Modernisation of validation in the European Statistical System (ESS)

A multidimensional approach

UNECE-CES Work session on Statistical Data Editing
The Hague, 24-26 April 2017

Topic (ii) – Theme (iv)

vincent.tronet@ec.europa.eu
Eurostat, Unit B1
Objective: providing comparable statistics at EU level

Focus on EU policies, harmonisation extended to nearly all statistical fields

Partnership between Eurostat and NSI in EU and EFTA countries

Cooperation with candidate countries, Commission services, ECB and International Organisations
ESS Vision 2020

- Adopted in May 2014
- European response to challenges of official statistics

5 key areas of work:
1. Focus on users
2. Strive for quality
3. Harness new data sources
4. Promote efficiency in production processes
5. Improve dissemination and communication
Data validation in production chain
Focus on data sent to Eurostat

Validation can take place in several points of the ESS statistical production process
Current situation

Member State

Design validation rules → Validate Data

Validation rules

Data

Validation report

No

Accept data?

Yes

Design validation rules → Validate Data

Validation rules

Eurostat
Main shortcomings in current situation

- No clear picture of who in the ESS is doing what as regards validation:
  - Risk of validation gaps
  - Time-consuming validation "ping-pong"
  - Potential subjectivity in assessment of data quality

- Lack of standards and of a common architecture for shareable and reusable validation services:
  - Duplication of IT development costs within the ESS
  - Manual work due to low integration between the different tools
## Medium-term goals for ESS validation

<table>
<thead>
<tr>
<th>Medium-term goals</th>
<th>Business Outcomes</th>
</tr>
</thead>
</table>
| **Goal 1:** Ensure the transparency of the validation procedures applied to the data sent to Eurostat by the ESS Member States. | Increase in the quality and credibility of European statistics
Reduction of costs related to the time-consuming validation cycle in the ESS ("validation Ping-Pong") |
| **Goal 2:** Enable sharing and re-use of validation services across the ESS on a voluntary basis. | Reduction of costs related to IT development and maintenance |
A multidimensional approach

A. Methodology
B. Processes/governance
C. Information standards
D. IT resources
E. Human resources
A. Methodology

Need for a common framework for validation in the ESS

=> methodological handbook:

- Definitions
- Classifications
- Metrics
B. Processes/Governance

Business architecture for ESS Validation

- Eurostat and Member States agree at domain specific Working Groups on validation rules, severity and roles
- Validation principles
- Standard process for determining acceptability of data
- Main data flows needed for efficient validation process
- Possible scenarios for reusing validation services in countries to validate data prior to official transmission
Key principles for validation in business architecture

- The sooner, the better (Eurostat cannot correct country data)
- Trust, but verify
- Well-documented and appropriately communicated validation rules
- Well-documented and appropriately communicated validation errors
- Comply or explain
- Good enough is the new perfect
To-be situation

Member State

Design Validation Rules

Validate Data

Data

Validation report

Accept data?

No

Yes

Validation rules

Eurostat
3 optional IT scenarios for countries

1. Connect to central repositories
   - Installation of service or tool in own environment
   - Only access to repositories (SDMX DSDs, validation rules)

2. Connect to CSPA compliant service
   - Integration of service into own workflow

3. Connect to the process
   - Transmission to Eurostat required
   - Eurostat runs service and delivers report
C. Information standards

Need for standardization of the 3 types of objects exchanged with Member States:

- **Data**: SDMX information model (any structured data)
- **Validation rules**: VTL (Validation and Transformation Language) for statisticians and machines
- **Standard validation reports** for statisticians and machines
D. Information Technology in ESS
=> 3 CSPA services, 2 Registries

Validation Rule Manager

Data Structure Registry (e.g. SDMX registry)

Structural Validation Service

Validation Rule Registry

Content Validation Service
D. Information Technology in ESS => 3 CSPA services, 2 Registries

- **Structural Validation** service based on a Data Structure Registry (e.g. SDMX Registry)

- **Content Validation** service that checks consistency and plausibility of data based on rules stored in VTL in a Validation Rule Registry

- **Validation Rule Manager** to view, create, modify and manage VTL validation rules stored in the Validation Rule Registry.
E. Human resources

- **Communication** (e.g. Validation Portal)
- **Training** (ESS course on 21-22 Nov and 4 Regional conferences planned between end 2017 and Q1/2018)
- **Work** to describe validation rules, agree on them and keep them up to date
Concrete use case: National Accounts

**Objective:** Re-design of the National Accounts statistical Production System using Eurostat corporate (CSPA compliant) services for managing and validating incoming data files.
Initial situation (3 parallel processes)

FAME
Oracle RDBMS
Oracle OLAP
Phase 1: Common structural validation

- SOA Service
- FAME
- Oracle
- Oracle OLAP
Phase 2: Common structural and content validation

- **SOA Service**
- **FAME**
- **Oracle**
- **Oracle OLAP**
Future: More common statistical services
Conclusions

- Modernisation of validation in the ESS will contribute to lower the burden, increase quality and transparency
- Cooperation and International standards are essential
- Components are progressively being put in place
- Limited focus in a first stage: Data sent to Eurostat
- Deliverables potentially extended in future to cover the whole statistical process in countries and at Eurostat
- First results already visible
QUESTIONS ?