violence resulting in 25-999 BRD/year, where the state is one side to the conflict such as minor armed conflict (intrastate and interstate), MID.

(0) Database does not contain information on the sub cluster “state, medium-level conflict”.
(1) Database contains information on the sub cluster “state, medium-level conflict”.

**Sub cluster C – state, high-level**
Provides information if database contains information on the sub cluster “state, high-level conflict”, i.e. violence resulting in more than 1000 BRD/year, where the state is one side to the conflict such as civil war, intrastate war, internationalized war (interventions in civil wars); one-sided violence by the state.

(0) Database does not contain information on the sub cluster “state, high-level conflict”.
(1) Database contains information on the sub cluster “state, high-level conflict”.

**Sub cluster C – non-state, low-level**
Provides information if database contains information on the sub cluster “non-state, low-level conflict”, i.e. violence resulting in less than 25 BRD/year, where the state is not a side to the conflict such as “social conflict among non-state actors”, riots, violent demonstrations, crime, homicide, individual terrorism.

(0) Database does not contain information on the sub cluster “non-state, low-level conflict”.
(1) Database contains information on the sub cluster “non-state, low-level conflict”.
Sub cluster C - non-state, medium-level
Provides information if database contains information on the sub cluster “non-state, medium-level conflict”, i.e. violence resulting in 25-999 BRD/year, where the state is not a side to the conflict such as intercommunal violence or organized terrorism.

(0) Database does not contain information on the sub cluster “non-state, medium-level conflict”.
(1) Database contains information on the sub cluster “non-state, medium-level conflict”.

Sub cluster C - non-state, high-level
Provides information if database contains information on the sub cluster “non-state, high-level conflict”, i.e. violence resulting in more than 1000 BRD/year, where the state is not a side to the conflict such as gang wars or organized terrorism.

(0) Database does not contain information on the sub cluster “non-state, high-level conflict”.
(1) Database contains information on the sub cluster “non-state, high-level conflict”.

Sub cluster C - peace
Provides information if database contains information on the sub cluster “peace”, i.e. information that goes beyond the mere absence of direct physical violence.

(0) Database does not contain information on the sub cluster “peace”.
(1) Database contains information on the sub cluster “peace”.

Cluster Divisions
Provides information if database contains information on the cluster “divisions”.

(0) Database does not contain information on the cluster “divisions”.
(1) Database contains information on the cluster “divisions”.

Sub cluster D – attributed, demographic constellation
Provides information if database contains information on the sub cluster “divisions – attributed, demographic constellation”, i.e. characteristics attributed by birth (physical appearance and traditions) such as ethnicity, religion, linguistics, cultural fractionalization, gender, age; e.g. Ethnolinguistic Fractionalization (ELF)-Index.

(0) Database does not contain information on the sub cluster “divisions – attributed, demographic constellation”.
(1) Database contains information on the sub cluster “divisions – attributed, demographic constellation”.

Sub cluster D – attributed, quality of relations
Provides information if database contains information on the sub cluster “divisions – attributed, quality of relations”, i.e. the political salience of attributed identities or any other information beyond demographic constellations such as level of confrontation between ethnic, religious, etc. groups; access to power, example: PREG Index (Politically Relevant Ethnic Groups).

(0) Database does not contain information on the sub cluster “divisions – attributed, quality of relations”.
(1) Database contains information on the sub cluster “divisions – attributed, quality of relations”.

**Sub cluster D – acquired, demographic constellation**
Provides information if database contains information on the sub cluster “divisions – acquired, demographic constellation”, i.e. characteristics acquired throughout life (political, social, economic distinctions) such as ideological, political, social, economic; e.g. Gini-Coefficient.

(0) Database does not contain information on the sub cluster “divisions – acquired, demographic constellation”.
(1) Database contains information on the sub cluster “divisions – acquired, demographic constellation”.

**Sub cluster D – acquired, quality of relations**
Provides information if database contains information on the sub cluster “divisions – acquired, quality of relations”, i.e. the political salience of acquired identities such as level of confrontation between ideological groups or supporters of political parties.

(0) Database does not contain information on the sub cluster “divisions – acquired, quality of relations”.
(1) Database contains information on the sub cluster “divisions – acquired, quality of relations”.

**Cluster Surrounding Conditions**
Provides information if database contains information on the cluster “surrounding conditions”.

(0) Database does not contain information on the cluster “surrounding conditions”.
(1) Database contains information on the cluster “surrounding conditions”.

**Codebook**
Availability of code including link and remark if necessary (e.g. reduced accessibility through uncommon language or dead link).

**Data Feature Article**
Existence of a data feature article including name full reference. A data feature article is an article in which the database as such is presented. If several such articles exist, the latest version is named.

**Source of Data**
Sources on which the data builds on are named, in particular reports by international organizations, news reports by news agencies etc. In case of index construction, sources are mentioned in the databases of which the index consists. In case of expert assessments, the source of information and data type are identical.

**Type of Database**
Indicates the type of database, whether it is an original collection of data, an index construction, or a mixed version. Originally collected data: data collected originally on the basis of coding of documents, polls, etc.; index construction: existing data used to construct a new index; mixed: both original
collection and index construction. This question applies only to originally coded or constructed variables in the dataset.

**Methodological Type of Data**
Indicates whether data is subjective, objective or expert assessment, or a mixture. Subjective = data from survey/opinion pulls (attitudes, self-report); Objective = data taken from statistics or coding on the basis of news reports/year books/documents; Expert assessment (with no particular source of information).

Indicates further whether data is quantitative or qualitative. Quantitative = data are presented in numbers (at least ordinal scale level; binary: 1/0) Qualitative = data are presented verbally only.

**Scope**

**Level of Aggregation**
Indicates how disaggregated the data is in terms of time, geography, and actors, e.g. country/year or group/year etc. Time: day, period, or year; Geography: polygon/cell, administration unit, country; actor dimension included: Individual actors or collective actors? (e.g. ordinary people, warring parties, elites).

**Time Span**
Indicates the years or period covered by the database.

**Geographical Scope**
Contains the geographical units covered/excluded by the database. If readily available, number of countries included are mentioned.

**Availability**
Indicates if the database is readily available. 1. Publicly available?, 2. For free?; 3. Online (if available, link is added), 4. File format: e.g. Excel, STATA, SPSS.

**Methodological Remarks**

**Methodological Challenge 1: Scope/Time**
Indicates whether database covers more or less than 25 years.

(0) Database covers less than 25 years.
(1) Database covers more than 25 years.

**Methodological Challenge 2: Scope/Geography**
Indicates whether database has or has not a global scope.

(0) The database has no global scope.
(1) The database has a global scope.

**Methodological Challenge 3: Level of Aggregation/Time**
Indicates whether the temporal level of aggregation is yearly or more detailed (daily, weekly, monthly).
(0) Temporal level of aggregation is yearly.
(1) Temporal level of aggregation is monthly/weekly/daily/etc..

**Methodological Challenge 4: Level of Aggregation/Groups**
Indicates whether the level of aggregation of actors is at the state level or more detailed (individuals, non-state groups etc.).

(0) The level of aggregation of actors is at the state level.
(1) The level of aggregation of actors is at the group/organization/individual/etc. level.

**Methodological Challenge 5: Level of Aggregation/Geography**
Indicates whether the geographical level of aggregation is at the country level or more detailed (city, community, administration unit etc.).

(0) The geographical level of aggregation is at the country level.
(1) The geographical level of aggregation is at the city/community/administration unit/etc. level.

**Methodological Challenge 6: Transparency**
Indicates whether database has transparency problems or not. A database has transparency problems e.g. if codebook is not publicly available, if there are imprecise/unclear thresholds for inclusion or lack of precise description. A database does not have transparency problems if codebook is available with precise thresholds and descriptions. A short reasoning for choice for the respective methodological problem is provided in the next cell labeled “reasoning”.

(0) The database has transparency problems.
(1) The database has no transparency problems.

**Methodological Challenge 6: Availability**
Indicates whether database has availability problems. A database has availability problems if data cannot not be obtained online or by email to the author, if it is only for sale, or if it is available only in book form. A database has no availability problems if data is readily available online, by email, and for free. A short reasoning for the choice of respective methodological problem is provided in the next cell labeled “reasoning”.

(0) The database has availability problems.
(1) The database has no availability problems.

**Innovation**
Provides short reasoning if database is particularly innovative.

**Missing data**
Provides short description if data is obviously missing.

**Comments**
Open section. Indicates if database has severe problems with reliability or validity. Those can include that the database uses solely expert assessments or secondary sources without intercoding; if the authors changed measurement; if data is extrapolized; if concepts are operationalized by a single indicator only; if concepts are poorly operationalized; if indicators do not have corresponding concepts; if some elements of concepts are not operationalized; etc.