

# PA-X CORPUS Codebook

Version 2, September 2024



Cite as:

*Bell, Christine, Sanja Badanjak, Robert Forster, Astrid Jamar, Sean Molloy, Kevin McNicholl, Kathryn Nash, Jan Pospisil, Laura Wise (2023). PA-X Corpus Codebook. Peace and Conflict Resolution Evidence Platform, University of Edinburgh, Edinburgh. [pax.peaceagreements.org](https://pax.peaceagreements.org)*

## Accessing and downloading the Corpus

The PA-X Corpus is available via the PA-X main search page, at [Search PA-X | The University of Edinburgh \(peaceagreements.org\)](https://pax.peaceagreements.org). Any search will result in a list of agreements and options for download. It is possible to reduce the corpus to only the features one is interested in by using the search tool, for example, if only the agreements from a particular region or conflict type are of interest. Selecting no search criteria produces a list of all documents and a .csv file containing all agreement documents in the PA-X collection. The file is in comma-separated format, which is easily readable into R and similar software. However, we do not recommend using Excel to open the file, due to Excel's limits on the number of characters per cell.

The corpus can be merged with the main PA-X quantitative data via agreement ID, which allows further aggregation and merges with other relevant data sources, relying on the country- and conflict codes.

All texts are in English, translated by professional translators or native speakers with expertise in the subject matter.

There are some peculiarities about the data that users should be aware of. While it is common to remove English stopwords, we recommend that users consider further removals that may be useful to them, in particular those words that are very common in peace agreement formats, such as 'shall', 'agreement', 'signed', names of locales, and a range of Arabic and Roman numerals. Additionally, the documents greatly vary in length, and we suggest that this be considered as a relevant factor in research design. Finally, the corpus includes a non-exhaustive segment of local peace agreements, which do not represent a census of local agreements in the way that the remainder of the database is a census of major formal peace agreement documents. Local negotiation practices differ from national- and international level peace processes, and it may in some cases be useful to remove the local agreements from the corpus before conducting analyses.

## Variables

### Con: Country/Entity

The country or jurisdiction in which the conflict originated; where the conflict has a territorial element, the entity to which the conflict relates. This is the case for both intrastate agreements in

intrastate conflict and for interstate agreements in intrastate conflicts, i.e. 'pure' interstate agreements signed as a result of intrastate negotiations in a conflict.

For 'pure' interstate agreements, the states signing or publicly agreeing the agreement are listed.

Special cases:

- Northern Ireland intrastate agreements are always listed as "Ireland/UK/Northern Ireland".
- All agreements related to conflicts in Slovenia, Croatia, Bosnia-Herzegovina, Serbia and Kosovo have been allocated the entity 'Yugoslavia (former)' in addition to the country names. This is in order to reflect the cross-border cooperation between other Yugoslav successor republics and self-declared autonomous republics or non-state actors within Bosnia, and the *complex* conflict dyads throughout Yugoslavia's collapse.
- The country and entity names do not always match the official country names, for reasons related to the particular conflicts and peace processes.

### Contp: Conflict type

Type of conflict the agreement is addressing:

- Government
- Territory
- Government/territory
- Inter-group
- Other

'Government' refers to ideological or political disputes, and 'Territory' to purely territorial disputes.

All intrastate disputes over territory are coded 'Government/territory', as they involve an incompatibility over the territory and over the government and are sometimes resolving by dealing with just one of these issues.

Agreements are coded as 'territorial' even beyond secessionist conflicts, where claims contest or are related to governance of specific regions, even if secessionist or irredentist claims are not made.

Inter-group conflicts are between non-state groups; the agreements pertaining often deal with provisional revolutionary interim governments, e.g. in cases when a government is defeated militarily and the non-state groups 'divide the spoils'.

#### Special Cases:

- Darfur: Due to the strong territorial dimension of the conflict, all Darfur agreements have been coded as government/territory even though the conflict is not a classical secessionist dispute.
- Somalia: Most of the Somalia agreements have been coded as government/territory due to the strong federal dimension of the conflict as it has progressed. Nevertheless, due to the ideological underpinnings of the initial conflict, some of the early Somalia agreements have been coded as 'government' only.

#### PP: Peace Process Number

The number code for the peace process

#### PPName: Peace process Name

Name of peace process.

#### Reg: Region

Region in which the conflict addressed by the agreement is taking place.

Regions listed: Africa (excl. MENA), Americas, Asia and Pacific, Europe and Eurasia, Middle East and North Africa, Cross-regional, Other

#### AgtId: Agreement ID

Unique identifying number for each of the agreements.

#### Ver: PA-X release/version number

Indicates which PA-X release agreement was published in.

#### Agt: Agreement Name

The name of the agreement as found in the text of the agreement. Where appropriate, popular names of the agreements are included in parenthesis.

#### Dat: Date Signed

Listed in the dataset as YYYY-MM-DD

The date on which the agreement was signed or agreed.

Sometimes, particularly in the case of 'declarations' from meetings, the database contains a date period, e.g. 4-8 May 2015. In this case the last date is entered as the date of the agreement.

### Status: Agreement Definition and Status

Character variable, describes the type of agreement.

- Multiparty signed/agreed. Agreement signed or clearly assented to by more than one group of the opposing protagonists (but not necessarily all of the opposed groups). It is irrelevant that a party later recanted, if they previously agreed to the agreement. We also count initial agreement which required further things to happen (e.g. wider referendum etc.) as 'signed/agreed' if there is evidence that the parties had agreed to it, albeit with implementation contingent on future processes.

- This category includes agreements by political leaders (where these are separate from military actors) where there are clear connections and communications and relationships between political leaders and military protagonists and disputes over who really represents the divided groups. Constitutions and Constitutional Amendments which are part of the database are often included on this basis.

- Unilateral agreement: agreement produced by 'one side' but in response to agreement with other side (e.g. as part of an 'exchange of documents' or 'informal choreography'). These also include documents produced by international actors, after talking to both sides, to reflect the 'agreement' between them, or inter-state agreements which *set* the frame for negotiations or provide for their implementation, for example, UN Security Council Resolutions. These 'agreements' reflect agreement between the parties, even though they are not parties to the resolution.

- Status unclear: if it is unclear how much of the agreement was signed, or who agreed it, but there is some indication from surrounding documentation that it was signed/agreed, its status is coded as 'status unclear'. Where entered as status unclear, the database itself includes the reason as part of the description of the agreement. Accessing the database entries for such agreements is required for these clarifications, as they are not part of the database output.

- Agreement with Subsequent Status: a proposed agreement which has been based on talking to the parties and a 'best guess' of what they will agree to which, while not accepted by the parties as an agreement, nonetheless became the basis for subsequent developments, for example: because it formed a 'pre-negotiation exploration' which *set* the frame of later negotiations, or because it was later adopted as a framework for government. These are thus coded as 'agreement with subsequent status'. For these agreements, the "Stage" variable is necessarily coded as "Other", as they often look comprehensive in nature, but in fact operate as 'pre-negotiation' soundings.

NOT INCLUDED:

- Agreement clearly not signed or agreed
- Agreement signed by one side and not the other, which has no clear subsequent status

#### Lgt: Agreement length (pages)

Agreement length, in pages. This is taken from the agreement pdf, and includes any title pages, empty pages, pages with photos, and similar.

#### N\_characters: Agreement length (characters)

Agreement length, by number of characters in agreement text.

#### Agtp: Agreement/conflict type.

This is a character variable, describing the primary nature of the agreement and conflict.

- Inter: Interstate/interstate conflict. Interstate treaty relating to interstate conflict. E.g. Iraq/Kuwait, or North/ South Korea

- InterIntra: Interstate/mixed or intrastate conflict. These agreements are clearly only inter- state in nature (treaty), but rather than referring to a single inter-state conflict, they refer to intra-state conflict/conflicts, whose major components originate within existing (de- facto or legal) state borders. Parties to these are most often states or international actors e.g. Israel-Jordan Treaty in support of the PLO- Israel Peace Agreements; African Great Lakes agreements; European Union unilateral Action Plan on Former Yugoslavia; communiqués, conclusions, and declarations produced by international conferences on Afghanistan. This category enables differentiating such interstate agreements from 'pure interstate agreements', which relate to international conflict.

- Intra: Intrastate agreement relating to intrastate conflict; refers mainly to conflicts within a state's borders, e.g. Northern Ireland. Please note that these may have more than one state party, but if they are not pure inter-state agreements, but have internal parties, then they are coded as 'intrastate agreements'.

- IntraLocal: These are agreements that are related to or an intrastate conflict, but are aiming to resolve local issues rather than what is perceived as a conflict-wide issue. Please note that some of these do not satisfy the 25 battle- related deaths criterion for conflict. The agreements that do not satisfy this criterion can be identified as they do not have an attached UCDP conflict code.

Note on Bosnia: Most agreements relating to the former Yugoslavia are categorised as 'intrastate/intrastate', unless clearly inter-state in nature, relating to intrastate conflict (e.g. European Union unilateral plan on Former Yugoslavia). Whilst acknowledging the declared independence of several republics (Croatia, Bosnia and Herzegovina), to classify inter-former republic agreements as 'inter-state/inter-state' makes a judgement on the timing of statehood and disguises the influence of neighbouring governments over self-declared autonomous regions (e.g. the FRY and the Republic of Serbian Krajina in Croatia). Therefore, we treat these as intra-state secessionist conflicts with reference to the dissolution of former Yugoslavia.

## Stage: Agreement stage

This is a character variable that provides information on the stage of the peace process that the agreement is signed in. We have identified seven main stages of the process that agreements may be part of: pre-negotiation/process; substantive-comprehensive; substantive-partial; implementation/renegotiation; renewal; ceasefire/related; other.

- Pre: Pre-negotiation/process. Agreements that aim to get parties to the point of negotiating over the incompatibilities at the heart of the conflict. Note that these agreements can happen at any time in a process (even after a comprehensive peace agreement has been signed), if parties have moved back into 'talking about how they are going to talk'. Therefore, this characterization does not imply temporal precedence.
- SubPar: Framework-substantive, Partial. Agreements that concern parties that are engaged in discussion and agreeing to substantive issues to resolve the conflict, but only deal with some of the issues in ways that appear to contemplate future agreements to complete.
- SubComp: Framework-substantive, Comprehensive. Agreements that concern parties that are engaged in discussion and agreeing to substantive issues to resolve the conflict and appear to be set out as a comprehensive attempt to resolve the conflict.
- Imp: Implementation/renegotiation. Aiming to implement an earlier agreement. Note that this category does not include ceasefires.
- Ren: Renewal. These are short agreements (typically of just one page), which do nothing other than 'renew' previous commitments. Note that this category does not include ceasefire renewals which are contained in the ceasefire category (see below).
- Cea: Ceasefire/related. This category contains agreements which provide in their entirety for a ceasefire, or association demobilisation, or an agreement that is purely providing a monitoring arrangement for, or extension, of a ceasefire.
- Other: This is a residual category, capturing all agreements that do not fit the remaining definition. The user is advised to refer to the database for examples of the agreements in this category.

## StageSub: subcoding of stage

Each of the categories in the previous variable is split into subcategories. This means that the subcategories are category-specific. Only agreements in, for example, ceasefire category can be classified into a ceasefire subcategory.

Pre-negotiation/process subcategories:

- Proc: Process; agreement sets out a process for talking or having meetings.
- Prin: Principles; sets out principles for talks, or the main issues for talks.

- Conf: Confidence-building measure; deals with a specific issue that even if substantive is related to the *object* of getting people into talks (e.g. opening an airport, holding a particular inquiry, or releasing particular prisoners)
- PreMix: Mixed; a mixture of process, principles, and/or confidence building measures.
- PreOth: Other; any other pre-negotiation or process oriented subject matters.

#### Framework/substantive partial subcategories:

- Iss: Core issues – deals with a core issue, e.g. human rights, or powersharing
- MultIss: Partial but multiple issues – clearly not comprehensive, but deals with a range of issues rather than one issue
- FrparOth: Other partial framework - another form of framework/substantive agreement that is not comprehensive and that does not fit in the other categories.

#### Framework/Substantive Subcategories

- FrAg: Comprehensive – purports or appears to be comprehensive, as laying out a broad framework for resolving the conflict (note that no agreement is entirely comprehensive, as they all leave some matters to future processes or for development); this group does not include constitutions
- FrCons: Constitution - document that operates as a comprehensive interim or final constitution in name or function.

#### Implementation/renegotiation subcategories:

- ImpMod: Implementation modalities – establishes implementation modalities, e.g. mechanisms or timetables etc. for implementing commitments, or if an inter-state agreement – mechanisms for support for implementation (e.g. peace-keeping).
- ExtSub: Substantive Extending - extends agreement to new or outstanding issues, extends the agreement to incorporate any new issues that have arisen
- ExtPar: Partial Extending - extends agreement to new parties, extends past commitments to a new group
- ImpOth: Other – any other agreement that predominantly addresses implementation but does not fit into any other category

#### Renewal subcategories:

- Reimp: Renewal of an implementation agreement
- Repr: Renewal of a pre-negotiation agreement
- Resub: Renewal of an agreement dealing with substantive issues



- Reoth: Renewal of other type of agreement

Note that ceasefire renewals are treated as a subcategory of ceasefire agreements.

Ceasefire/related sub-categories

- Ceas: ceasefire agreement, mainly or wholly providing for a ceasefire, including issues such as definitions of violations.

- Rel: Ceasefire-related - agreements that are related to ceasefires, for example, monitoring or extension arrangements

- CeaMix: Ceasefire-mixed - agreement that combines ceasefire and provisions related to or touching on broader issues of process.

### Part: Parties

The parties who directly sign the agreement, often the parties to the conflict and their regional allies. This was coded in 'cut and paste' form, exactly as written in the original peace agreement.

Some agreements are not signed, for example, in a 'National Dialogue' forum, or where the 'agreement' is in fact a joint declaration. These cases are noted, and the parties are listed, for example: 'Not signed, agreement mentions the following parties as having produced it: [text from agreement]'

Finally, some agreements just mention that the participants are a generic group, e.g. 'for the participants in the UN talks on Afghanistan'. In this case, the names and positions of the people who have actually signed the document are noted.

Constitutions are a special case: where signed, the signatories are listed; where unsigned – this is noted and the political body/bodies which passed/ratified the constitution is stated.

### ThrdPart: Third Parties

Parties who sign the agreement as third parties, such as observers, guarantors, or witnesses; these can be states, international organizations, and nongovernmental organizations.

The actual term used for the third party is entered in the database with a colon: and the name of the party. In other words, the name of the third party is preserved exactly as written in the original

peace agreement. For example, 'Agreement witnessed by: H.E. Lt. Gen. Riek Machar Teny-Dhurgon (PHD, Vice President of Southern Sudan).'

### OthAgr: Other Agreement

Names of any other agreements that the document in question mentions, as found in the text.

### Loc1ISO

Indicates primary location of intrastate conflict and returns the ISO country designation. In cases of interstate conflict, indicates one of the two main locations, assigned randomly to location 1 or location 2. The fields are left empty in cases of agreements that deal with multiple locations and conflicts, such as the Great Lakes Agreements.

### Loc2ISO

Only used in cases of interstate conflict, indicating one of the primary locations of interstate conflict, and returning the ISO country designation. The two main locations are assigned randomly to location 1 or location 2. The exception to this common usage of the variable is in cases of local agreements which deal with cross-border communities and local peace practices therein. In these cases, the locations of both countries are noted, with random assignment of each to Location 1 or Location 2.

### Loc1GWNO

Indicates primary location of intrastate conflict and returns the Gleditsch-Ward country code. In cases of interstate conflict, indicates one of the two main locations, assigned randomly to location 1 or location 2. The fields are left empty in cases of agreements that deal with multiple locations and conflicts, such as the Great Lakes Agreements.

For more on GWNO codes, see: Gleditsch, Kristian S., and Michael D. Ward. 1999. "Interstate System Membership: A Revised List of the Independent States since 1816." *International Interactions* 25(4): 393–413.

### Loc2GWNO

Only used in cases of interstate conflict, indicating one of the primary locations of interstate conflict, and returning the Gleditsch-Ward country code. The two main locations are assigned randomly to location 1 or location 2. The exception to this common usage of the variable is in cases of local agreements which deal with cross-border communities and local peace practices therein. In these cases, the locations of both countries are noted, with random assignment of each to Location 1 or Location 2.

For more on GWNO codes, see: Gleditsch, Kristian S., and Michael D. Ward. 1999. "Interstate System Membership: A Revised List of the Independent States since 1816." *International Interactions* 25(4): 393–413.

### UcdpCon

Uppsala Conflict Data Program Conflict ID. Note that the Conflict IDs applied are from 2017 and onward, after the UCDP introduced a new system of conflict designation.

It is possible that this field is empty this happens in cases of conflicts not registered in the UCDP conflict data.

For more information on the UCDP data and conflict codes and definition see:

Pettersson, Therése, and Kristine Eck. 2018. "Organized Violence, 1989–2017." *Journal of Peace Research* 55(4): 535–47. Also see <http://ucdp.uu.se/downloads>

### UcdpAgr

Uppsala Conflict Data Program peace agreement ID. For more information on these data, see:

Högbladh, Stina. 2011. "Peace Agreements 1975-2011 - Updating the UCDP Peace 21 Agreement Dataset Peace in the UCDP Dataset." *States in Armed Conflict 2011: Uppsala University: Department of Peace and Conflict Research Report 99*: 39–56.

Pettersson, T., Högbladh, S. and Öberg, M. (2019) 'Organized violence, 1989–2018 and peace agreements', *Journal of Peace Research*, p. 002234331985604. doi: 10.1177/0022343319856046.

Also see <http://ucdp.uu.se/downloads/>

### PamAgr

Kroc Institute's Peace Accords Matrix agreement ID. It is possible that this field is empty – this happens in cases of agreements not part of the PAM data. The PAM focuses on comprehensive agreements only, so the bulk of PA-X documents do not have the PAM agreement code.

For more information on PAM, see:

Joshi, Madhav, Jason Michael Quinn, and Patrick M Regan. 2015. "Annualized Implementation Data on Comprehensive Intrastate Peace Accords, 1989–2012." *Journal of Peace Research* 52(4): 551–62. <http://journals.sagepub.com/doi/10.1177/0022343314567486>.

Also see <https://peaceaccords.nd.edu/research>

### CowWar

Correlates of War: War Number. Many PA-X agreements come from peace processes in conflicts that do not meet the high battle-related deaths criterion of the Correlates of War project, so the war numbers are not provided for many of the agreements.

For more information on the Correlates of War war list, see:

Sarkees, Meredith Reid, and Frank Whelon. Wayman. 2010. Resort to War : A Data Guide to Inter-State, Extra-State, Intra-State, and Non-State Wars, 1816-2007. Washington D.C.: CQ Press. [https://www.worldcat.org/title/resort-to-war-a-dataguide-to-inter-state-extra-state-intra-state-and-non-state-wars-1816-2007/oclc/768956479&referer=brief\\_results](https://www.worldcat.org/title/resort-to-war-a-dataguide-to-inter-state-extra-state-intra-state-and-non-state-wars-1816-2007/oclc/768956479&referer=brief_results) (March 30, 2018).

Also see <http://cow.dss.ucdavis.edu/>

### Interim: Interim Arrangements

The variable returns three categories: Y, N, or Maybe.

If 'Y', the agreement is considered to be putting forward an interim, temporary, transitional institutional arrangement, in line with the definition provided by International IDEA:

"Interim governance arrangements are an institutional framework established to create a 'bridge' from a situation of governance that has been interrupted by political or violent crisis often situated in an authoritarian past, towards a more peaceful, inclusive and democratic government." If 'N', the agreement is not part of an interim arrangement. The third option, 'Maybe', is used very sparingly, for cases that are ambiguous.

See: Interim Governance Arrangements in Post-Conflict and Fragile Settings by International IDEA, at <https://www.idea.int/publications/catalogue/interimgovernance-arrangements-post-conflict-and-fragile-settings>

### Agreement text

Full text of the agreement.