Sunstone Institute

Counting Schools Around the World

Data Methodology

Executive Summary

This project consolidates the data from various data sources such as government data portals, government released statistics, government educational reportings, and for some cases reputed news outlets across 193 United Nations member states. The database provides a national count of schools for each member state.

Coverage and Results
As of December 6, 2024, school data was collected for 190 countries, and a total count of 5,919,463 schools.

Definition and Scope
For this study, a "school" is defined as an educational establishment that is either a preschool, primary school, secondary school, or a combination of these (ISCED levels 0–3) or for certain countries which do not have a formal school classification under ISCED, were classed according to the country's definition of preschool, primary, and secondary schools, further referred as sunstone map. School counts are aggregated at the national level, including both subcategories of public and private schools where available.

Data Sources
Data was collected from official government sources, including National Education Ministries (42%), statistical reports and yearbooks (31%), the Eurydice Network for European countries (9%), National Statistical Offices (8%), and Open Data portals (6%). Each record includes source URLs for verification and traceability.

Data Quality
An internal standardized quality framework categorizes data into four ratings: Good (57%), Excel (22%), Okay (18%), and Re-Check (3%).

Geographic Scope
The scope of selected countries was limited to current, active UN member states. The final list was created by extracting and filtering the available UN membership data, a process that involved parsing membership dates and verifying active status. Additionally, ISO 3166-1 alpha-2 codes were added for each country.

Temporal Scope
Data spans multiple years from 2001 to 2024, with the majority of records (62%) collected from 2020-2024. The "Time-frame" field in the dataset indicates the specific reference year for each country's data.

Limitations

The study encountered challenges including inconsistent reporting across official portals, data loss due to poor website maintenance, missing data for recent years, and inconsistent ISCED-aligned classifications. For three countries (Cuba, Turkmenistan, Comoros) we have not found data.

Data Schema

Data Field	Type	Description	Example
Country	string	Country name	Afghanistan
Alpha-2 Code	string(2)	ISO 3166-1 alpha-2	AF
Alpha-3 Code	string(3)	ISO 3166-1 alpha-3	AFG
Date of Admission	date	UN membership start date	1946-09-11
Amount	float	School count	2088.0
Source	string	Data source URL	https://...
Data Quality	enum	Quality assessment	Good/Excel/Okay/Re-Check
Time-frame	string	Data collection period	2020, 2021-22

Collection Methodology

Data Sources

Data was collected from multiple authoritative sources, prioritized in the following order:

1. **National Education Ministries**

   Official government portals providing national education statistics (e.g., moe.gov.af, education.go.ke, deped.gov.ph)

2. **National Statistical Offices**

   Official statistical agencies publishing education indicators (e.g., stat.gov.kz, bps.go.id)

3. **Eurydice Network**

   European Commission's network providing standardized education system information for European countries

4. **Open Data Portals**

   Government open data initiatives (e.g., Open Data for Africa)

5. **Secondary Sources**

   World Bank reports and national news outlets (used only when official sources were unavailable)

Data Collection Process

For each UN member state, we followed a systematic collection process:

1. **Source Identification**
   We searched for the country's official Ministry of Education website. If unavailable, we searched for the National Statistical Office, followed by UNESCO data and regional databases.

2. **Data Location**
   We navigated to the education statistics section and located the total school count, noting whether data was disaggregated by education level or school type.

3. **Data Extraction**
   We recorded the total school count, copied the full source URL for verification, and noted the reference year or academic year.

4. **Quality Assessment**
   We assigned a data quality rating based on source reliability, data recency, and completeness.

Source Validation

Each source underwent validation to ensure authenticity:

- **Domain Verification**

  We verified official government domains (.gov, .gob, .gouv patterns)
- **Authority Confirmation**

  We confirmed the source was the official national ministry or statistical authority
- **Provenance Check**

  We identified the specific publication, noted publication dates, and verified authorship
- **Plausibility Check**

  We reviewed outliers and validated school counts against population ratios

Multi-Source Countries

Some countries require aggregating data from multiple sources. For example: Burkina Faso — Four sources combined (pre–school, primary, and secondary reported separately)

Classification Framework

ISCED Alignment

The dataset categorizes school counts according to the International Standard Classification of Education (ISCED) 2011 framework, covering levels 0 through 3:

National Classifications (Sunstone map)

In cases where countries lacked formal alignment with ISCED classifications, schools were categorized based on national definitions of preschool, primary, and secondary levels. The list of school level categories considered for each country is available in the "School classification" sheet of the dataset.

In the final dataset, the column in the "ISCED Rating" field indicates the classification method used:
- **ISCED** — Direct ISCED classification from source
- **Sunstone Map** — Mapped based on the national classification
- **No data** — Education level breakdown not available

Data Quality Framework

Quality Ratings

Each country's data was assigned one of four quality ratings:

- **Good** (57% of records)
  Data from official government ministry or statistical office sources, collected from 2020 or later, with direct URL access and clear methodology documentation.

- **Excel** (22% of records)
  Data extracted from downloadable spreadsheets or PDF documents, typically from official statistical yearbooks or can be extracted in .csv format.

- **Okay** (18% of records)
  Data from secondary sources or older publications (pre-2020), where the classification of "school" may be less precise.

- **Re-Check** (3% of records)
  Data requiring verification due to inconsistencies, changed source URLs, or discrepancies with other sources.

Coverage Statistics

- Countries with school data: 190 of 193 (98.4%)
- School count range: 10 (Nauru) to 1,489,115 (India)
- Median school count: 6,438

Data Processing

Cleaning Procedures

1. **Numeric Standardization** – Removed formatting (commas, spaces) and converted values to integers
2. **Country Name Alignment** – Standardized country names to official UN nomenclature
3. **ISO Code Enrichment** – Added ISO 3166-1 Alpha-2 codes
4. **Quality Rating Normalization** – Standardized quality ratings

Aggregation Rules

1. School counts are aggregated at the national level
2. Public and private schools are combined where both are reported
3. Education levels are summed for the total count (Pre-primary, primary and secondary)
4. For countries with multiple sources, data was aggregated by education level to avoid double-counting.

Data Limitations

The study encountered certain challenges during data collection:

1. **Inconsistent reporting across official portals**

   Countries use different reporting formats and definitions

2. **Data loss due to poor government website maintenance**

   Some historical data is no longer accessible

3. **Missing data for recent years**

   Not all countries publish annual statistics

4. **Limited data accessibility to specific geographic regions**

   Some regions have restricted data access

5. **Data velocity issues**

   Infrequent updates mean some data is several years old

6. **Inconsistent ISCED alignment**

   Not all countries report using ISCED classifications

7. **Lack of clarity on school types**

   Public versus private breakdown unavailable for many countries

8. **Lack of specification on education levels**

   Pre-primary, primary, and secondary breakdown unavailable for many countries

Temporal Limitations

Data years range from 2001 to 2024. Direct cross-country comparisons should account for different reference years. The oldest data in the dataset is from the Democratic Republic of Congo (2001) and Iraq (2001).

Counting Methodology Variations

Countries may count schools differently, (for e.g., based on buildings, administrative units etc..) and must be referred for the counting methodology used by the country itself.

Data Dictionary

Field	Description
Country	Official UN member state name
Country Code	ISO 3166-1 alpha-2 code (2 letters)
Date of Admission	Date of UN membership
Amount	Total count of schools
Public Schools	Government-funded schools (or "-" if unavailable)
Private Schools	Privately-funded schools (or "-" if unavailable)
Pre-school	ISCED Level 0 institutions
Primary	ISCED Level 1 institutions
Secondary	ISCED Levels 2-3 institutions
Source	URL(s) to original data source
Data Quality	Assessment rating (Good / Excellent / Okay / Re-Check)
Time-frame	Reference year for the data
ISCED Rating	Classification method used

Updates and Maintenance

This dataset is maintained and updated as new official statistics become available. Update priorities are:

1. Countries with "Re-Check" quality rating
2. Data older than 3 years
3. Countries with incomplete education level breakdowns
4. Newly released national education statistics

For questions about the methodology or to report data corrections, please contact the Sunstone Institute data team.