Data Validation and Transformation Support in SDMX

Validation and Transformation Language (VTL)

Stefano Pambianco – European Central Bank
Vincenzo Del Vecchio - Bank of Italy
SDMX TWG & VTL Task Force
Agenda

- Background and Objectives
- Core values
- Ongoing initiatives
- Future evolutions
The context

28 January 2021
SDMX Expert Group
Main original goals

Standardise the definition of **Transformations**

(and **Validations** as particular cases of Transformations)

by specialised **Subject Matter Experts** (non-IT definers)

For standardising, documenting and industrialising

the **Official Data Processing**

(inside and across institutions - data analysis by final users out of scope)

Through a language:

- **Declarative** (functions, not procedures)
- **Formal** (allowing parsing and translation in other languages for execution)
- **IT neutral** (independent of commercial IT tools and languages)
Agenda

• Background and Objectives
• Core values
• Ongoing initiatives
• Future evolution
Expression based

Transformation:

Result Data Set (percentage difference)

Operand
(Data Sets)

Expression

DS_r := (DS_2 - DS_1) / DS_1 * 100

Operand (non persistent)

Assignment

Operators
**Data set:** 1\textsuperscript{st} order function

Functional link between independent (e.g. date, country) and dependent (e.g. population) variables

**Operator & Transformation:** 2\textsuperscript{nd} order function

Functional link between input & output data sets

**Main benefits for users**

**Reproducible:** given one input \(\rightarrow\) one possible output

**Understandable:** simpler and clearer behaviour

**Transparent:** nothing hidden in complex code
**VTL 2.0 – main categories of operators**

- **Join** *(with functional constraints)*  
  - inner, left, full, cross

- **Set** *(of data points)*  
  - filter, union, intersection, diff., symm.diff.

- **on the Data Set space**  
  - various kinds of operations on variables

- **Validation**  
  - on datapoints, var., data sets, also according rulesets

- **Aggregation**  
  - according to hierarchical relationships or operations like  
    avg, count, max, sum … with group by, having clauses

- **Analytics**  
  - e.g. max/min, median, sum … and so on …  
    with partitioning, ordering, windowing clauses

- **String** *(concatenation / substring / pattern …)*, **Numeric** *(+, -, *, /, abs, mod, log, power and so on …)*, **Boolean** *(and / or / xor / not) … … …*

- **For the Time series**  
  - time shift, flow-stock conversions …

- **Comparison**  
  - equal, greater, between, element of, exist in …

- **User defined operators** and **External routine call**

- **Reusable rulesets**  
  - datapoint and hierarchical relationships
Agenda

• Background and Objectives
• Core values
• Ongoing initiatives
• Future evolutions
VTL over time

- March 2015: VTL 1.0
- Nov 2016: VTL 1.1 draft
- April 2018: VTL 2.0
- June 2019: SDMX adoption
- Implementing and sharing VTL tools and services

28 January 2021
SDMX Expert Group
Some ongoing VTL initiatives

ESCB
- BIRD – Banking Integrated Reporting Dictionary

NSOs - NCBs
- **Italian Institute of Statistics**: VTL ISTAT Framework
- **Bank of Italy**: VTL engine & editor
- **Statistics Finland**: Editing and imputat.service – (SAS engine)
- **Statistics Poland**: VTL - SQL translation
- **Statistics Norway**: Java-VTL

International Organisations
- **Eurostat**: VTL CONVAL VRM
- **European Central Bank**: Balance Sheet Items Pilot
  VALICE
• Background and Objectives
• Core values
• Ongoing initiatives
• Future evolutions
Upcoming activities

• Improvement of the language (e.g. new operators, new features)

• Upgrade of the SDMX implementation (VTL for SDMX 3.0)

• Compatibility kit (TBD)

• Support about VTL use (increasing requests)

• Support about VTL implementations (growing interest)
How to best support the future evolution of the standard?

VTL TF / SDMX TWG (language improvement and upgrade of SDMX implementation)

? User Support group (it should be permanent)

? Organisation for handling relationships with:
  ? VTL users / user community
  ? Developers of VTL tools (including private companies, e.g. software vendors)

? Other Suggestions
VTL

(Validation and Transformation Language)

International language for data validation and transformation

Thanks for your attention!

stefano.pambianco@ecb.europa.eu
vincenzo.delvecchio@bancaditalia.it