Codebook for the MSDD Dataset*

David E. Cunningham

University of Maryland, College Park

Peace Research Institute Oslo

Leo Bauer

University of Maryland, College Park

Sloan Lansdale

University of Maryland, College Park

Megan Lloyd

University of Maryland, College Park

Version 1.0 (June 30, 2025)

*Data collection was funded through the National Science Foundation (grant 1948669). We thank Mingyang Su, Ellen Tuttle, Fatima Younis, Mason Blackburn, and Cindy Sung for research assistance.

Overview

The Mediation in Self-Determination Disputes (MSDD) dataset introduces the first cross-national dataset on mediation events in self-determination (SD) disputes irrespective of war. The dataset is of global reach, representing a random sample of SD movements active from 1991 to 2015 both in and out of conflict. The dispute-dyads on which the MSDD is based are taken from Cunningham's (2013) list of SD movements. The MSDD is an event dataset and thus takes the mediation event as its unit of analysis.

A total of 51 disputes were coded in random order.[1] Of the 51 disputes, 5 were only coded as to whether or not they received mediation on the dispute-year level.[2] Of the remaining 46 disputes, 26 received mediation and are included in the MSDD. These disputes were the recipients of 1,284 mediation events between the years of 1991 and 2015. An additional 20 disputes did not receive mediation. Some of these disputes saw third-party involvement that did not rise to the level of mediation.[3] The data release package contains a supplementary file that records all dispute-years for the 51 disputes coded in random order. It eases the conversion of the event data into time-series data and includes a variable denoting which disputes were only coded on the dispute-year level. An accompanying readme-file explains its utility in more detail.

Each mediation event recorded in the MSDD is identified in relation to a specific dispute, method of mediation, and must be continuous in time. All participants from the disputant sides and third parties present in a mediation event are included in the same observation. The MSDD codes three different types of mediation events: bilateral mediation, shuttle mediation, and direct mediation.

[1] One dispute, China-Tibetans (KGCID 106), was dropped from the sample due to the extreme volume of news articles related to the dispute. Four non-randomly chosen test cases were also coded but are not included in the MSDD.

[2] This was done due to the high volume of news articles related to these disputes.

[3] These edge cases are excluded from the final dataset but a repository of these events is maintained and can be accessed at the researcher's request.

Basic Definitions

Mediation: A meeting between one or both parties to an SD dispute and a third party in which (1) the dispute is discussed and (2) the third party shows a clear intent to have the dispute resolved peacefully, either via proclamation or action. While the parties to the dispute are either the state government or a direct affiliate, or an organization that is part of the SD movement, third parties acting as mediators may be individuals, state governments, international or non-governmental organizations, or other political entities.

Bilateral mediation: A mediation event where either, but not both, the state or the SD group is present along with at least one third party acting as a mediator, discussing the dispute with clear intent to resolve the dispute.

Shuttle mediation: A mediation event where both sides of the dispute are represented along with at least one third party acting as a mediator, discussing the dispute with clear intent to resolve the dispute. However, the two representatives of the disputants are not in the same location and the mediator travels between them to facilitate mediation.

Direct mediation: A mediation event where both sides of the dispute are present along with at least one third party acting as a mediator, discussing the dispute with clear intent to resolve the dispute in the same location at the same time.

Data Generation

In a first step, a random sample was drawn from the list of SD disputes compiled by Cunningham (2013), which served as the population of interest for data collection. SD disputes were then randomly assigned to coders. Coders relied on the news article database NexisUni to code mediation events.[4] Search strings comprising disputant names and a fixed set of keywords related to mediation were used systematically across all coded disputes.

[4]For a small set of initial disputes, the Factiva news article database was used as well.

Example search string: (Moldova) w/seg [Gagauz*] w/seg [mediation OR talks OR meeting OR visit OR envoy OR "shuttle diplomacy" OR "good offices" OR ceasefire OR settlement OR "peace agreement" OR "special representative" OR representative OR delegat* OR diploma* OR ambassador OR "peace deal" OR dialogue OR fact-finding OR "fact finding" OR bilateral OR multilateral OR multiparty OR resolution OR "round table" OR round-table OR workshop]

In a second step, the search string was used to query NexisUni. The resulting set of articles was preprocessed with machine learning to reduce coder workload. An initial set of disputes was hand-coded and with the resulting mediation events a series of language models were trained to classify news articles on whether or not they contain mentions of mediation events, utilizing the NexisUni API. Based on model performance a DistilBERT model was selected to classify all articles for the remaining disputes. As mediation events are rare, the model was configured particularly to ensure a low number of false negatives.

In a third step, coders then hand-coded mediation events from the resulting set of articles for each dispute. As mediation is not always explicitly described in news articles, in a first step coders recorded all events in which one or both disputants met with a third party and talked about the dispute or met directly in a foreign country with that country's government acting as the host. Yet, as the dispute may only be a secondary topic during a meeting (as in, e.g., general proclamations about domestic peace and stability during an economic summit), coders recorded whether or not the event constituted mediation only in a second step. This procedure allowed for tracking edge cases with third-party involvement short of mediation and improved coder training.

In a last step, the final number of recorded events was subset to only those which were coded as constituting mediation. All machine learning and final data cleaning tasks were performed in Python. The dataset contains information on the title and publication date of the news articles from which a mediation event was coded, so that researchers may reproduce coding decisions. Additional information about sources can be provided upon request.

Variables

Variable name	Description	Type
event_id	Unique mediation event identifier	String
kgcid	SD movement code, KGCID ID system from Cunningham (2013) is used	Integer
dispute	SD dispute name, formatted as country-movement	String
gwcode	Identifier of country in which dispute occurs, Gleditsch and Ward country code is used	Integer
start_date	Start date of mediation event, formatted as YYYY-MM-DD	String
end_date	End date of mediation event, formatted as YYYY-MM-DD	String
start_precision	Precision indicator of start date of mediation event: 1 = YYYY-MM-DD 2 = YYYY-MM 3 = YYYY	Integer
end_precision	Precision indicator of start date of mediation event: 1 = YYYY-MM-DD 2 = YYYY-MM 3 = YYYY	Integer
year	Start year of mediation event	Integer

Variable name	Description	Type
event_type	Indicator of the type of mediation that occurred: 1 = Bilateral mediation 2 = Shuttle mediation 3 = Direct mediation	Integer
state_present	Binary variable of whether the state side was present at mediation event	Boolean
sd_present	Binary variable of whether the SD movement side was present at mediation event	Boolean
participants	Names of state and SD movement participants in mediation event. Sides separated by semicolon. Specific actors for both sides listed in parentheses, separated by comma if multiple.	String
thirdparty_a	Name of main third-party mediator. Specific actors listed in parentheses, separated by comma if multiple.	String
thirdparty_other	Name of additional third-party mediator. Separated by semicolon if multiple.	String
location	Location of mediation event. Formatted as city, country if city is available.	String
article_title	Title of the news article from which event was coded	String
article_date	Publication date of the news article from which event was coded, formatted as YYYY-MM-DD	String

References

Cunningham, Kathleen Gallagher. 2013. "Actor Fragmentation and Civil War Bargaining: How Internal Divisions Generate Civil Conflict." *American Journal of Political Science* 57 (3): 659–672. https://doi.org/10.1111/ajps.12003.