# Codebook for IOEval Data set

*The data set in the repository is called "IOEval.csv"*

**When using the data, please cite**: "Eckhard, Steffen; Jankauskas, Vytautas; Leuschner, Elena; Burton, Ian; Kerl, Tilman; Sevastjanova, Rita (2023). The Performance of International Organizations: A New Measure and Dataset Based on Computational Text Analysis of Evaluation Reports. DOI: 10.1007/s11558-023-09489-1.

IOEval is a data set at the sentence-level containing evaluation reports from international organizations (IOs).

This version of the dataset comprises of reports from nine major UN system IOs: ILO, UNDP, UNICEF, FAO, UNESCO, WHO, IOM, UNHCR, and UN WOMEN. We chose these IOs to gain a diverse set of organizations while ensuring that each published a large number of reports: First, these IOs vary in their policy fields, staff size and constellations, as well as budgetary scale and scope. Second, each IO published between 43 and 244 reports between 2012 and 2021.

As members of the UN Evaluation Group (UNEG), these IOs are subject to the same system-wide evaluation norms and guidelines (UNEG, 2016), making their evaluation reports comparable. To verify that the definition of evaluation and the evaluation criteria indeed matched the UNEG standards, we restricted the data collection for each organization to the period for which we could access their evaluation policies.

The raw text read in from the PDFs was cleaned by applying standard procedures of natural language processing (e.g., removal of special characters and numbers) and split into sentences.

The final *IOEval* dataset includes a total of 1,082 evaluation reports published from 2012 to 2021 and 923,694 distinct sentences, indicating their order in the original report. In addition, the *IOEval* dataset also includes metadata variables at the level of reports: report *title*, *publication date*, *evaluation type* (project, program, institutional or thematic), *evaluation level* (country (specifying its name), regional, global), and *commissioning unit* (centralized or decentralized). At a sentence level, we specify to which *text section* a sentence belongs (executive summary, main text, appendix).

Below we describe every variable included in the IOEval data set.

| Variable name | Description |
|---|---|
| *id* | String containing a unique ID for every report |
| *title* | String identifying the title of the evaluation report |
| *eval_type* | String denoting what type of evaluation a report is classified as, includes evaluations that were labelled as "Impact", "Project/Programme", "Thematic, "Regional", "Report", "Strategic/Policy Synthesis of Evaluations", "Outcome/Country", and "Other" |
| *countries* | String containing the country in which the project of an evaluation was conducted in |
| *IO* | String containing the abbreviation of the international organization that commissioned the evaluation report |
| *agency* | String identifying the international organization that commissioned the evaluation report |
| *year* | Numeric value of the year in which an evaluation report was published |
| *sent_id* | Numeric value of the order in which a sentence appears in a report |
| *sentence* | String containing an identified sentence of the evaluation report |
| *completion_date* | String denoting in which year and month an evaluation report was published ("YYYY-m") |
| *decentral* | Numeric dummy variable turning 1 if an evaluation was |
| *control* | String containing under which control structure the evaluation was conducted, takes on the values "MS" for member states, "IPA" for international public administrations and "mixed" for mixed cases |
| *control_num* | Numeric value denoting the degree of control of the IPA ranging from 0 to 1 |
| *sentiment* | Numeric value denoting the predicted sentiment produced by the trained language model BERT, turns 0 if a sentence is predicted to be "neutral", -1 if a sentence is predicted to be "negative", and turns 1 if a sentence is predicted to be "positive" |
| *sentiment_chr* | String representation of sentiment |
| *accuracy_positive* | Predicted accuracy score that a sentence is "positive", ranges from 0 to 1 |
| *accuracy_negative* | Predicted accuracy score that a sentence is "negative", ranges from 0 to 1 |
| *accuracy_neutral* | Predicted accuracy score that a sentence is "neutral", ranges from 0 to 1 |
| *meas_uncert* | Numeric value of the accuracy score (obtained from the trained BERT model) of the sentiment prediction with the highest accuracy score of the sentence |
| *section* | String stating whether a sentence is part of the executive summary ("executive") or main text ("main") |
| *word_nr* | Sum of words contained in the sentence |

| | |
|---|---|
| *fao* | Numeric dummy variable that turns 1 if the report was commissioned by FAO and is 0 otherwise |
| *ilo* | Numeric dummy variable that turns 1 if the report was commissioned by ILO and is 0 otherwise |
| *iom* | Numeric dummy variable that turns 1 if the report was commissioned by IOM and is 0 otherwise |
| *undp* | Numeric dummy variable that turns 1 if the report was commissioned by UNDP and is 0 otherwise |
| *unesco* | Numeric dummy variable that turns 1 if the report was commissioned by UNESCO and is 0 otherwise |
| *unhcr* | Numeric dummy variable that turns 1 if the report was commissioned by UNHCR and is 0 otherwise |
| *unicef* | Numeric dummy variable that turns 1 if the report was commissioned by UNICEF and is 0 otherwise |
| *unwomen* | Numeric dummy variable that turns 1 if the report was commissioned by UN WOMEN and is 0 otherwise |
| *who* | Numeric dummy variable that turns 1 if the report was commissioned by WHO and is 0 otherwise |