# International Systems Dataset Version 2.2: Codebook and Best Use Instructions

## Key Definitions

### Identifying states

A state is a recognizable political entity that possesses:

1. A population of at least 10,000 (previously 100,000)
2. Autonomy over a specific territory
3. Sovereignty that is either uncontested or acknowledged by the relevant international actors.

In our conception, a sovereign state includes both a domestic and international dimension. Domestically (or internally), it has a political apparatus that is hierarchic, force-wielding, and in control over a given territory or set of places. Internationally (or externally), the state is recognized (formally or informally) by other states and it has control over its foreign policy.

The first criterion regarding population size (10,000) is simply an arbitrary number used to identify a population minimum. This is for practical reasons as smaller states are harder to research in general. Moreover, smallish entities are gradually less viable as states. Although any population threshold is, in a sense, arbitrary, we chose 10,000 because it is focal, because it was not too small to hinder our data-collection efforts, and because it provided consistency across the entire 1816–2016 period given that many UN member states have populations that barely exceed(ed) 10,000.

- The first version of the ISD used 100,000 as a population threshold.
- The coding template provides an opportunity to identify polities between the 10,000 and 100,000 limits. Please code states if they fall within this range.

The second criterion regarding autonomy over a specific territory is meant to filter fictional states, governments in exile, etc. We are looking for a hierarchical, administrative infrastructure with the right and capacity to mobilize coercion, extract some taxes (including, for example, tributes) and administer some areas of law-making and jurisprudence. This should be over a stable territory or set of places. The state need not possess clearly defined borders; many states had fuzzy boundaries. Furthermore, it need not have control over all areas of governance, and, in practice, there is significant variation within states in terms of how decision-making powers are distributed. For example, the Oyo Empire in 19th-century West Africa controlled little more than external relations, leadership succession, and judicial authority over capital crimes in relation to its vassals, while they retained wide-ranging autonomy over decision-making in relation to taxation methods, policing, and local law-making

The third criterion detects recognition by the "relevant actors" for that particular state. Inevitably, this decision has a subjective character to it and notes are recorded in the coding narratives. We define the relevant international actors as those states that, for all practical purposes, are the key legitimizers of the state in question. A regional

approach also allows us to be sensitive to varying patterns of order and local systems of recognition. We require that no other state has the formal right to determine the polity's external affairs. If we found evidence that neighboring states or major powers in the region recognized a state as independent, then we treated it as independent. Of course, this can be ambiguous in cases where other states gradually take control of another state's external affairs. In such cases, we recorded a state exit (death) when the larger state and other relevant actors began to treat the state in question as nonsovereign

- Some states may only possess low levels of diplomatic intercourse with other states and live relatively hermetic lives. But if their sovereignty is uncontested we identify them as a state.
- When contestation does exist it is usually a former metropole or conquering party that disputes the territory and contests the sovereignty.
- For cases of secession, we impute sovereignty if the state acquires tacit recognition from its metropole. If, however, the metropole continues to contest the secession, a start is coded if conflict ceases for at least one year and the remaining relevant international actors accept the state. Officially, the start should be backdated to the cessation ofconflict and recognition by other states.
- Following Fazal, a state dies (exits) when it loses formal control over foreign policy to another state. This rule holds even if the state later regains its independence (e.g. France in 1940). This can be difficult to code in cases where a larger power gradually takes de facto control over the state in question. In such cases, code for a state death when the larger state and relevant actors begin to treat the state as non-sovereign.

## Variables and Definitions

### COWNum

The number assigned to the political unit in the Correlates of War (COW) dataset. For new cases, a COWNUM was created.

### COWID

A three-letter abbreviation given by the COW dataset. For new cases, a COW ID was created.

### StateName

The name used for the state in the ISD dataset.

### Other Name

Other names by which the state is known.

**ISD_Category**

1 = states have same start and end years as ISD V1

2 = states were in ISD V1 but have changed start and/or end years

3 = states are new states that were not in ISD V1

**Region**

The COW index number for the geographic region in which the state is located.

1. Africa
2. South and Central Asia
3. East and Southeast Asia (including Oceania)
4. Americas
5. Eastern and Western Europe
6. The Middle East

**Start**

The year the state entered the system – i.e. met all of the criteria for statehood listed above. For states that existed at the beginning of the time period for the dataset, 01-01-1816 is used as the start.
Format: dd-mm-yyyy

**Start_Am**

Was there ambiguity in the coding of the start?

1 = Yes

0 = No

**EStart**

What year did the state start if earlier than 1816?
Format: yyyy

**EStart_Am**

Was there ambiguity regarding the early start?

1 = Yes

0 = No

**Declare****

Is there evidence of a declaration of independence? This can be a pronouncement or a document. By listing "1" for this variable you are directing the future researcher to a case where something occurred.[1]

1 = Yes

0 = No

**DecDate****

Format: dd-mm-yyyy Date of the declaration of independence (year and date provided if possible).

**Population**

Population for the start year or as close as possible to it.

**PopDate**

Year population was counted.

**Pop_Am**

Ambiguity surrounding whether the state crossed the 10,000 population threshold?

1 = Yes

0 = No

**PopulationHigh**

Upper end of population estimates

**PopulationLow**

Lower end of population estimates

**StartType**

How was the state created?

1 = Territorial Consolidation: state created over space where none previously existed.

2 = Secession: state separates from larger state.

3 = Dissolution: state is born as a fragment of a larger state that broke apart and ceased to exist (e.g. Austro-Hungarian Empire).

[1]** Indicates variables that were more experimental. -9 is used throughout the dataset to indicate "Not available or relevant"

4 = Unification: state born from the voluntary merging of several (typically equally sized) states.

5 = Decolonization: state born from decolonization

6 = Other (please provide information in case narrative)

**StartType_Am**

Was there ambiguity on the start type?

1 = Yes

0 = No

**StartSettle**

Was there a treaty or formal settlement attending the birth of the state?

1 = Yes

0 = No

**End**

Year the state exited the system
Format: dd-mm-yyyy

**End_Am**

Was there ambiguity on the end date?

1 = Yes

0 = No

**EndType**

How did the state end?

1 = Colonization: Imperial, non-contiguous colonization.

2 = Conquest / annexation: State death via foreign take-over.

3 = Unification: State dies through process of voluntary unification.

4 = Dissolution: State death via dissolution of state.

5 = Partition: State death via partition by outside powers.

6 = Other: (please provide information in case narrative)

**EndType_Am**

Was there ambiguity on the end type?

1 = Yes

0 = No

**EndSettle**

Was there a treaty or formal settlement attending the death of the state?

1 = Yes

0 = No

**Sovereignty_Am**

Was there ambiguity regarding the criterion that the sovereignty of the state is either uncontested or acknowledged by the relevant international actors.

1 = Yes

0 = No

**EuroDip****

Is there evidence that the state had diplomatic relations with a European power? If so, provide information (and citations) in the case narrative.

1 = Yes

0 = No

**Borders****

Is there evidence that the state had clearly defined borders separating it from other states or sources of authority?

1 = Yes

0 = No

**Borders_Am**

Was there ambiguity on the issue of borders?

1 = Yes

0 = No

**Latitude**

Capital latitude coordinates (Degrees Minutes Seconds)

**Longitude**

Capital longitude coordinates (Degrees Minutes Seconds)

**Capital**

Name of the most important capital, used to calculate latitude and longitude coordinates

**VioStart****

Was the state born violently? Violence here is defined as the killing of civilians and combatants.

1 = Yes

0 = No

**VioStart_Am**

Is their ambiguity about the starting violence?

1 = Yes

0 = No

**VioEnd****

Did the state die violently? Violence here is defined as the killing of civilians and combatants.

1 = Yes

0 = No

**VioEnd_Am**

Is their ambiguity about the ending violence?

1 = Yes

0 = No

**DestinationA − > DestinationU**

COW numbers for the territories of modern states that this state was located in.

## Broad Methodological Approach

Our data collection efforts were conducted over a five-year period from 2016 to 2021, with a significant update in 2022 and 2025. The team consisted of two Principal Investigators (PIs) and 13 Research Assistants (RAs) who were doing doctoral or MA work in one of the social sciences. We began by consulting a variety of secondary historical sources to identify states and code their features over the 1816–2016 period. Key sources for the first sweep of cases included the Cambridge History Series, the UNESCO History of Africa Series, the General History of Africa series, and the Correlates of War Colonial Entities List. Once a potential case was identified, RAs consulted secondary sources relating to those cases to establish whether the case met our criteria for statehood. Coding these first cases often led to the discovery of new states and the process snowballed until we concluded that we had saturated the area for all states meeting our criteria.

Once a state was added to our list, a number of additional variables were coded (see above). Coders were assigned cases based on regions, which led to the development of expertise and familiarity with sources related to the region. Most states in the data were independently coded by at least two coders. Each state included in the dataset has a separate case narrative that provides a description of the state and outlines the sources used to reach coding decisions. The final list was produced after we, the PIs, read each case narrative and coding, and reconciled this information with the coding criteria to ensure the consistency of the data in relation to the variable operationalization across coders, regions, and time periods. Each variable has an ambiguity score of 0 or 1 that reflects whether a coder thought that a data point was highly uncertain due to conflicting sources or scarce information.

We refer users to the 2020 article:

*Butcher, C., & Griffiths, R. D. (2020). States and their international relations since 1816: introducing version 2 of the International System (s) Dataset (ISD). International Interactions, 46(2), 291-308.*

And our book (especially Chapter 3):

*Butcher, C., & Griffiths, R. D. (2025). Before Colonization: Non-Western States and Systems in the Nineteenth Century. Columbia University Press.*

## Best Use Guidelines

The ISD are designed, and best used, for exploring and analyzing broad patterns in statehood across time and regions. Although we have invested considerable time and resources into identifying entities that cross our threshold of statehood, the data remain a work in progress. We know that there will be other political entities that would meet our definition of statehood that are not currently included in the dataset. As we continue to work on the project, we will also continue to update our state list. We always welcome feedback and suggestions from users about how we might improve the dataset. Please contact Charles Butcher for suggestions and feedback.

We are also not taking a normative or teleological position on statehood. Put dierently, we are not taking a position on whether having a state is good or bad or whether states are a "higher" development of political organization. Rather this is an empirical project that aims to identify states to enable researchers to understand changes in the international

system over time, their features, or the longer-term legacies of historical state systems. We do not code states in, for example, Australia or Papua New Guinea in 1816, but this does not mean there is anything deficient about governance in these areas. We refer users to our book *Before Colonization: Non Western States and Systems in the Nineteenth Century* where we argue, and show, that governance systems across numerous regions were often ecient adaptations to local environmental, social, and cultural conditions. As such our data should also not be used to adjudicate on the legitimacy of claims to prior statehood or the legitimacy of claims to self-determination.