EM-DAT Documentation

Documentation   EM-DAT Project   Public Data   HDX   GitHub   About

Protocols

Data Collection Processes

- 1: Entry Criteria
- 2: Encoding, Quality Control, and Validation Procedure
- 3: Economic Adjustment

1 - Entry Criteria

Minimal Requirements for a Disaster To Be Entered in EM-DAT

The EM-DAT definition of a disaster considers unintended hazards with a substantial impact unforeseen by a community (see General Definitions and Concepts). For management and operational purposes, EM-DAT has a set of entry criteria that specify what substantial impact means. EM-DAT disaster records related to natural and technological hazards meet at least one of the following inclusion criteria:

EM-DAT Inclusion Criteria

- At least ten deaths (including dead and missing).
- At least 100 affected (people affected, injured, or homeless).
- A call for international assistance or an emergency declaration.

There are, however, secondary criteria, especially for past events where quantitative data were not available (e.g., "the worst disaster in a country or region" or "an event that resulted in considerable damage").

2 - Encoding, Quality Control, and Validation Procedure

How EM-DAT Data Is Encoded and Controlled?

Source identification and collection can be facilitated and partially automated thanks to online services offered by specific sources (e.g., email alert systems, news feeds, or APIs). However, data collection and encoding are always supervised manually by the database manager. The database manager controls the sources that are selected, the classification of the event, its spatiotemporal delimitation, and the identification of the impact figures. Data encoding and validation in EM-DAT is a three-step process:

> ### EM-DAT Updating Routines
>
> 1. Daily Encoding.
> 2. Quality Control and Annual Validation.
> 3. Periodic Thematic Reviews.

In addition, Automated Procedures and Constraints prevent or check for abnormal values or data formats in the database.

Daily Encoding

The database manager checks daily information using the preferred source list (see EM-DAT sources). Whenever a new disaster is identified based on a source, it is added to the database. In the first stage, the event is not made public. It becomes so when an entry criterion is met and confirmed by at least two sources. The figures remain subject to changes. Any publication or modification made on a public disaster entry will be visible to the user after the weekly update routine. This routine is usually executed on mondays but may be triggered by the manager if deemed necessary, e.g. in the case of faulty figures or typos.

The published impact variables could be selected from one or more sources. An event can therefore be validated from several sources of information. For example, the human impact can come from an OCHA report and the economic data from a reinsurance report, depending on specific expertise. If the figures differ between the sources, the database manager decides which ones to attribute to the disaster. The choice depends on several elements: the figure itself and the area and period to which it refers, the sources' chronology, and its degree of reliability. Because this task is complex and case-dependent, there is no pre-determined rule for selecting figures, and

the database manager makes the final choice. Some examples of general, however, not systematic, decision rules are illustrated in the box below.

> ## Informal Rules used to Prioritize Sources and Select Loss Statistics
>
> - Sources offering loss statistics on a broader spatial and temporal scale are preferred to sources providing figures for a restricted locality and time frame. For example, a national report is prioritized over a news brief reporting the losses for a specific day in a local district.
>
> - The latest and more mature sources are preferred to earlier ones as they are likely to better reflect the aftermath of the disaster.
>
> - Sources from research institutes, governmental organizations, or humanitarian NGOs are preferred to those from the press and early emergency alerting systems as their figures are assumed to be more accurate.

The rules here are only informal and the database manager may disregard them. For example, suppose a report mentions 200 deaths and a news article says, "regional authorities estimate the death toll now stands at 243 dead and missing". Although this is a newspaper article, the precision of the statement suggests that this figure is more reliable than the one in the official report.

Quality Control and Annual Validation

Quality control and annual validation are systematic checks of all the entries starting in a specific year. It typically takes place at the beginning of the following year. During this validation, all the disasters that took place in the previous year are reviewed to consolidate the data, identify possible additional sources, and modify the published figures accordingly. In addition, the georeferencing, i.e., the more precise attribution of the disaster to GAUL level 1 or 2 zones, is also finalized during the annual validation period (see GAUL Index and Admin Levels).

Thematic Reviews

The CRED periodically conducts thematic reviews of disasters to mitigate the database's weaknesses (see Known Issues). This task involves systematically checking entries for a type of disaster over a given period or region. The revision can be a data analysis for further quality control, a systematic review of the scientific literature, or a comparison with other existing databases.

From September 2023 onward, these substantial database content updates are planned to be notified on the EM-DAT website and in the documentation release notes for tracking purposes (see Introduction). The `Entry Date` and `Last Update` column have also been introduced in the EM-DAT Public Table.

Automated Procedures and Constraints

In 2023, the constraints of the EM-DAT database have been strengthened. These constraints define the domain of values that can be encoded and, if correctly set, can prevent encoding errors in the value or its format. In addition, there are automated procedure checks for anomalies that are likely to be an error, i.e., those without enough certainty to be a constraint but with sufficient likelihood to be notified to the database manager for verification.

Currently, constraints and automated routines check for the consistency of date and time fields, latitude and longitude values, and hazard magnitude values. These procedures are implemented incrementally each time an error that could be prevented is discovered. Hence, by detecting and reporting issues, users may contribute to developing these routines and improving EM-DAT data quality.

Issue Reporting

The CRED encourages any external initiative that can help us improve our data quality. Any problem with the data content (such as errors, or missing data) can be reported to CRED by email using the contact address (see Send Us an Email). For missing entries, you should be aware that the CRED only publishes disasters for which reliable figures are available and when these are corroborated by at least two sources (see Daily Encoding). The CRED reserves the right to modify the data according to subsequent notifications. For issues concerning specific disaster events, the problem should be explained and notified with a list including the related `Dis No.` values (see Column Description).

For less specific issues, users who have analyzed the quality of EM-DAT are encouraged to share their reports or scientific studies with the EM-DAT team to improve the database (see Contributing).

3 - Economic Adjustment

How are Economic Impact Variables Converted and Adjusted for Inflation?

All economic damage in EM-DAT is expressed in thousands of US$ (see Column Description and Economic Impact Variables). For each disaster, the registered figure corresponds to value of the damage at the moment of the event. If a source reports damage in another currency, it is first converted to US$ before it is entered in EM-DAT using a converter (see this example), at the exchange rate when the damage occurred.

The adjusted economic losses in EM-DAT provide a monetary value in US$ that has been adjusted for inflation (see Column Description). The adjustment is linearly proportional to the OECD Consumer Price Index (CPI) provided by the Organization for Economic Cooperation and Development (OECD). The CPI reflects the change in prices of a basket of goods and services that are typically purchased by specific groups of households. EM-DAT relies on the total CPI, i.e., including both food and energy products in the basket, defined for the USA.

In practice, OECD provides the US CPI with 2015 as the base year ($CPI_{2015} = 100$, see https://data.oecd.org/chart/6RMa). In EM-DAT, the CPI is rescaled to use the last year as a reference, e.g., $CPI_{2021} = 100$. The rescaling is performed when the OECD CPI value becomes available. The economic losses given for the current year are therefore not adjusted and not reported in the column `Total Damages, Adjusted ('000 US$)`.

How to Compute an Adjusted Economic Impact

For a given year Y and its rescaled CPI value CPI_y, a monetary loss X is adjusted into a year-specific X_{adj} based on the formula:

$$Xadj, 2021 = CPI_{2021} \times \frac{X}{CPI_y}$$
$$\text{where}^* : CPI_{2021} = 100$$

*2021 is here given as an example. In the EM-DAT Public Table, the annual `CPI` values are adapted and rescaled from OECD values to ensure that we have the latest year available at the 100 value. This does not affect the end result as the CPI ratio between years is conserved.

Examples

For instance, on December 25, 2022, a disaster entry that occurred in 2021 would have a `CPI` value of 100. A disaster that has occurred in 2022 will not have a `CPI` value; hence, no economic losses will be mentioned in the `Total Damages, Adjusted ('000 US$)` column.

If a disaster has an economic loss of 1 million US$ in 2019 with a rescaled `CPI` value of 93, the adjusted damage would be 100×1,000,000/93=1,075,000 US$.