

## **ESS Validat**

**Single Grant Agreement B2794-2018-ESS-VAL-IMPL**

<http://www.cros-portal.eu/>

Single Grant Agreement Number **553-826006-2018-PT-ESS-VAL-IMPL**

### **Hybrid Validation Implementation**

#### **Deliverable 5.6.2**

**List of main types of VTL rules to translate to SQL**

**Version: 1.0**

## 1. VTL rules to translate to SQL

A set of the main types of validation rules<sup>1</sup>, covering more than 90% of the validation situations was identified and revised both by the Task Force on Validation<sup>2</sup> and the ESSnet ValiDat Integration<sup>3</sup>. These rules cover both structural and content validation. One of the inputs was an inventory of existing validation rules covering the following pilot statistical domains: National Accounts, Short Term Business Statistics, Energy Statistics, Asylum statistics, animal and milk statistics.

In the current Hybrid Validation Implementation Project<sup>4</sup> INE is interested in the content validation rules that can be pre-formatted as in a template and then filled with the correct parameters for any domain. Furthermore as in the main types of validation rules is studied the possibility to express the rule through VTL it is our aim to understand which and how the rules can be translated to SQL.

To this purpose we selected a domain selected for testing during the current project, Agriculture and Animal, which is already integrated in our Statistical Data Warehouse (SDW) on Oracle databases. Validation there is currently performed through SQL rules. In fact it is for now the only validation rules repository at INE. This repository is not centralized but rather organized and maintained by domain. However the possibility to have the main types of validation rules that could be parameterized for any domain would be a step towards centralizing and documenting the validation processes and also an opportunity to introduce it in domains where it is not yet available.

The 21 main types or rules identified were grouped by us according to their usability and the possibility of their being tested in the Agriculture and Animal domain.

Rule	Usage	In Test
(EVA) Envelope is Acceptable	Structural Validation	No
(FLF) File Format	Structural Validation	No
(FDD) Fields Delimiter	Structural Validation	No
(DES) Decimals Separator	Structural Validation	No
(FDT) Field Type	Structural Validation	No
(FDL) Field Length	Content Validation	YES
(FDM) Field is Mandatory or empty	Content Validation	YES
(COV) Codes are Valid	Content Validation	YES
(RWD) Records are Without Duplicate id-keys	Content Validation	YES
(REP) Records Expected are Provided	Content Validation	YES
(RTS) Records are all present for Time Series	Content Validation	YES

<sup>1</sup>

[https://ec.europa.eu/eurostat/cros/system/files/02c\\_main\\_types\\_of\\_data\\_validation\\_rules\\_and\\_fictive\\_domain\\_0.pdf](https://ec.europa.eu/eurostat/cros/system/files/02c_main_types_of_data_validation_rules_and_fictive_domain_0.pdf)

<sup>2</sup> <https://circabc.europa.eu/faces/jsp/extension/wai/navigation/container.jsp>

<sup>3</sup> [https://ec.europa.eu/eurostat/cros/content/essnet-validat-integration\\_en](https://ec.europa.eu/eurostat/cros/content/essnet-validat-integration_en)

<sup>4</sup> ESS Grant: 553-826006-2018-PT-ESS-VAL-IMPL

(RNR) Records' Number is in a Range	Content Validation	YES
(COC) Codes are Consistent	Content Validation	YES
(VIR) Values are in Range	Content Validation	YES
(VCO) Values are Consistent	Content Validation	YES
(VAD) Values for Aggregates are consistent with Details	Content Validation	YES
(VNO) Values are Not Outliers in Time Series	Content Validation	YES
(VSA) Values for Seasonally Adjusted data are plausible	Content Validation	No
(RRL) Records Revised are Limited in number or %	Eurostat's Reception	No
(VRT) Values are Revised within a Tolerance level	Eurostat's Reception	No
(VMP) Values for Mirror data are Plausible	Eurostat's Reception	No

All types of content validation with the exception of the seasonal adjustment will be examined. This means 12 types of validation rules will be inspected.

The selected types of rules may target microdata as well as aggregate data, meaning that they can be extended for our own national validation processes. Special attention will be given to the consistency of values to other data sets or even domains/sources as the data we will be targeting is integrated in the SDW. The results will be discussed in deliverables D10\_T13\_2 and D11\_T14\_2.