

# UNEP Live Global database aggregation methods

---

## Introduction

Where sufficient data is available, aggregations are performed for all indicators which share a common unit and are believed to be internationally comparable. Indicators which are expressed in national currency or another national unit are not aggregated.

Aggregations are calculated for sub-regional, regional, global and special groupings. The same aggregations are used for the UNEP regions available on UNEP Live and the aggregations of UNEP data in the UNSD SDG database. All aggregates on UNEP Live have been calculated by UNEP. This is to ensure consistency across the membership of the regional groupings and across the methods.

UNEP employs the following strategy for calculating aggregate values: (1) determining if the aggregate grouping meets the minimum data criteria for aggregation; (2) imputing country data for missing years when data for one or more other years is available based on the criteria described later in this document; and (3) calculating an aggregate sum, average or weighted average for indicators that meet the minimum data criteria.

## Minimum data criteria

In order to ensure that regional aggregates are only computed when sufficient data is available the following criteria are used.

1) For each indicator a basis for the determination of data availability is determined. For most indicators, the basis for the determination of this criteria is either (a) population; (b) land area;

(c) GDP; or (d) the number of countries. The variable for this determination is denoted as: **X**

where the aggregate of X for each aggregation region is denoted as ***X<sub>agg\_region</sub>***. The criterion for number of countries is based on the number of UN Member States with available data.

2) Each data point is categorized as either (1) having available data for a specific year, (2) not having data for a specific year, but having data available that can be used to interpolate or extrapolate the data for the specific year, or (3) having no data and no nearby data that can be used for interpolation or extrapolation. Note that the interpolation and extrapolation procedures are described in the next section and the basis for extrapolation is that the extrapolation should not be more than three years forward or backward and interpolation should not occur when the data gap is more than ten years.

3) The volume of **X** that is represented with available data or data which can be imputed for each aggregation region is denoted as ***D<sub>agg\_region</sub>*** for data available and ***I<sub>agg\_region</sub>*** for data which can be imputed.

4) These calculations are used to ensure that more 30% of a regional, sub-regional or economic aggregation grouping has data available for the year which is being aggregated and more than 55% of the aggregation grouping has either data which is available or can be imputed. Thus an aggregation is only calculated when:

$$D_{agg\_region} / X_{agg\_region} > 0.3 \text{ and } (D_{agg\_region} + I_{agg\_region}) / X_{agg\_region} > 0.55.$$

These criteria for meeting the minimum data availability requirements for aggregation are less stringent than the criteria that are used by some other international organizations. The reason for considering the representation of countries with data that can be interpolated or extrapolated is due to the fact that many environment related indicators are not calculated annually but are available every 2-5 years and this method provides a means for creating aggregations in these circumstances.

## Imputing missing years using interpolation and extrapolation

Missing values are imputed for missing years based on interpolation or extrapolation. Missing data are only used in the calculation of aggregates and are not incorporated into national level data.

Interpolation is used to fill in a data gap of ten or less years. Extrapolation is used to forecast or backcast for a maximum of 3 years forward or backward, respectively.

Interpolation is based on the following method:

- 1) If both the data point preceding and following a particular data gap are positive, non-zero values then the data gap is interpolated using an exponential growth rate.
- 2) If a data point preceding or following a particular data gap is negative or zero then the data gap is interpolated using a linear growth rate.
- 3) If the data gap is more than ten years, then extrapolation will be used for three years after the first data point and three years before the last data point. However, the remainder of the gap is not imputed.

Extrapolation is based on the following method:

- 1) If there are two data points that within 5 years preceding or following a particular year then the data is extrapolated using either the exponential or linear growth rate. The exponential growth rate is used if the values are positive, non-zero numbers. The growth rate is used to extrapolate for up to three years before or after the last available data point.
- 2) If there is only one data point that can be used to extrapolate, then the figure is carried forward or backward for up to three years (no growth rate is applied).
- 3) Data are not extrapolated for more than 3 years.

## Aggregates

Sums, averages and weighted averages are calculated using simple arithmetic formulas. For weighted averages, missing weights are imputed using the same process described above. If the weight is missing after imputation, then that country is omitted from the calculation of the weighted average.

## Calculating “totals” based on other aggregates

For some indicators, the total for a subregional or regional aggregate is based on a weighted average of another variable. For example, the total population living in poverty is calculated as the proportion living in poverty for an aggregate grouping multiplied by the total population. This method provides a mechanism for ensuring that the “totals” and the “average” are comparable. It also provides a way of ensuring that the “total” is not grossly underestimated.

## Metadata

The specific details on the basis for meeting the minimum data criteria and the weights used in computing weighted averages are included in the metadata for each indicator.

For more information contact: [unenvironment-science-sdgs@un.org](mailto:unenvironment-science-sdgs@un.org)