

# Reporting of establishment information under the Seveso Directive (eSPIRS)

**Quality controls documentation** 

\_\_\_\_\_



Cover design: EEA

Cover photo: Photo by Victor on **Unsplash** 

Layout: EEA

# **Acknowledgments**

This data model documentation was prepared by the <u>European Environment Agency</u> (EEA) in cooperation with the <u>European Topic Center for Health and Environment</u> (ETC/HE). The data model was produced using Unified Modelling Language (UML). Please contact <u>seveso.helpdesk@eea.europa.eu</u> for any questions, clarifications or comments.

## **Version control**

Version ID	Description	Date
0.1	Internal draft for discussion	May 2025
0.2	Internal draft for discussion	September 2025
0.3	Version for sharing with the Member State Reporters	October 2025

# **Contents**

Acknow	wledgments	2
Versio	n control	2
Conter	nts	3
1 Ir	ntroduction	4
2 Q	uality checks classification	5
3 Q	uality checks severity levels	6
4 V	alidation process	7
5 Q	uality checks description	9
5.1	Custom checks on text fields lengths	11
5.2	Custom checks on InspireId format	11
5.3	Custom checks on productionSite data consistency	12
5.4	Custom Spatial Checks	13
5.5	Custom checks on linked tables data consistency	13
5.6	Information Level checks	15
Annex	1 - Glossary of Terms	16

## 1 Introduction

Quality control (QC) is a key component of effective data management, ensuring that reported data meets fundamental quality criteria of accuracy, completeness, consistency, and uniqueness. Providing timely and comprehensive feedback on data quality issues helps minimize the number of corrections cycles and reduces the overall time required to achieve high-quality data submission.

The Reportnet 3 platform includes an integrated automatic quality control framework that delivers immediate feedback to users upon data submission. This document outlines the general principles of the quality control mechanisms implemented in Reportnet 3. It explains how reporters can use the system to verify their data and interpret QC feedback to improve the quality of their submissions.

In addition, this document provides a detailed description of the specific quality controls developed for the eSPIRS dataflow.

Please note that the current version of this document covers only QC checks assessing the content of a single data delivery. Future versions will include additional controls to compare data across reporting periods or to cross-reference other dataflows such as the EU Registry.

In addition to this document, two further related documents will be prepared which address other aspects of the data flow, namely the following:

- The Data Mode Documentation
- The Data Exchange Format Specification
- A Manual for Reporters, which will guide reporters on the practicalities of the reporting.

# 2 Quality checks classification

Quality checks (QC) in Reportnet 3 can be categorized in several ways to help users understand their purpose and impact. Two primary dimensions of classification are type of implementation (automated vs. manual) and level of data structure (field, record, or table level).

When looking at implementation types, we have **Automated** quality checks that are generated automatically by the system based on the structure and metadata of the dataset.

#### Examples of automated checks are:

- Ensuring that values in a column match the expected data type (e.g., numeric fields contain only numbers).
- Enforcing the provision of mandatory fields.
- Verifying the uniqueness of Primary Key values in a table.

Automated checks are integral to the validation process and help identify basic structural issues early in the submission workflow.

**Manual** quality checks, on the other hand, are used for more complex, rule-based validations defined by domain experts at the dataset design stage. Manual checks are tailored to address specific aspects of the data.

#### Examples of manual checks are:

- Verifying that a numeric value falls within a valid range, possibly dependent on the content of other columns.
- Implementing logical rules that go beyond simple structural validation.

When looking at data structure level, we have **Field** level checks that validate individual data elements, such as ensuring a value complies with the expected type or is not missing in a mandatory field. **Record** level checks evaluate conditions across multiple fields within the same row (record). For instance, if the value in one column imposes constraints on another column in the same row, a record-level check would apply. Finally, **Table** level checks involve the entire content of a table. They are typically used to confirm the completeness of the dataset, for example, verifying that all required records are present.

The level at which a QC is applied affects how messages are displayed in the Reportnet 3 user interface, which will be illustrated later in this document.

# 3 Quality checks severity levels

Each quality check in Reportnet 3 is assigned to a severity level, which determines the significance of the issue and its impact on the reporting process. Figure 1 shows the symbology associated to the error severity levels. There are currently four severity levels:

**INFO** level checks are used for Informational messages that highlight noteworthy aspects of the data. They may include summaries, report unusual values, or significant changes compared to previous submissions.

No action is required by the reporter.

**WARNING** level checks Indicate a potential issue that cannot be conclusively flagged as an error due to system limitations or ambiguous dataset definitions. For example, a field might be conditionally mandatory under complex rules that are difficult to evaluate programmatically.

• It is up to the reporter to assess the warning and decide whether corrections are needed.

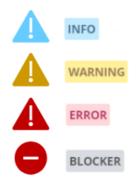
**ERROR** level checks signal a clear issue that, while not critical, still requires correction. An example could be a missing mandatory field that is not used in downstream processes or regulatory assessments.

Reporters are expected to correct these issues. Data requesters may follow up to ensure resolution.

**BLOCKER** level check represents critical problems that prevent the data from being processed or understood. This can be caused by incorrect data types (e.g., text instead of numbers) or invalid date formats.

• BLOCKER errors must be resolved by the reporter before the dataset can be submitted. A single blocker will prevent dataset release.

Figure 1. Error level severity errors and their representations



# 4 Validation process

The validation process is automatically performed during the release of the data; however, the reporter can trigger validation at any time during the reporting process. Validations are performed separately for each dataset. The reporter needs to separately validate the **ReportInfo** and **Establishments** datasets to perform a full validation of the dataflow content. Validation of a dataset starts by clicking on the **Validate** button [A] in the Reportnet 3 user interface Figure 2.

The validation functionality is only available when the table is not in edit mode, thus the user needs to first exit edit mode and then start the validation. Edit mode can be enabled and disabled using the rightmost button in position [C] in Figure 2.

The execution of the quality checks for the eSPIRS dataflow normally takes a few minutes, depending on the size of the dataset and the global workload the platform is experiencing. Upon completion of the process, the reporter can access the results of the validation by clicking on the **Show validation** button [**B**] which will show the validation dialog. The validation dialog presents a table listing all the QC error messages together with their severity level, their identification code, their entity level type, the table and field (if any) they refer to, and the number of records affected. The user can navigate the validation table by using the controls placed at the bottom [**D**] (see Figure 3 )or sort the columns by clicking on the header [**E**], or filter the content of the table using the controls at the top of the table. After applying the filters [**F**] the original content of the table can be restored by clicking the **Reset** button [**G**]. Clicking on a specific line of the validation table [**H**] will show the records in the table affected by the error.

After completion of the dataset validation process, error messages and icons are shown in the tab of any table where a QC rule fails. When viewing the table content, Field or Record entity type QC errors in a record are highlighted by an icon, corresponding to the highest severity level, placed in the **Validations** column [I] (see Figure 4). Hovering on the icon shows a tooltip bubble presenting the combined error messages of all the QC failing for the row. Field level errors are marked by icons in the specific field [J], while Table level errors are shown by an icon placed on top of the table [K].

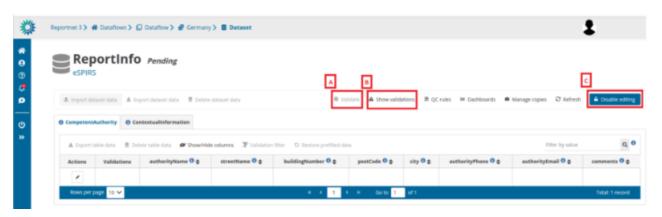


Figure 2. Reportnet 3 user interface. Controls relevant to dataset validation.

Figure 3. Validations dialog



Figure 4. Validation errors representation in tables



# 5 Quality checks description

In the eSPIRS dataflow, the **Establishments** dataset currently features over a hundred QCs, while the **ReportInfo** dataset has just around twenty. Most of these QCs are automated, and the vast majority have a BLOCKER severity level. Error messages associated with automatic QC are self-explanatory and usually not customized. For additional clarity, we provide, for each group of automated checks, a description of the underlying problem. In the following section, each group of checks is identified by the first part of its code and, when needed, by its name.

**FT (Field Type)**: checks that the value in the field is a valid representation of the field data type (INTEGER, DECIMAL, URL, DATE). An error is reported when the textual representation of the value cannot be converted to the underlying data type.

All FT QCs are BLOCKERs except the ones checking the validity of URL data type that are classified as ERROR.

#### **Examples:**

- A double decimal place in a decimal number 14..3444 instead of 14.3444
- A date provided in a different format 03-APR-2025 instead of 03/04/2025

Solution: correct the data to comply with the format specification. If the error persists, contact the helpdesk.

**FC (Field Cardinality)**: checks that a value is provided for a field marked as mandatory.

All FC QCs are BLOCKERs except the ones checking the presence of mandatory Establishment Address components (streetName, buildingNumber, city, postCode), and those checking the mainIndustryTypeNace, publicInformationURL that are classified as ERRORs. FC QC related to substance quantity declaration (quantityTNE) is also classified as ERROR.

Solution: Provide the missing data. If the error persists, contact the helpdesk.

**TB (Mandatory Table records check)**: checks that at least one record is provided in a table that is marked as mandatory

#### All TB QCs are BLOCKERs.

Solution: Provide the missing data if the data is not missing contact the helpdesk.

**TC (Field Type LINK)**: checks that LINK fields, when not empty, contain values matching an entry in the linked table. To solve these kinds of errors, the reporter should verify that the value inserted in the field corresponds to a value in the linked field. Linked tables can be tables within the same dataset or tables in other datasets or reference datasets. Please note that in some cases, in addition to the actual value used in the link, an additional element of the linked table is shown (as a label). However, only the value used in the link should be entered in the field.

#### All TB TCs are BLOCKERs.

Solution: Verify that the value in the link table exists and that there are no spelling errors in the specified value. If a value is missing in a reference dataset table or if the error persists even if the value seems to be correct, contact the helpdesk.

In addition to the automatic checks, several manual checks are also present for both the **Establishments** and **ReportInfo** datasets. The manual checks are designed to identify issues specific to the nature of the reported data. In the rest of the section, we describe each one of the manual checks and provide instructions on how to solve the underlying issues

## 5.1 Custom checks on text fields lengths

A first group of manual checks are designed to verify lengths of textual fields where there is a length constraint.

- **01\_FL\_establishmentName\_length.** Verifies that **establishmentName** does not exceed 255 characters.
- **02\_FL\_parentCompany\_length.** Verifies that **parentCompany** does not exceed 255 characters.
- **03\_FL\_streetName\_length.** Verifies that **streetName** does not exceed 255 characters.
- **04\_FL\_buildingNumber\_length.** Verifies that **buildingNumber** does not exceed 100 characters.
- **05\_FL\_city\_length.** Verifies that **city** does not exceed 60 characters.
- **06\_FL\_postCode\_length.** Verifies that **city** does not exceed 30 characters.
- **07\_FL\_publicInformationURL\_length.** Verifies that **publicInformationURL** does not exceed 255 characters.
- **08\_FL\_generalURL\_length.** Verifies that **generalURL** does not exceed 255 characters.
- **09\_FL\_lastInspectionURL\_length.** Verifies that **lastInspectionURL** does not exceed 255 characters.
- **10\_FL\_ comments\_length.** Verifies that **comments** does not exceed 500 characters.
- 11\_FL\_identifierScheme\_length. Verifies that identifierScheme does not exceed 80 characters.
- **12\_FL\_identifier\_length.** Verifies that **identifier** does not exceed 80 characters.
- 13\_FL\_siteIdentifierScheme\_length. Verifies that siteIdentifierScheme does not exceed 80 characters.
- 14 FL siteIdentifier length. Verifies that siteIdentifier does not exceed 80 characters.
- 17\_FL\_physicalProperties\_length. Verifies that physicalProperties does not exceed 255 characters.
- **18\_FL\_substanceComments\_length.** Verifies that **substanceComments** does not exceed 500 characters.

All FL QCs are classified as ERRORs except those concerning the inspireID of Establishments or Production Site that are BLOCKERs. FL QC concerning inspireIDs are treated separately.

Solution: Verify that the field length does not exceed the maximum allowed size.

## 5.2 Custom checks on InspireId format

The following manual checks verify that the format of the InspireId identifier is correct.

- **15\_FL\_inspireId\_length**. Checks that the total length of the **inspireId** does not exceed limit of 161 characters given by the sum of the maximum length of its components: **namespace** (80 characters), **localId** (80 characters) plus one separation character (forward slash).
- **01\_FC\_inspireId\_Format**. Checks that only valid characters are used. The only allowed characters are uppercase and lowercase letters, digits, colon [:], forward slash [/], dash [-] and underscore [\_]. No spaces or tabs are allowed.

**02\_FC\_inspireId\_Format**. Checks that the **inspireId** can be separated into its underlying components namespace and localId. The inspireId is built by concatenating the **namespace** and **localId** placing a forward slash between them.

**InspireIds** should contain at least one isolated forward slash; the text at the left of the last forward slash is considered the **namespace**, the text at the right is the **localid**. No forward slash can be used as the last character of the **inspireId** and the forward slash cannot appear next to another forward slash.

#### All InspireID QCs including those on Field Length are classified as BLOCKERs.

Solution: check that only allowed characters are used in the inspired and that a single forward slash is used to separate the namespace from the localld. Verify the total length of the inspired.

Similar checks (16\_FL\_siteInspireId\_length, 01\_FC\_siteInspireId\_Format, 02\_FC\_siteInspireId\_Format) are applied also to siteInspireId.

Example of InspireId composition:

Namespace: IT.CAED Localid: EST 00134234 Inspireld: IT.CAED/EST 00134234

## 5.3 Custom checks on productionSite data consistency

As explained in more detail in the data exchange format specification document, the eSPIRS data model was simplified in Reportnet 3 by combining the **establishment** and **productionSite** information in one table. When more than one establishment shares the same **productionSite**, the productionSite specific information (name, coordinates, thematic identifier, confidentiality classification) is repeated more than once, leading to possible inconsistencies. A set of custom QCs was introduced to detect when the records with the same **siteInspireId** contain different values of site related information.

- **01\_CC\_siteName**. Checks that the **siteName** field has the same values for records in **SevesoEstablishment** table having the same **siteInspireId**.
- **02\_CC\_siteIdentifierScheme**. Checks that the **siteIdentifierScheme** field has the same values for records in the **SevesoEstablishment** table having the same **siteInspireId**.
- **03\_CC\_siteIdentifier**. Checks that the **siteIdentifier** field has the same values for records in the **SevesoEstablishment** table having the same **siteInspireId**.
- **04\_CC\_siteLongitude**. Checks that the **siteLongitude** field has the same values for records in the **SevesoEstablishment** table having the same **siteInspireId**.
- **05\_CC\_siteLatitude**. Checks that the **siteLatitude** field has the same values for records in the **SevesoEstablishment** table having the same **siteInspireId**.
- **01\_CC\_siteNameConfidentiality**. Checks that the **siteNameConfidentiality** field has the same values for records in the **EstablishmentConfidentiality** table linked to records in **SevesoEstablishment** table having the same **siteInspireId**.
- **02\_CC\_siteLocationConfidentiality**. Checks that the **siteLocationConfidentiality** field has the same values for records in the **EstablishmentConfidentiality** table linked to records in **SevesoEstablishment** table having the same **siteInspireId**.

#### All consistency checks on productionSite are classified as BLOCKERs.

Solution: Correct the submission, making sure that the information on the same ProductionSite provided in different rows of the **Establishment** or **EstablishmentConfidentiality** table is the same.

## 5.4 Custom Spatial Checks

Approximate coordinates of the center of the areas belonging to establishments and production sites need to be reported in the **Establishments** dataflow. A number of QCs are in place to make sure that the reported coordinates are within the boundaries of the reporting country and that the locations of establishments and related production sites are compatible in terms of distance.

- **01\_SS\_siteLocationCountry**. Checks that the location of the reported **ProductionSite** is within 1km from the country border. The 1 km buffer zone is used to avoid errors due to different representation of complex boundary geometry.
- **02\_SS\_establishmentLocationCountry**. Checks that the location of the reported **SevesoEstablishment** is within 1km from the country border.
- **03\_SS\_EstablishmentSiteLocation**. Checks that the distance between **Establishment** and the **ProductionSite** does not exceed 500 m. This rule applies to 'isolated' Establishments where the **ProductionSite** is linked to a single **Establishment**.
- **04\_SS\_NonIsolatedEstablishmentSiteLocation**. Checks that the distance between **Establishment** and the **ProductionSite** does not exceed 2000 m. The larger threshold value applies to 'non-isolated' Establishments where multiple entities belong to the same **ProductionSite**. In these cases, it is expected that the point representing the site could be further away from the point representing the **Establishment**.
- **05\_SS\_longitude\_accuracy**. Checks that the **Establishment** longitude is provided with at least 5 digits after the decimal separator.
- **06\_SS\_latitude\_accuracy**. Checks that the **Establishment** latitude is provided with at least 5 digits after the decimal separator.
- **07\_SS\_siteLongitude\_accuracy**. Checks that the **ProductionSite** longitude is provided with at least 5 digits after the decimal separator.
- **08\_SS\_siteLatitude\_accuracy**. Checks that the **ProductionSite** latitude is provided with at least 5 digits after the decimal separator.

#### All spatial checks are classified as BLOCKERs.

Solution: verify and correct ProductionSite and/or Establishment coordinates. Contact helpdesk if the error persists.

## 5.5 Custom checks on linked tables data consistency

This group of custom checks verifies the consistency of entries reported in different tables.

**TR\_07\_ConfidentialityDeclarations**. Checks that each record in Establishment has a record in **EstablishmentConfidentiality** with the same establishmentId. Confidentiality declarations should be

provided for every Establishment even if all the confidentiality fields are empty to avoid unintended release or confidential information.

- **01\_CC\_industryType**. Checks that at least one value is provided either for SPIRS or NACE secondary industry Types in each row of **EstablishmentIndustryType**.
- **01\_TU\_IndustryTypeSPIRS**. Checks that the same SPIRS code is not declared multiple times for the same **Establishment**.
- **01\_TU\_IndustryTypeNACE**. Checks that the same NACE code is not declared multiple times for the same **Establishment**.
- **02\_CC\_industryTypeNACE**. Checks that the NACE code used as main industry type is not repeated as a secondary one.
- **02\_CC\_industryTypeSPIRS**. Checks that SPIRS code used as main industry type is not repeated as a secondary one.
- **01\_TU\_ThematicIdentifier**. Check that the combination of **thematicIdentifierScheme** and **thematicIdentifier** is not repeated in the **SevesoEstablishment**

#### All custom checks on linked tables consistency are classified as BLOCKERs.

Solution: Check the underlying data. Contact helpdesk if the error persists.

#### 5.6 Information Level checks

Information level checks provide the reporters summary information on the data delivery that could help identify potential issues, in particular with confidentiality declarations.

- **01\_TR\_EstablishmentCount**. Reports the total number of **Establishments** in the current data delivery.
- 02\_TR\_EstablishmentCountByTier. Reports the number of Establishments declared as Tier I and Tier II.
- **03\_TR\_EstablishmentCountByStatus.** Reports the number of **Establishments** declaring different values of operationalStatus.
- **04\_TR\_EstablishmentCountByConfidentialityFields.** Reports the number of **Establishments** declaring confidentiality in **EstablishmentConfidentiality** for each confidentiality field.
- **05\_TR\_EstablismnetReportingSubstances.** Reports the number of **Establishments** having at least one **Substance** declared
- **06\_TR\_ConfidentialSubstanceCount.** Reports the number of **Establishments** declaring at least one substance confidential and the number of confidentiality substance declarations.
- **07\_TR\_SecondaryIndustryDeclaration**. Reports the number of **Establishments** declaring at least one NACE or SPIRS secondary industry type in **EstablishmentIndustryType**.
- **08\_TR\_SiteCountByConfidentialityFields**. Reports the number of **ProductionSites** declaring confidentiality in **EstablishmentConfidentiality** for each confidentiality field.

# **Annex 1 – Glossary of Terms**

This section contains explanations of the key terms used in this document.

**Boolean attribute** – These attributes are evaluated as either true or false.

**Code list** – a defined set of values which can be used to populate an attribute.

**Data type** – A data model element that defines both characteristics of data and which operations can be performed on the data.

**Feature type** – Represents a class of data together with relevant attributes.

**Geospatial information** – Data about a physical object that can be represented by numeric values in a geographic coordinate system.

**INSPIRE** – Infrastructure for Spatial Information in the European Community. The INSPIRE Directive (2007/2/EC) aims to establish an infrastructure for the sharing of environmental spatial data within the European Union. This will enable sharing among public sector organisations, facilitate public access to spatial data across Europe, and will aid in cross-boundary policy making.

**Multiplicity** – A definition of cardinality - i.e. the permitted number of elements - of some collection of elements.

**NACE** - The 'statistical classification of economic activities' in the European Community, abbreviated as NACE, is the classification of economic activities in the European Union (EU).

**Production Site** – Represents the geographical location of the Seveso Establishment or a piece of land where the facility was, is, or is intended to be located.

**Seveso Establishment** – Represents the whole location under the control of an operator where dangerous substances are present in one or more installations, including common or related infrastructures or activities according to DIRECTIVE 2012/18/EU.