Validation rules for air emissions accounts
2023 data collection

Version: April 2023

Contents

1. Introduction

Air emissions accounts (AEA) record flows of gaseous and particulate materials (six greenhouse gases including CO_2 and seven air pollutants) by emitting economic activities. Economic activities include production activities (hierarchically classified by NACE) and household consumption activities. AEA national totals deviate from the totals as defined in national emission inventories. AEA reconciles totals with national inventories through so-called 'bridging items'.

This document refers to the data transmissions for air emissions accounts (AEA) under Regulation (EU) 691/2011. Those transmissions are due to adopt SDMX standards, according to Implementing Regulation (EU) 2015/2174. This is the opportunity to make use of new validation services, which allow automatic data validation at a basic level. These checks will be integrated in the SDMX data transmission process without need for human intervention. The general set up of the SDMX data transmission schemes is described in document[1] discussed at the meeting of the working group environmental accounts on 15 May 2019.

This document presents the validation rules for air emissions accounts (AEA) which were discussed and approved by the working group environmental accounts already in 2018.

The document has been updated in September 2022 by adding rules for cross-domain plausibility of air emissions accounts (see chapter 3.9) and in March 2023 by updating chapters 3.3 and 3.5.

2. Types of validation results

A validation rule is a logical statement applied to data. Whenever a validation rule is applied on a dataset, it issues a validation result. The SDMX validation tools allow for four types of validation results:

OK   This means that the transmitted data passed the validation rule and no specific follow up is required. At this stage of validation, no further explanations are expected. However, it might be possible that questions will be asked during a later stage of the validation process. In general, validation rules that have 'OK-result' are not listed in the validation report.

ERROR   This means that a serious issue related to format, completeness, coherence, consistency and/or plausibility was detected in the transmitted data. In the fully automatized SDMX data transmission process an 'ERROR-result' implies the refusal of the data transmission and consequently the need of a

---

[1] See ENV/EA-MESA/WG/06(2019)

new transmission. In other words, the data are automatically sent back to the sender, without Eurostat having looked at them.

WARNING This means that some reported element is 'suspicious'. It might refer to an individual data point, combination of data points or calculations based on reported data points. It highlights an issue of attention and for which a valid explanation might exist. It does not mean that the data is not coherent or that a new data transmission is required. It merely indicates an element of special attention. A 'WARNING-result' will usually trigger questions to the sender seeking for an explanation and clarification the reported values. Countries can accelerate the validation procedures by providing footnotes explaining these warnings.

INFO This means that some reported element is simply an issue of attention to be highlighted. A further clarification or explanation is not required.

The application of validation rules to a dataset generates a validation report, which is usually a list of observations (records) presenting validation results.

This document only addresses ERROR and WARNING rules. Only ERROR results will trigger the refusal of data in the SDMX data transmission process.

3. Validation rules for AEA

This chapter presents a list of validation rules for AEA as discussed and agreed by the working group environmental accounts at their meeting in 2020 and updated in 2021. The list is based on the experiences gained in past AEA data collections.

The validation rules are presented by categories. For each validation rule the type of validation result is indicated in rectangular brackets [ERROR, WARNING].

This document expresses the rules in 'human language' close to the terms of the Excel questionnaire (such as questionnaire cells, symbols, footnotes, etc.). This is the terminology familiar to most of the AEA compilers and working group members. The SDMX terminology is different and the approach to define some elements is different (e.g. there are no questionnaire cells but records, values and footnotes are reported as combinations of SDMX attributes obs_value and obs_status, etc.

3.1. Completeness

This check gives an overview of completeness of the dataset for each pollutant.

Rule 1. The dataset must include all mandatory characteristics, for the five mandatory reporting years (5 years mentioned in Regulation (EU) 691/2011,

Annex I, section 4, paragraph 6). Missing data for all gases and pollutants except F-gases, bridging items and item NACE L68A trigger an [ERROR] result.

Rule 2. Missing data for F-gases, bridging items and item NACE L68A, for the five mandatory reporting years, trigger a [WARNING] result.

Eurostat seeks that this rule is upgraded to [ERROR] in a few years. This will be re-discussed with the working group environmental accounts.

3.2. Symbols

Symbols are the alphanumeric characters of a reported value in the cells of the questionnaire. This check identifies invalid symbols among the reported values that have passed the completeness check.

Rule 3. Valid symbols are:

- In the bridging items 'national residents abroad' and 'other adjustments and statistical discrepancy': positive value, zero value, negative value, colon (:) (i.e. 'not available')

- In any other cell: positive values, zero value, colon (:) (i.e. 'not available')

Invalid symbols trigger an [ERROR] result.

3.3. Internal consistency

This set of rules relates to the hierarchically classified characteristics, namely economic activities and bridging items. It checks that the sum of components equals the corresponding total or subtotal item.

Rule 4. The rule checks that each aggregate of activities is equal to the sum of the reported disaggregated activities:

4.1	A_U (Total industries) = A + B + C + D + E + F + G + H + I + J + K + L + M + N + O +P +Q +R + S + T + U
4.2	A = A01 + A02 + A03
4.3	C = C10_C12 + C13_C15 + C16 + C17 + C18 + C19 + C20 + C21 + C22 + C23 + C24 + C25 + C26 + C27 + C28 + C29 + C30 + C31_32 + C33
4.4	E = E36 + E37-E39
4.5	G = G45 + G46 + G47
4.6	H = H49 + H50 + H51 + H52 + H53

4.7	J = J58 + J59_J60 + J61 + J62_J63
4.8	K = K64 + K65 + K66
4.9	L68A < L
4.10	M = M69_M70 + M71 + M72 + M73 + M74_M75
4.11	N = N77 + N78 + N79 + N80_N82
4.12	Q = Q86 + Q87_Q88
4.13	R = R90_R92 + R93
4.14	S = S94 + S95 + S96
4.15	HH total = HH_TR + HH_HEAT + HH_OTH
4.16	AEA total = A_U + HH
4.17	Total AEA +/- Bridging Items = Total UNFCCC/CLRTAP
4.18	TER_NRES = TER_NRES_LTR + TER_NRES_WTR + TER_NRES_ATR
4.19	RES_ABR = RES_ABR_FWTR + RES_ABR_LTR + RES_ABR_WTR + RES_ABR_ATR
4.20	H50 > RES_ABR_WTR
4.21	H51 > RES_ABR_ATR
Rules for PM10, PM10_ROAD and PM2.5, PM2.5_ROAD	
4.22	For each cell, the value reported for PM10 must bigger or equal to the value reported for PM2_5
4.23	For each cell, the value reported for PM10_ROAD must bigger or equal to the value reported for PM2.5_ROAD
Rules for 'ROAD' sheets of the questionnaire	
4.24	For each cell, the value reported in 'ROAD' sheets must be smaller or equal to the value reported in the 'mother sheet' of the same pollutant.

This is an [ERROR] rule. Differences smaller than 0.00001 are accepted, i.e. they do not trigger an error.

3.4. Deletion of previously reported data

This check identifies the deletion of previously reported data outside the 5-year period established in Regulation (EU) 691/2011. Deletion means replacing previously transmitted data by missing values in the current data transmission.

Rule 5. Removing values previously transmitted (i.e. overwriting with missing value) and mandatorily requested in Regulation (EU) 691/2011 generates an [ERROR] result. This is for data for the period 2008 (=first mandatory year) or subsequent years according to the 5-year period established in Regulation (EU) 691/2011.

Rule 6. Removing values previously transmitted and not requested in Regulation (EU) 691/2011 generates a [WARNING] message. This is for data before 2008.

3.5. Footnotes

There are two types of footnote symbols: symbols for pre-defined footnotes (letter) and symbols for free text footnotes (numbers). This rule checks the correctness of symbols of pre-defined footnotes.

Rule 7. The valid pre-defined footnotes are the following:

b)	Break in series	Break occurring when there is a change in the standards for defining and observing a variable over time. The flag 'b' is to be attached to the first time period after the break.
c)	Confidential	Confidential data are data which are subject to confidentiality clauses. Where possible, further details should be provided in the national quality report via ESS-MH, under the concept 7. 'Confidentiality'.
d)	Secondary confidentiality	Secondary confidential data are data made confidential in order to prevent third parties to indirectly calculate the data points genuinely flagged as confidential. Please note that this footnote converts into confidential (c) when published in Eurostat's online database.
e)	Estimated data	The 'e' (estimate) flag shall be used only if one or several data points have been calculated using a significantly different methodology and/or sources than the rest of the data points in the questionnaire.
p)	Provisional	The 'p' (provisional) flag shall be used when a data point value is expected to be revised and submitted to Eurostat before the next data collection. In the case of early

		estimates, the flag 'e' is deemed sufficient and the 'p' flag can be omitted. Notice all 'p' (provisional) flags sent during a given data collection will be systematically removed during the subsequent data collection, unless the 'p' flags are again resubmitted.

The use of non-valid footnote symbols triggers an [ERROR] result.

Rule 8. The following combinations of pre-defined footnotes are allowed:

Combinations of pre-defined footnote symbols in Excel questionnaire	Combinations of pre-defined footnotes
))b)e)	Break in time series whilst estimated data
))b)p)	Break in time series whilst provisional
))b)e)p)	Break in time series whilst estimated data whilst provisional
)c)b)	Confidential whilst break in time series
)d)b)	Secondary confidentiality whilst break in time series
))e)p)	Estimated data whilst provisional

The use of different combinations of pre-defined footnotes triggers an [ERROR] result.

Rule 9. The following combinations of symbols of pre-defined footnotes and value will trigger a [WARNING]: pre-defined footnote symbols p) or e) in combination with the value (:) (i.e. 'not available').

See also rule 12 about confidentiality.

Rule 10. A free text footnote symbol with no text defined for it in the footnote area of the questionnaire triggers an [ERROR] result.

3.6. Confidentiality

This check verifies certain rules related to confidentiality.

Rule 11. Confidentiality at NACE section (i.e., 1 digit NACE level) is not allowed. This is an [ERROR] rule.

Rule 12. Cells with a value (:) (i.e. 'not available') cannot be confidential (pre-defined footnote c). This is an [ERROR] rule.

Rule 13. The figure flagged as confidential (pre-defined footnote c) could be derived through calculation. This is a [WARNING] rule.

3.7. Plausibility of reported time series

This check detects implausible changes between consecutive years in time series of the same reporting. Implausible changes are identified against a maximum range of change that is still considered plausible for purposes of the automatic validation.

Rule 14. The existence of implausible changes triggers a [WARNING] result. 'Implausible changes' are defined as follows:

- Totals of NACE sections (i.e., NACE A, B, C, etc.) which vary more than 30% between consecutive years if they are bigger than 10% of all industries (total A_U) in the second year.

- NACE industries A01-A03, E36-E39 and H49-H53 which vary more than 80% between consecutive years if they are bigger than 30% of total A, E and H respectively.

- NACE industries C10-C33 which vary more than 80% between consecutive years if they are bigger than 5% of total C.

- Any other two-digit NACE industry which varies more than 80% between consecutive years if it is bigger than 5% of all industries (total A_U) in the second year.

- Two-digit Households positions which vary more than 30% between consecutive years if they are bigger than 50% of Total Households in the second year.

- Total Households which vary more than 30% between consecutive years if they are bigger than 10% of Total Industries plus Households in the second year.

- Bridging items which vary more than 50% between consecutive years if they are bigger than 50% of the superior total, i.e. to National residents abroad or National residents in the territory.

Rule 15. Implausible changes must be explained with a free text footnote or, in the case of breaks in the series, with a pre-defined footnote b) and the corresponding free text footnote. The absence of a footnote triggers an [ERROR] result.

3.8. Plausibility of revisions

This check detects implausible revisions between the current and previous year's questionnaire. Implausible revisions are identified against a maximum range of revisions that is still considered plausible for purposes of the automatic validation.

Rule 16. Implausible revisions trigger a [WARNING]. 'Implausible revisions' are those bigger than 40% if the cell contributes at least to 10 % of Total Industries or Total Households.

3.9. External consistency (cross-domain plausibility)

This series of checks compare AEA to other reported and external data.

Please note, that the following cross-domain plausibility checks, except for rule 17, are not expected to be performed by countries before the data is transmitted.

Rule 17. The UNFCCC or CLRTAP inventory total reported in the AEA questionnaire (TOT_CONV) is compared to the UNFCCC or CLRTAP inventory total published by European Environment Agency (EEA). If the two totals do not equal, the country is asked for clarification on the data source used in the questionnaire. This is a [ERROR] rule.

Rule 18. There is a comparison between TOT_CONV and TOT_NACE_HH. For emissions of CO_2, SO_x, NO_x, CO, PM2.5 and PM10, if TOT_CONV (EEA total) equals to the reported TOT_NACE_HH (Total NACE plus households), it is very likely that the residence principle was not applied to AEA. In this case, the country must provide a footnote, clarifying why the residence principle was not applied. This is a [WARNING] rule.

Rule 19. Implausible differences between AEA data and PEFA (Physical energy flow accounts) data trigger a [WARNING] result

For checking the coherence between AEA and PEFA, one needs to relate tonnages of emissions to energy flows expressed in terajoules (TJ). For the resulting emission factors – i.e. tonnes of emissions per TJ of energy use – one may define plausible ranges around IPCC standard emission factors.

Air Emission Accounts (AEA) vs. PEFA Table C	
19.a	If CO2 (AEA) > 0 then P00 (PEFA Table C) > 0 If CO2 (AEA) = 0 then P00 (PEFA Table C) = 0 for 'A' to 'U', 'HH', 'HH_HEAT', 'HH_TRA', 'HH_OTH'
19.b	CO2:H51 (AEA) / P15:H51 (Table C) = 72 tCO2/TJ +/-40%
19.c	CO2:H50 (AEA) / (P13:H50 + P17:H50 + P18:H50 + P19:H50) (Table C) = 75 tCO2/TJ +/-40%
19.d	[CO2 + CO2_BIO (AEA)] / P00 (Table C) = annual change rate not beyond +/- 30% for 'A' to 'U', 'HH', 'HH_HEAT', 'HH_TRA', 'HH_OTH'

Rule 20. Implausible differences between AEA data and OECD's air transport CO_2 emissions trigger a [WARNING] result.

This check compares AEA data reported for air transport (NACE H51) and the related bridging items with the respective OECD data on CO_2 emissions from air transport[2].

AEA vs. OECD's CO_2 emissions from air transport data and UNFCCC data	
20.a	AEA H51 vs OECD H51 threshold +/- 10%;
20.b	AEMIS_RES_ABR_ATR vs OECD Bridging item - Residents abroad threshold +/- 10%;
20.c	AEMIS_TER_NRES_ATR vs OECD Bridging item - Non-residents on territory threshold +/- 10%;
20.d	AEA H51 vs [UNFCCC 1.A.3.a - AEMIS_TER_NRES_ATR + AEMIS_RES_ABR_ATR] threshold +/- 10%;

Rule 21. Implausible differences between AEA data and Energy statistics trigger a [WARNING] result

Detailed energy statistics on the final energy consumption of households (solid fuels + natural gas + oil and petroleum products + primary solid biofuels for space heating + space cooling + water heating + cooking) [nrg_d_hhq] are compared with Heating/cooling activities by households and Other activities by households of CO_2 emission item in AEA. As NRG comes in TJ and AEA is in tCO_2, an artificial emission factor is calculated that should fall in the range of +/-40% around an average CO_2 emission factor (65 tCO_2/TJ).

AEA vs. Energy statistics	
21.a	AEA_CO2_HH_HEAT vs NRG_HEAT = 65 tCO2/TJ +/-40%
21.b	HH_OTH vs NRG_HH = 65 tCO2/TJ +/-40%

Rule 22. Implausible differences between AEA data and Supply and use table trigger a [WARNING] result

The check verifies economic activity for each NACE division for which AEA reports emissions (value bigger than zero). In the past, it happened that AEA reported emissions for NACE divisions for which the SUT reported zero economic activity.

---

[2] https://stats.oecd.org/Index.aspx?DataSetCode=AIRTRANS_CO2