

# Codebook: Domestic LGBT+ NGO Database

Siddhartha Baral

2026-05-15

## Contents

<b>1</b>	<b>General Overview</b>	<b>2</b>
1.1	Inclusion Criteria . . . . .	2
1.2	Original Assembly of LGBT+ Organizations . . . . .	3
1.3	RA Verification and Expansion . . . . .	3
1.4	Limitations . . . . .	3
<b>2</b>	<b>Dataset Overview</b>	<b>4</b>
<b>3</b>	<b>Variable Documentation</b>	<b>5</b>
3.1	Group 1: Organization identificatory variables . . . . .	5
3.1.1	organization . . . . .	5
3.1.2	organization_alt_name . . . . .	6
3.1.3	country . . . . .	6
3.1.4	region . . . . .	7
3.2	Group 2: Organization (Self-)description . . . . .	8
3.2.1	mission_en . . . . .	8
3.2.2	mission_native . . . . .	8
3.2.3	about_en . . . . .	9
3.2.4	about_native . . . . .	9
3.2.5	language . . . . .	9
3.3	Group 3: Temporal Presence . . . . .	10
3.3.1	year_first_record . . . . .	10
3.3.2	source_first_record . . . . .	11
3.3.3	link_source_first_record . . . . .	12
3.3.4	year_last_record . . . . .	12
3.3.5	source_last_record . . . . .	12
3.3.6	link_source_last_record . . . . .	13

3.4	Group 4: Online Presence & Contact . . . . .	13
3.4.1	online_presence . . . . .	13
3.4.2	website_link . . . . .	14
3.4.3	social_media_link . . . . .	14
3.4.4	org_email . . . . .	14
3.5	Group 5: Leadership . . . . .	15
3.5.1	leader_name . . . . .	15
3.5.2	leader_role . . . . .	15
3.5.3	leader_contact . . . . .	16
3.6	Group 6: Organizational Status . . . . .	16
3.6.1	independent_org . . . . .	16
3.6.2	umbrella_org . . . . .	17
3.6.3	issue_focus . . . . .	17
3.7	Group 7: Sources & Verification . . . . .	18
3.7.1	source_1 . . . . .	18
3.7.2	source_2 . . . . .	18
3.7.3	both_sources . . . . .	19
3.8	Group 8: Research Process Metadata . . . . .	19
3.8.1	drop_flag_raw . . . . .	19
3.8.2	notes . . . . .	20
4	Missingness Across All Variables . . . . .	20

---

# 1 General Overview

The Domestic LGBT+ NGO Database is the largest cross-national dataset of lesbian, gay, bisexual, and transgender (LGBT+) non-governmental organizations operating at the domestic level. The dataset was assembled by Dr. Kristopher Velasco and Dr. Tara Gonsalves, the Principal Investigators (PIs) of the project. The goal of the dataset is to provide a systematic, large-scale inventory of domestic LGBT+ civil society organizations that can support comparative research on LGBT+ organizing across countries and over time. At the time of cleaning, the dataset contains 9,150 organizations spanning 218 countries.

## 1.1 Inclusion Criteria

For inclusion in the dataset, an organization had to meet three criteria. First, working on behalf of LGBT+ people had to be a primary aim of the organization. Second, the organization had to be independent — LGBT+ caucuses or interest groups embedded within broader unions, political parties, government, or other workplaces were excluded. Third, the organization had to be non-profit — LGBT+ businesses were therefore excluded. And fourth, the geographical focus of the organization had to be domestic — those that operate within a single country — as opposed to international or transnational LGBT+ bodies.

## 1.2 Original Assembly of LGBT+ Organizations

The dataset was built in two steps. In the first step, the principal investigators assembled a preliminary list of domestic LGBT+ NGOs by consulting over 200 sources across nine broad categories. First, membership rosters were retrieved from the annual reports and websites of LGBT+ international non-governmental organizations (INGOs) identified from the Yearbook of International Organizations, covering as many years as could be found. This source alone proved highly productive: ILGA, the leading LGBT+ INGO, has over 1,000 member organizations (ILGA 2020). Second, government databases of registered charities were searched using keywords drawn from a pre-established set used to identify LGBT+-related content in newspaper archives. Third, the Encyclopedia of Associations was consulted to identify additional organizations (Schofer and Longhofer 2010). Fourth, directories and grantee lists produced by NGOs and the United Nations were incorporated. These included ILGA’s online directory of thousands of LGBT+ organizations worldwide; grantee records from Mama Cash, the Arcus Foundation, and the Astraea Foundation (sourced from their websites and IRS Form 990 tax filings); country-specific reports from the UN Development Programme’s “Being LGBT+ in Asia and the Pacific” project, each of which included lists of domestic LGBT+ organizations; and signatory lists from NGO-sponsored UN resolutions, including two statements produced by ARC International in 2006 and 2014 that were collectively signed by 1,123 organizations. Fifth, membership rosters of domestic umbrella organizations were incorporated where available — for example, the Consortium of LGBT+ Voluntary and Community organizations in the United Kingdom, which has over 200 members. Sixth, academic books and articles on LGBT+ activism were consulted, with particular attention to historically oriented sources such as the Greenwood Encyclopedia of LGBT+ Issues Worldwide. Seventh, proper nouns were extracted from a large corpus of newspaper data using text analysis techniques; after removing individual names and geographic locations, the remaining list was manually inspected to identify references to LGBT+ organizations, a method that was particularly effective for identifying smaller organizations that appeared only in local news or were more active in earlier time periods. Eighth, the websites and social media of already-identified organizations were searched for references to other groups, typically listed under “Partners” or “Alliances” sections. Finally, country-specific Google searches were conducted to identify any remaining organizations not captured by the preceding steps.

## 1.3 RA Verification and Expansion

In the second step, a team of research assistants — each assigned responsibility for a specific set of countries — systematically reviewed every organization in the preliminary list. Assistants verified and updated existing information (such as the year of first record and online presence), and added information not present in the original dataset (such as organizational contact details, leadership information, and the year of last record). New organizations were also added through two channels: country-specific searches conducted during the RA review process, and new organizational names gathered through a cross-national survey of LGBT+ NGOs that drew on an earlier version of this dataset. Data collection took place across multiple waves between 2023 and 2026. Because collection was distributed across a large team over an extended period, some variation exists in how template fields were interpreted and completed — particularly in free-text fields. The cleaning and coding decisions documented in this codebook were made to harmonize these differences.

## 1.4 Limitations

The dataset is subject to several limitations inherent in its construction. The reliance on internet-accessible and digitally archived sources introduces a systematic bias toward organizations with a digital footprint, and likely undercounts organizations in regions with lower internet penetration or in earlier historical periods before widespread digitization. Moreover, organizations that are more recently established (such as in 2025 or 2026) and are yet to establish much digital footprint may also escape being captured in this dataset. Additionally, some LGBT+ organizations may deliberately avoid public identification for reasons of safety or political context, and these organizations will not be captured. However, because most comparative research on LGBT+ civil society is concerned with organizations that have some broader national movement

impact — and, therefore, notable direct or indirect digital presence — this approach is argued to reasonably capture the universe of organizations relevant to most research questions. Note that despite our best efforts, variables in the dataset come with different degrees of missingness, a fact we document in the codebook systematically for transparency and better understanding of the extant data. Finally, we did not collect any data for the United States.

---

## 2 Dataset Overview

For each variable the codebook provides:

- **Label** — in brief, what the variable captures
- **Description** — what the variable captures in more detail, and how to interpret it
- **Coding notes** — for variables recoded from free text, the decision rules applied
- **Statistics** — missingness for free-text and URL fields; frequency tables for categoricals; min/median/mean/max for numeric variables

## Total organizations: 9147

## Total variables: 30

Table 1: organizations by region

Region	N	%
Europe	3906	42.7
Latin America and Caribbean	1775	19.4
Asia	858	9.4
NA	838	9.2
Africa	754	8.2
North America	526	5.8
Oceania	469	5.1
Global/Multinational	13	0.1
Other	6	0.1
Middle East	2	0.0

Table 2: Top 20 countries by number of organizations (218 countries total)

Country	N	%
United Kingdom	1148	12.6
Canada	559	6.1
Germany	427	4.7
Brazil	377	4.1
Australia	361	3.9
Argentina	287	3.1
Spain	284	3.1
France	281	3.1

Country	N	%
Mexico	277	3.0
Italy	258	2.8
Colombia	245	2.7
Belgium	241	2.6
India	239	2.6
Netherlands	208	2.3
Philippines	158	1.7
Kenya	127	1.4
New Zealand	120	1.3
South Africa	111	1.2
Ecuador	107	1.2
Uganda	102	1.1

The map below shows the geographic coverage of the dataset. Countries are shaded by the number of organizations recorded; grey indicates countries with no organizations in the dataset. A square-root colour scale is used to reveal variation across both low-count and high-count countries simultaneously.

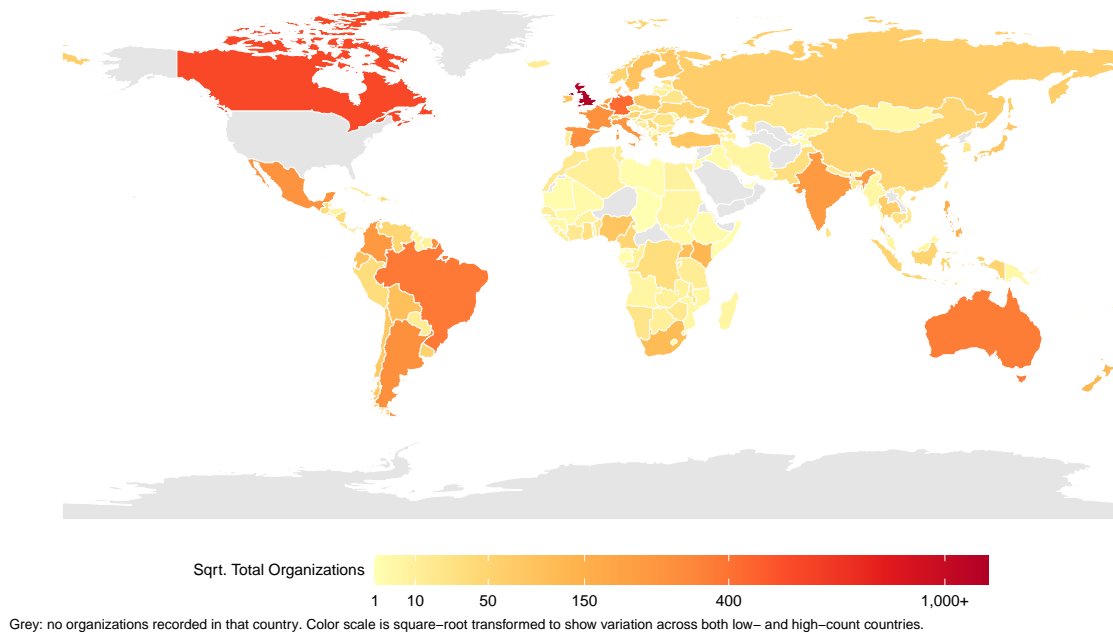


Figure 1: Geographic coverage of the Domestic LGBT+ NGO Database. Color intensity reflects organization count on a square-root scale. Grey countries have no organizations in the dataset.

### 3 Variable Documentation

#### 3.1 Group 1: Organization identificatory variables

##### 3.1.1 organization

**Label:** Name of the organization

**Description:** The primary name of the organization. Names were identified through the PI-led assembly process, which drew on numerous sources described earlier, including INGO membership rosters, government databases, NGO directories, newspaper archives, and web searches. During the subsequent RA verification phase, names were cross-checked against the organization’s own materials where possible (website, social media, official registration). No standardisation of capitalisation or transliteration has been applied; names appear as collected.

**Type:** String / free text

---

	N	%
Non-missing	9147	100
Missing	0	0
Total	9147	100

---

### 3.1.2 organization\_alt\_name

**Label:** Alternative name of the organization

**Description:** An alternative or former name for the organization, such as a name in a different language, an acronym, a former legal name, or a commonly used informal name. Populated during the RA verification phase where a meaningful alternative could be identified.

**Type:** String / free text

---

	N	%
Non-missing	812	8.9
Missing	8335	91.1
Total	9147	100.0

---

### 3.1.3 country

**Label:** Country the organization belongs to

**Description:** The country in which the organization primarily operates. Country was recorded as part of the original PI-led organizational data assembly; research assistants were subsequently assigned to specific countries to conduct verification and expansion. The dataset spans 218 unique countries.

**Type:** String / categorical (open)

Table 5: Top 25 countries (218 countries total)

---

	N	%
Non-missing	9112	99.6
Missing	35	0.4
Total	9147	100.0

---

Country	N	%
United Kingdom	1148	12.6
Canada	559	6.1
Germany	427	4.7
Brazil	377	4.1
Australia	361	3.9
Argentina	287	3.1
Spain	284	3.1
France	281	3.1
Mexico	277	3.0
Italy	258	2.8
Colombia	245	2.7
Belgium	241	2.6
India	239	2.6
Netherlands	208	2.3
Philippines	158	1.7
Kenya	127	1.4
New Zealand	120	1.3
South Africa	111	1.2
Ecuador	107	1.2
Uganda	102	1.1
Bolivia	101	1.1
Sweden	96	1.0
Ireland	95	1.0
Nigeria	87	1.0
Japan	81	0.9

### 3.1.4 region

**Label:** Region the organization belongs to

**Description:** The world region in which the organization operates. Derived from the raw `region` field entered by research assistants during the verification phase, which was standardised and consolidated into eight categories during cleaning.

**Type:** String / categorical (coded)

**Coding notes:** Raw region values were free text and inconsistent across researcher sheets (some recorded sub-regions, some city names, some used non-standard labels). The cleaning script extracted the most specific region component and mapped it to the following eight standard categories. Values that could not be assigned to a category (e.g. “DK”, ambiguous entries) are coded as missing.

Code	Meaning
Africa	Sub-Saharan and North Africa
Asia	East, Southeast, South, and Central Asia
Europe	All European countries including UK
North America	Canada
Latin America and Caribbean	Mexico, Central America, South America, Caribbean
Oceania	Australia, New Zealand, Pacific Islands
Middle East	Middle East and North Africa (where not coded Africa)
Global/Multinational	Explicitly cross-national organizations

---

Code	Meaning
------	---------

---

Value	N	%
Europe	3906	42.7
Latin America and Caribbean	1775	19.4
Asia	858	9.4
(Missing / NA)	838	9.2
Africa	754	8.2
North America	526	5.8
Oceania	469	5.1
Global/Multinational	13	0.1
Other	6	0.1
Middle East	2	0.0

---

## 3.2 Group 2: Organization (Self-)description

### 3.2.1 mission\_en

**Label:** Mission statement of the organization in English

**Description:** The organization’s formal mission statement in English, as found on its website, social media, or other official materials. Where the original language was not English, research assistants sourced an English-language version from the organization’s own bilingual materials or provided a translation. This field is free text and has not been standardised. High missingness (52%) reflects organizations for which no mission statement could be located.

**Type:** String / free text

	N	%
Non-missing	4335	47.4
Missing	4812	52.6
Total	9147	100.0

---

### 3.2.2 mission\_native

**Label:** Mission statement of the organization in native language

**Description:** The organization’s mission statement in its native (non-English) language. Only populated where the native-language version was available and distinct from the English version. High missingness reflects both English-speaking organizations (where no native version applies) and organizations for which no mission statement was found.

**Type:** String / free text

---

	N	%
Non-missing	1628	17.8
Missing	7519	82.2
Total	9147	100.0

---

### 3.2.3 about\_en

**Label:** About us statement of the organization in English

**Description:** A general “About Us” description of the organization in English, typically drawn from an About page, homepage bio, or similar source, distinct from mission statement. More descriptive and contextual than the mission statement, and may include founding history, programme areas, or membership information. Free text, not standardised.

**Type:** String / free text

---

	N	%
Non-missing	6437	70.4
Missing	2710	29.6
Total	9147	100.0

---

### 3.2.4 about\_native

**Label:** About us statement of the organization in Native language

**Description:** The native-language equivalent of the About Us description. Only populated where the native-language version was available and distinct from the English version.

**Type:** String / free text

---

	N	%
Non-missing	3093	33.8
Missing	6054	66.2
Total	9147	100.0

---

### 3.2.5 language

**Label:** Native language

**Description:** The primary language in which the organization operates, as identified during the RA verification phase. Reflects the language of the country or region in which the organization primarily works. 256 unique language values are present.

**Type:** String / categorical (open)

Table 13: Top 20 languages (languages with fewer than 10 entries not shown)

	N	%
Non-missing	4278	46.8
Missing	4869	53.2
Total	9147	100.0

Language	N	%
NA	4869	53.2
English	1244	13.6
Spanish	1060	11.6
French	400	4.4
Portuguese	246	2.7
Dutch	179	2.0
German	173	1.9
Italian	78	0.9
Swedish	50	0.5
Russian	37	0.4
Finnish	34	0.4
Japanese	29	0.3
DK	26	0.3
Polish	25	0.3
Turkish	22	0.2
Danish	21	0.2
English/Chinese	21	0.2
Norwegian	20	0.2
Croatian	19	0.2
Thai	19	0.2

---

### 3.3 Group 3: Temporal Presence

---

#### 3.3.1 `year_first_record`

**Label:** Year of first record

**Description:** The earliest year for which a record of the organization’s existence could be identified. If the year of first record derived from sources consulted during the PI-led assembly phase (such as INGO membership rosters from specific years, newspaper archives, or government registration databases), the RAs double-checked the value during the verification phase, and if missing, the RAs added the year through their own search. In case of conflict, RAs entered the year value they determined to be more accurate. While this variable may serve as a reasonable proxy for founding year, it may not necessarily capture the actual founding year. The precise meaning varies by the type of source, which is captured in `source_first_record`.

**Data quality note:** A small number of implausible values (years before 1940 or after 2026) were identified during cleaning. These have not been removed from the dataset.

**Type:** Numeric (integer year)

---

Statistic	Value	%
Non-missing	7223.0	79
Missing	1924.0	21
Minimum	202.0	NA
Median	2010.0	NA
Mean	2006.1	NA
Maximum	2026.0	NA

---

### 3.3.2 source\_first\_record

**Label:** Source for the year of first record

**Description:** The type of source from which `year_first_record` was obtained. Original entries were free-text responses; these have been coded into six standardised categories during cleaning (see coding notes below).

**Type:** String / categorical (coded)

**Coding notes:** Free-text source descriptions were assigned to one of six categories based on the type and provenance of the source. Where entries were ambiguous, unclear, or explicitly marked as unknown (“DK”, “no information”), the value was set to missing. The coding prioritises source specificity — for example, a social media registration date that explicitly belongs to the organization is coded as “Official organizational source”, not “Social media”.

---

Code	Description
Official organizational source	organization’s own website, About/History page, social media profile, or other materials directly managed by the organization
Government/legal registry	Formal government registration records, charity commission databases, incorporation filings, or similar official legal records
Third-party documentation	News articles, reports, directories, databases, or other materials produced by parties external to the organization
Network/coalition listing	LGBT+Q+ network member lists or umbrella organization directories (e.g. ILGA, IGLYO, TGEU)
Archives/academic	Archival sources, academic publications, Wikipedia, or Internet Archive/Wayback Machine snapshots
Other	Edge cases not classifiable into the above categories

---

---

Value	N	%
Official organizational source	5700	62.3
(Missing / NA)	1852	20.2
Third-party documentation	884	9.7
Other	419	4.6
Government/legal registry	120	1.3
Archives/academic	107	1.2
Network/coalition listing	65	0.7

---

### 3.3.3 link\_source\_first\_record

**Label:** Link for the source for the year of first record

**Description:** A URL linking directly to the source used to establish `year_first_record`. During cleaning, any cell not containing a valid URL (i.e. not containing `http`, `.com`, or `.org`) was set to missing. This field allows manual verification of first-record dates.

**Type:** String / URL

---

	N	%
Non-missing	6473	70.8
Missing	2674	29.2
Total	9147	100.0

---

### 3.3.4 year\_last\_record

**Label:** Year of last record

**Description:** The most recent year for which a record of the organization’s activity or online presence could be identified, as determined during the RA verification phase.

**Important interpretive note:** This variable does **not** necessarily indicate organizational closure or dissolution. Because data collection occurred between 2023 and 2026, many research assistants recorded the year of data collection as the last record for organizations that were still active at the time. The variable therefore has two distinct meanings: for organizations with last records in 2023–2026, it likely reflects continued activity; for organizations with last records in earlier years (e.g. 2015, 2018), it may indicate that the organization ceased operations or lost its online presence at that point.

**Type:** Numeric (integer year)

---

Statistic	Value	%
Non-missing	6814	74.5
Missing	2333	25.5
Minimum	1978	NA
Median	2024	NA
Mean	2023	NA
Maximum	2026	NA

---

### 3.3.5 source\_last\_record

**Label:** Source for the year of last record

**Description:** The type of source from which `year_last_record` was obtained. Coded using the same six-category scheme as `source_first_record`.

**Note on the “Government/legal registry” category:** For this variable, this category also includes deregistration, dissolution, closure, and liquidation records — formal evidence that an organization ceased to exist as a legal entity.

**Type:** String / categorical (coded)

Value	N	%
Official organizational source	5750	62.9
(Missing / NA)	2118	23.2
Third-party documentation	799	8.7
Other	330	3.6
Government/legal registry	88	1.0
Network/coalition listing	37	0.4
Archives/academic	25	0.3

### 3.3.6 link\_source\_last\_record

**Label:** Link for the source for the year of last record

**Description:** A URL linking to the source used to establish `year_last_record`. Cleaned using the same rules as `link_source_first_record`.

**Type:** String / URL

	N	%
Non-missing	6183	67.6
Missing	2964	32.4
Total	9147	100.0

## 3.4 Group 4: Online Presence & Contact

### 3.4.1 online\_presence

**Label:** Does the organization have any online presence (website, social media, etc)

**Description:** Whether the organization has any verifiable, direct online presence, including a website or an active social media account. Assessed by research assistants during the verification phase and coded as “yes”, “no”, or “unclear”. The “unclear” category covers cases where active online presence could not be confirmed or denied.

**Type:** String / categorical (coded)

Value	N	%
yes	6542	71.5
(Missing / NA)	1781	19.5
no	727	7.9

---

Value	N	%
unclear	97	1.1

---

### 3.4.2 website\_link

**Label:** Organization’s website link

**Description:** The URL of the organization’s primary website, recorded during the RA verification phase. During cleaning, any cell not containing a recognisable URL pattern was set to missing. Missing values indicate either that the organization has no website or that one could not be located.

**Type:** String / URL

---

	N	%
Non-missing	5547	60.6
Missing	3600	39.4
Total	9147	100.0

---

### 3.4.3 social\_media\_link

**Label:** Organization’s social media link

**Description:** A URL for the organization’s primary social media presence (Facebook, Instagram, Twitter/X, LinkedIn, YouTube, TikTok, etc.), recorded during the RA verification phase. Where multiple platforms were found, research assistants typically recorded one, though some recorded more than one, typically separated by semi-colons.

**Type:** String / URL

---

	N	%
Non-missing	6105	66.7
Missing	3042	33.3
Total	9147	100.0

---

### 3.4.4 org\_email

**Label:** Organization’s email ID

**Description:** A publicly listed contact email address for the organization, as found on its website or social media during the RA verification phase. Not verified for deliverability. Missing values indicate that no public email was found.

**Type:** String / email address

---

	N	%
Non-missing	6456	70.6
Missing	2691	29.4
Total	9147	100.0

---

## 3.5 Group 5: Leadership

---

### 3.5.1 leader\_name

**Label:** Organization’s leader name

**Description:** The name of a publicly identified leader of the organization, as found on its website, social media, or third-party sources during the RA verification phase. Where multiple leaders were identified, research assistants recorded the most senior or most visible person. High missingness (58%) reflects organizations for which no named leader could be identified from public sources.

**Type:** String / free text

---

	N	%
Non-missing	3895	42.6
Missing	5252	57.4
Total	9147	100.0

---

### 3.5.2 leader\_role

**Label:** Leader’s role in the organization

**Description:** The organizational role of the leader named in `leader_name`. Original entries were free text; these have been coded into nine standardised categories during cleaning (see coding notes below).

**Type:** String / categorical (coded)

**Coding notes:** Free-text role descriptions were assigned to one of nine categories. The coding applies a strict priority rule: founder status takes precedence over all other roles, followed by Executive Director/CEO, and so on. For example, for compound roles such as “Executive Director and Founder”, the person is coded as “Founder/Co-Founder” because founding represents a unique historical relationship to the organization that is analytically distinct from an operational title.

---

Code	Description
Founder/Co-Founder	Anyone involved in founding the organization, regardless of current operational role. Includes co-founders.
Executive Director/CEO	Top operational executive; includes Managing Director and General Director
President/Chair	Board-level president or chair roles
Board Member/Trustee	Other board governance positions

---

Code	Description
Director	Director roles not in executive or board chair positions; includes programme, artistic, and festival directors
Coordinator	Coordination roles at any level
Manager/Administrator	Management and administrative positions; includes “Head of” roles
Specialized Role	Creative, artistic, religious, or professional roles (e.g. Musical Director, Pastor, Editor)
Other	Remaining leadership positions not fitting the above categories

Value	N	%
(Missing / NA)	5718	62.5
President/Chair	1043	11.4
Founder/Co-Founder	600	6.6
Other	516	5.6
Executive Director/CEO	504	5.5
Director	283	3.1
Specialized Role	169	1.8
Coordinator	162	1.8
Manager/Administrator	82	0.9
Board Member/Trustee	70	0.8

### 3.5.3 leader\_contact

**Label:** Leader’s contact information

**Description:** Contact information for the named leader, as found in public sources during the RA verification phase. This may be an email address, phone number, or social media handle. High missingness (81%) reflects the limited public availability of direct leader contact information.

**Type:** String / free text

	N	%
Non-missing	1839	20.1
Missing	7308	79.9
Total	9147	100.0

## 3.6 Group 6: Organizational Status

### 3.6.1 independent\_org

**Label:** Is the organization an independent non-profit

**Description:** Whether the organization operates as an independent non-profit, as opposed to being a department, chapter, student group, or program of a larger institution (e.g. a university, government body, church, or political party). Coded as “yes”, “no”, or “unclear”. The intended scope of this dataset is

independent civil society organizations; the `independent_org` variable can be used to restrict the sample accordingly.

**Type:** String / categorical (coded)

**Coding notes:** Research assistants entered free-text assessments during the verification phase. “Yes” was assigned to clearly independent organizations. “No” was assigned to cases where the organization was explicitly described as part of, affiliated with, or operated by a larger institution. “Unclear” was assigned to ambiguous cases or where information was insufficient to make a determination. NA indicates the field was not completed. Missing/NA fields here can be generally interpreted as “yes”, since this was in large part, a result of coding inconsistency across RAs.

**Analyst guidance:** The `drop_flag_raw` column contains research assistants’ original rationale for uncertain cases and can be consulted alongside this variable when making case-by-case exclusion decisions.

---

Value	N	%
yes	5301	58.0
(Missing / NA)	3548	38.8
no	199	2.2
unclear	99	1.1

---

### 3.6.2 `umbrella_org`

**Label:** Is the organization an umbrella group of some other larger organization

**Description:** Whether the organization is itself an umbrella group, coalition, or network that encompasses other member organizations, as opposed to being a standalone service or advocacy organization. Coded as “yes”, “no”, or “unclear”. Missing/NA fields here can be generally interpreted as “no”, since this was in large part, a result of coding inconsistency across RAs.

**Type:** String / categorical (coded)

---

Value	N	%
no	4429	48.4
(Missing / NA)	3982	43.5
yes	657	7.2
unclear	79	0.9

---

### 3.6.3 `issue_focus`

**Label:** Issue focus of the organization

**Description:** The primary issue area(s) the organization focuses on, entered by research assistants during the verification phase. Entries are free text and highly inconsistent in format — some are single terms (e.g. “LGBT+”), some are comma-separated lists (e.g. “LGBT+Q+; HIV/AIDS; human rights”), and some are descriptive phrases. This field was not standardised in cleaning and should be treated as qualitative annotation rather than a coded categorical variable.

**Coverage note:** This variable was only present in a subset of researcher sheets and is missing for over 90% of observations. It should not be used as a basis for sampling or filtering without awareness of this severe coverage limitation.

**Type:** String / free text (open)

---

	N	%
Non-missing	840	9.2
Missing	8307	90.8
Total	9147	100.0

---

---

## 3.7 Group 7: Sources & Verification

---

### 3.7.1 source\_1

**Label:** First source verifying this organization's existence/information

**Description:** The primary source used to verify the organization's existence and basic information, recorded by research assistants during the verification phase. During cleaning, non-substantive entries (explicit unknowns such as "DK", email addresses, very short strings that are clearly not citations, and single-word entries without a URL) were set to missing. Substantive entries are typically URLs, publication titles, or descriptive source references.

**Type:** String / URL or citation

---

	N	%
Non-missing	6195	67.7
Missing	2952	32.3
Total	9147	100.0

---

### 3.7.2 source\_2

**Label:** Second source verifying this organization's existence/information

**Description:** A second source used to verify the organization's existence and information, recorded during the RA verification phase and cleaned using the same rules as `source_1`. Where a second source was not found, this field is missing.

**Type:** String / URL or citation

---

	N	%
Non-missing	4134	45.2
Missing	5013	54.8
Total	9147	100.0

---

---

	N	%
--	---	---

---

### 3.7.3 both\_sources

**Label:** Are both source columns filled in the dataset

**Description:** A derived logical variable indicating whether both `source_1` and `source_2` are non-missing. A value of `TRUE` indicates that at least two independent sources verifying the organization’s existence were identified during the RA verification phase. This variable can be used as a proxy for the overall reliability of the organizational record.

**Type:** Logical (TRUE / FALSE)

---

	N	%
Both sources		
FALSE (one or no sources)	5038	55.1
TRUE (both sources present)	4109	44.9

---

## 3.8 Group 8: Research Process Metadata

---

### 3.8.1 drop\_flag\_raw

**Label:** Research assistants’ notes about this organization generated during data collection

**Description:** Free-text notes entered by research assistants in the “Drop?” column of the raw data collection template. These notes typically explain why an organization might be excluded from analysis — most commonly because it is not an independent organization (e.g. affiliated with a university or government body), because it could not be verified, or because it is international rather than domestic in scope.

**Important:** Research assistants were not consistent in updating this column. This variable should be read as qualitative annotation to support analyst judgment. Its presence does not mandate exclusion. Combinations of `independent_org`, `umbrella_org`, and the two `source` variables provide a more reliable basis for programmatic filtering decisions.

**Type:** String / free text (qualitative annotation)

---

	N	%
Non-missing	2798	30.6
Missing	6349	69.4
Total	9147	100.0

---

### 3.8.2 notes

**Label:** Research assistants' notes about this organization generated during data collection

**Description:** General free-text notes entered by research assistants during the verification phase. Content varies and may include contextual information about the organization, caveats about data quality, explanations of ambiguous cases, or flags for follow-up. These notes are not standardised and should be treated as qualitative metadata rather than structured data.

**Type:** String / free text (qualitative annotation)

	N	%
Non-missing	1927	21.1
Missing	7220	78.9
Total	9147	100.0

## 4 Missingness Across All Variables

Table 38: Missingness summary (N = 9147 organizations)

Variable	N missing	% missing
organization_alt_name	8335	91.1
issue_focus	8307	90.8
mission_native	7519	82.2
leader_contact	7308	79.9
notes	7220	78.9
drop_flag_raw	6349	69.4
about_native	6054	66.2
leader_role	5718	62.5
leader_name	5252	57.4
source_2	5013	54.8
language	4869	53.2
mission_en	4812	52.6
umbrella_org	3982	43.5
website_link	3600	39.4
independent_org	3548	38.8
social_media_link	3042	33.3
link_source_last_record	2964	32.4
source_1	2952	32.3
about_en	2710	29.6
org_email	2691	29.4
link_source_first_record	2674	29.2
year_last_record	2333	25.5
source_last_record	2118	23.2
year_first_record	1924	21.0
source_first_record	1852	20.2
online_presence	1781	19.5
region	838	9.2
country	35	0.4

---

Variable	N missing	% missing
organization	0	0.0
both_sources	0	0.0

---