Business Architecture for ESS Validation

Table of Contents

1.0 Introduction

1.1 Purpose

In January 2013, the ESSC launched the ESS Vision 2020 Validation project. By providing foundational methodological and architectural frameworks, the ESS Vision 2020 Validation project, which ended in November 2015, laid the groundwork for the modernisation of the way data validation is performed in the ESS.

This Business Architecture document builds upon the deliverables of the ESS Vision 2020 Validation project. According to TOGAF, a widely used reference framework for Enterprise Architecture, the Business Architecture *"describes the product and/or service strategy, and the organizational, functional, process, information, and geographic aspects of the business environment"*. Its purpose is to provide a common understanding of how ESS validation should be conducted in the future – i.e. provide a comprehensive description of the target state for ESS validation. In order to accomplish this, the current document comprises several parts:

- Chapter 2 summarises the drivers for the modernising validation in the ESS. It also specifies the scope of this modernisation initiative.
- Chapter 3 defines the medium-term goals for validation in the ESS by identifying the main capabilities to be developed. It also contextualises these capabilities in the framework of the overarching ESS modernisation strategy as outlined in the ESS Vision 2020.
- Chapter 4 introduces a list of general validation principles. These principles provide the theoretical underpinning for all the major design decisions made in the course of the elaboration of the target to-be state.
- Chapters 5 and 6 provide a description of the target to-be state for validation in the ESS. Chapter 5 focuses on giving a high-level overview of this target state, while chapter 6 offers a more thorough description of its individual components.

This Business Architecture document follows the approach and principles set out in the ESS Enterprise Architecture Reference Framework (ESS EARF). When relevant, special attention has been given to ensure that the content of the Business Architecture is aligned with widely used reference standards such as GSBPM and GSIM.

1.2 Reader

The current Business Architecture document is designed to be a high-level communication tool on the objectives of the envisaged target state for validation in the ESS. The intended audiences are therefore ESS business and IT managers.

It should be however noted that some of the terminology used in this document is specific to the field of enterprise architecture and may sound foreign to readers who are not familiar with this discipline. Throughout the document, great care has been taken to define and explain technical terms when necessary. Moreover, a short glossary has been included at the end of the document. We refer the reader to the ESS EARF for more in-depth explanations.

2.0 Context

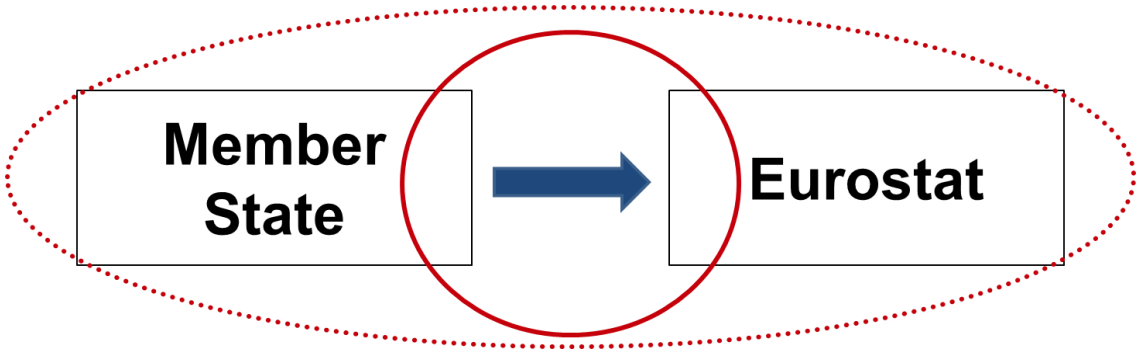
2.1 Scope

According to the Methodological Handbook on data validation written by the ESSnet Validat Foundation, data validation is “an activity verifying whether or not a combination of values is a member of a set of acceptable combinations”. Data validation focuses on the detection of errors in the data and is a distinct activity from data editing and data imputation, which focus on their correction.

One of the defining characteristics of the production of European statistics is the fact that the production process is distributed across several organisations. Data collection and a first round of processing are under the responsibility of ESS Member States. The data are then transmitted to Eurostat, where the data are processed further and finally disseminated at European level.

Data validation activities occur at several points in this statistical production chain. However, one key step in the ESS statistical production is the validation of the data sent to Eurostat by Member States. This is the step that ensures that the data coming from different national authorities abide by common consistency and coherence requirements and is thus essential in turning national statistics into European statistics. This is the step that this Business Architecture document concentrates primarily on.

While the focus of this business architecture is the validation of the data sent by Member States to Eurostat, the IT solutions envisaged in this document aspire to be reusable, on an optional basis, for the modernisation of national validation processes. In order to ensure conformity with national requirements and to maximise the reuse potential of the IT solutions to be developed, the Business and IT Architecture takes into account the outcomes of the 2015 ESSnet "ValiDat Foundation" and the recommendations of the Validation Task Force launched by the WG Methodology in 2016. Cooperation with Member States will continue throughout the implementation of the proposed architecture.



The continuous line represents the scope of this business architecture. The IT solutions are however meant to be applicable outside this original scope, as represented by the dotted line.

2.2 Drivers for change

The validation of data sent by Member States to Eurostat is a joint effort involving both national data providers (i.e. NSIs or other national administrations) and Eurostat. Together, these organisations must ensure that the coherence and consistency of the data they exchange is in line with expected quality standards. The overall quality of the ESS data validation process is therefore heavily dependent on the quality and depth of the collaboration between Eurostat and national data providers.

While they vary considerably between different domains, current validation practices exhibit shortcomings which could be corrected through strengthened collaboration. The main such shortcomings are listed below:

1. In several domains, the lack of a clear repartition of validation responsibilities among the different partners involved in the production process leads to double-work in the ESS and to the risk of "validation gaps", i.e. to cases where essential validation procedures are not carried out by any of the actors.
2. The lack of shared and easily accessible documentation on validation procedures can lead to time-consuming misunderstandings between Eurostat and ESS data providers when data validation problems arise (this phenomenon has been dubbed "validation ping-pong"). It can also lead to difficulties in assessing whether the quality assurance mechanisms applied to data sent to Eurostat are "fit-for-purpose".
3. The lack of common standards for validation solutions leads to a duplication of IT development and integration costs in the ESS. Moreover, the ESS is currently incurring high opportunity costs by not exploiting the general trend in the IT world towards Service-Oriented Architecture (SOA) and its potential benefits in terms of reuse and sharing of software components.

3.0 Objectives

3.1 Vision for ESS validation

The drivers for change identified in the previous chapter highlight the flaws in the current validation process for the data sent to Eurostat by Member States. The following vision statement articulates the response to these drivers and provides a compact description of the goals for validation in the ESS.

Response to drivers 1 and 2

Establish a transparent business process for validation in the ESS and support it via an integrated IT architecture allowing for the sharing and reuse of validation services among ESS members.

Response to driver 3

The vision statement above represents a translation of the general ESS Vision 2020 goals into validation-specific goals. In particular, as highlighted by the quotes below, the modernisation of data validation in the ESS will contribute to the implementation of two ESS Vision 2020 key areas: quality and efficient statistical processes.

> *"We will further enhance the existing approach to quality assurance with appropriate and effective quality assurance tools for all elements of the statistical life cycle."*
>
> ESS Vision 2020, Key area: Quality

> *"We will intensify our collaboration by further intensifying the sharing of knowledge, experiences and methodologies but also by sharing tools, data, services and resources where appropriate. The collaboration will be based on agreed standards and common elements of technological and statistical infrastructure."*
>
> ESS Vision 2020, Key area: Efficient statistical processes

3.2 Target business capabilities

The goals stated in the previous section's vision statement can also be expressed in terms of business capabilities. Business capabilities express what an organisation wants to be able to do in the future. Achieving them requires a combination of several key dimensions: human resources, methodology, information standards, IT and processes/governance.

In the table below, the vision statement has been translated into two target capabilities. These two validation-specific capabilities are mapped to the ESS EARF's Business Capability Model in order to correctly position the objectives for ESS validation in the wider context of the implementation of the ESS Vision 2020.

Target capabilities	
Capability 1: Ability to ensure the transparency of the validation procedures applied to the data sent to Eurostat by the ESS Member States.	**Capability 2:** Ability to share and reuse validation services across the ESS on a voluntary basis.

Mapping to ESS EARF capabilities	
Process & Workflow design	Design production system, statistical processing services and rules
Expected benefits	
Increase in the quality and credibility of European statistics Reduction of "validation ping-pong"	Reduction in IT maintenance and development costs

4.0 Validation principles

In 2016, Eurostat launched a Validation Task Force to support the implementation of the deliverables of the ESS Vision 2020 Validation project. The Validation Task Force created a list of 6 basic principles to be kept in mind when designing validation processes. The complete list of these principles can be found in Annex A.

These principles are generally applicable to all stages of data validation in all statistical domains, both between and within organisations. The target state for validation in the ESS outlined in this document is based on these principles. When relevant, the document will explain how these principles have been applied and translated into specific design decisions.

The validation principles take into account and were in part inspired by the generic principles listed in the ESS EARF. In particular, great care has been taken to make sure that the validation principles properly reflect the ESS EARF principles referring to the management of information in a statistical process.

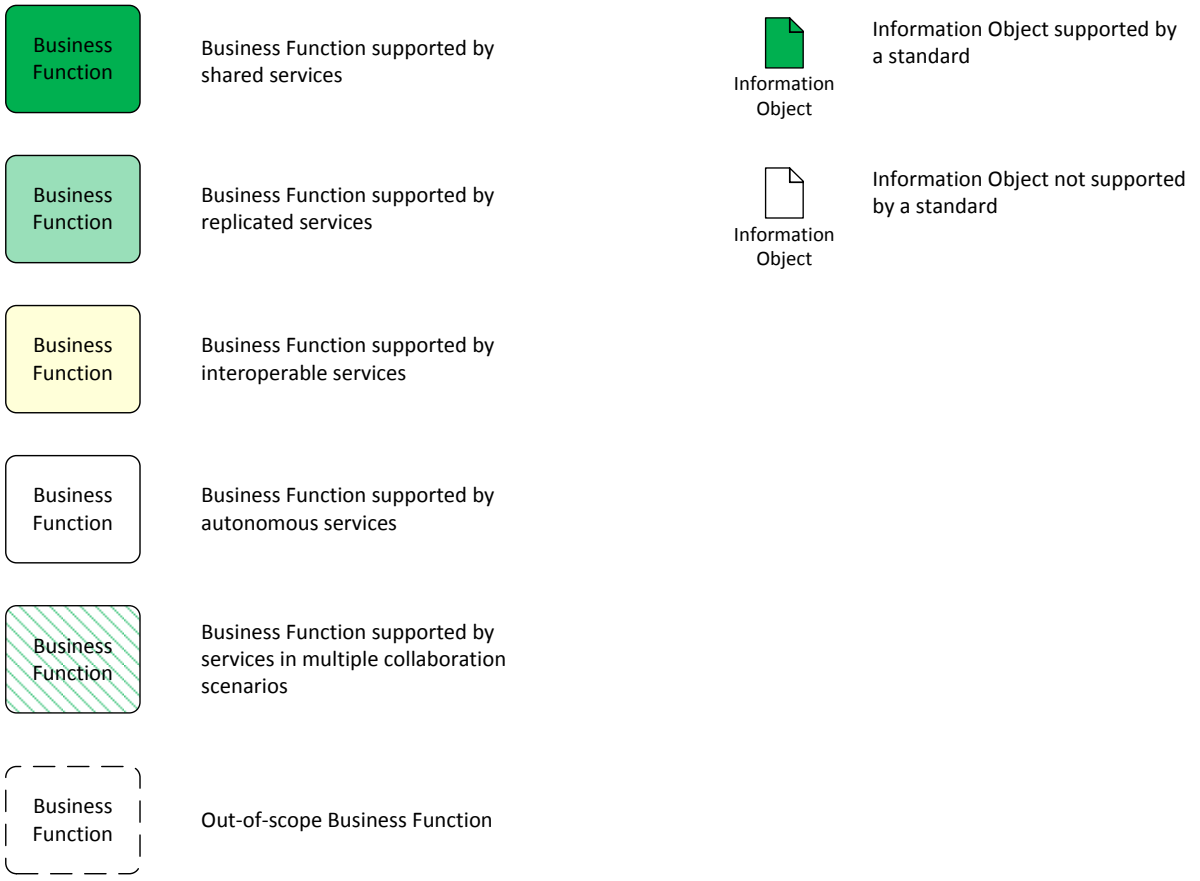
5.0 To-be state for validation in the ESS – overview

This section is dedicated to giving a high-level overview of the target to-be state for validation in the ESS. In particular, this section will describe:

- The target business process and the main business functions involved (a definition of business functions can be found in the glossary).
- The properties and characteristics of the information objects that are expected to be created and exchanged over the course of the validation process.

- The main IT services foreseen to support the target business process and their mutual interaction.

For all the business process diagrams in this section, the legend below will be used. Business Functions will be marked differently depending on the collaboration scenario of the services that support them (see the glossary for a definition of each collaboration scenario). Information objects are also marked differently according to whether or not they are supported by a standard.



5.1 Business view

The target business process for validation in the ESS is represented in the diagram on page 9. It comprises the following business functions:

- *Design data structures*: Eurostat and Member States jointly define, at Working Group level, the structure and format of the data to be sent by Member States to Eurostat. The data structure is documented using accepted metadata standards.

- *Design validation rules*: Once the data structure is agreed upon, Eurostat and Member States jointly define, at Working Group level, the validation rules which

must be applied to the data. Eurostat and Member States assign a severity level to each validation rule and determine which organisation is responsible for applying it. The validation rules are documented using a common ESS standard.

- *Validate data:* Prior to sending the data to Eurostat, Member States apply the validation rules they are responsible for. When Eurostat receives the data, it verifies that Member States have discharged their validation duties as expected and applies the validation rules under its own responsibility. The outcome of Eurostat's validation procedure is a validation report which is sent back to the Member States. This validation report follows a standard structure.

- *Accept data:* Based on the validation report produced by the validation procedures, Eurostat determines whether the data can be accepted or not. This acceptance procedure follows a standard and agreed upon process (see section 6.4). If the data are accepted, the validation business process ends. If the data are not accepted, a request for clarification or for resubmission of the data is sent to the Member State.

5.2 Information view

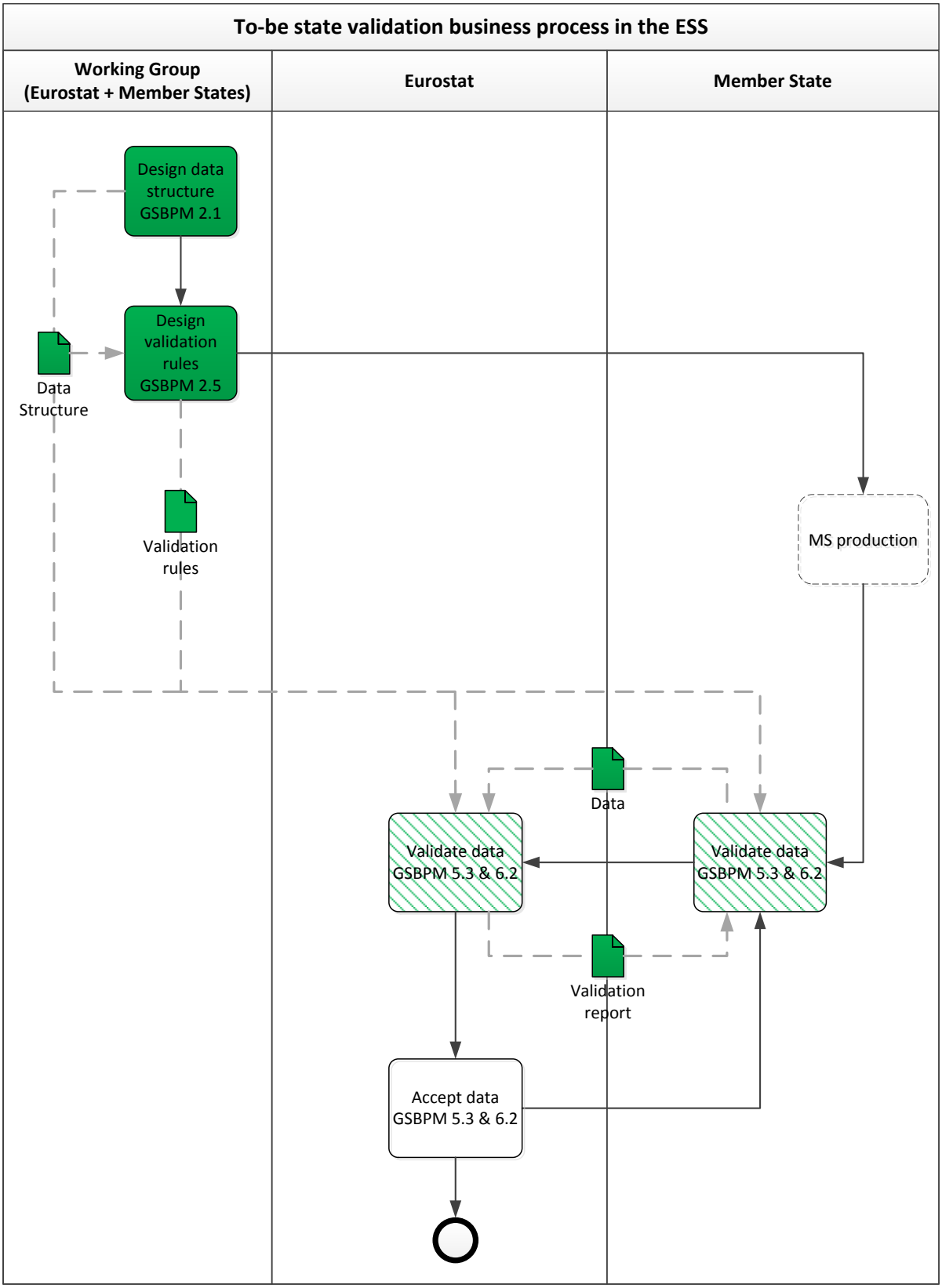
The target business process described in the preceding section foresees the creation and exchange of four kinds of information objects: data structures, data, validation rules and validation reports. It is expected that, in the target state, these information objects will be described using specific standards. A short definition of each information object is provided below.

- *Data structure:* this information object comprises the structural metadata that are needed to identify, use, and process data matrixes and data cubes, e.g. names of columns or dimensions of statistical cubes.

- *Data:* in the current context, the term "data" indicates the statistical information Member States provide to Eurostat.

- *Validation rule:* validation rules are mathematical expressions defining acceptable combinations of values. Data are said to satisfy the rules when the combination expressed by the rules is not violated. Besides the mathematical expressions themselves, this information is also understood to comprise relevant additional information regarding how a rule should be processed (e.g. the severity level and the responsible actor) or facilitating a classification of the rules (e.g. the type of rule).

- *Validation report:* a validation report is a document summarising and communicating the outcomes of a data validation procedure. It indicates which combinations of data values failed to satisfy specific rules. A validation report may also contain relevant additional metrics about the validation process (e.g. the overall processing time).


To-be state validation business process in the ESS
Working Group (Eurostat + Member States)
Eurostat
Member State
Design data structure GSBPM 2.1
Design validation rules GSBPM 2.5
Data Structure
Validation rules
MS production
Data
Validate data GSBPM 5.3 & 6.2
Validate data GSBPM 5.3 & 6.2
Validation report
Accept data GSBPM 5.3 & 6.2

The table below specifies the inputs and outputs of the business functions discussed in the previous section in terms of the four information objects defined above.

	Data structure	Data	Validation rule	Validation report
Design data structures				
Design validation rules	✓			
Validate data	✓	✓	✓	
Accept data				✓

Table 1- Business function inputs

	Data structure	Data	Validation rule	Validation report
Design data structures	✓			
Design validation rules			✓	
Validate data				✓
Accept data				

Table 2 - Business function outputs

The information objects can also be classified according to their degree of sensitivity. In particular, data and validation reports may in some cases be sensitive or confidential (validation reports related to confidential data may contain confidential information). Data structures and validation rules are never sensitive or confidential.

5.3 Application view

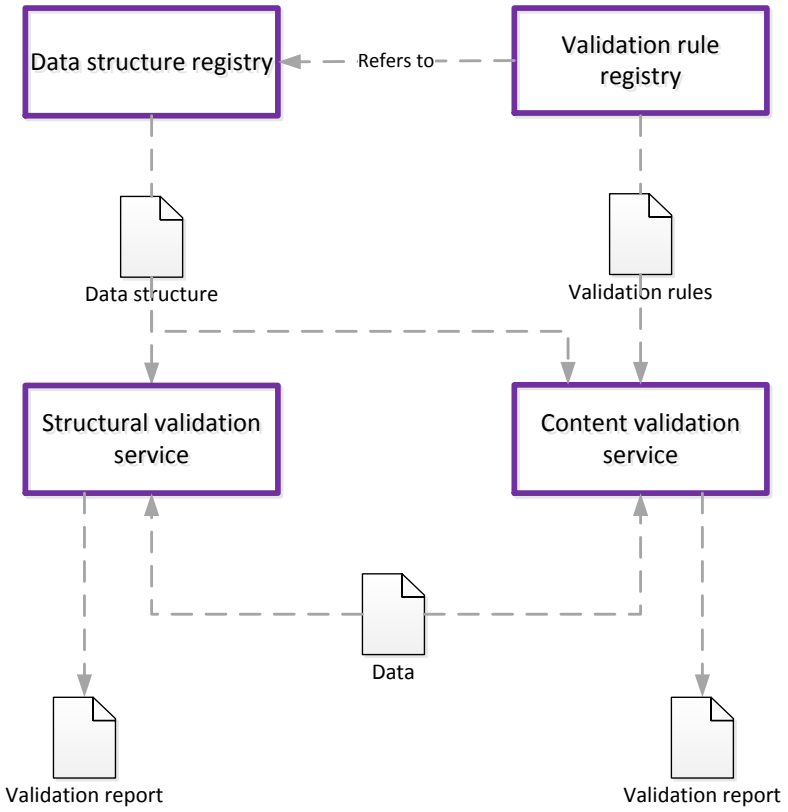
The different business functions foreseen in the target to-be state are expected to be supported by IT services. The ESS Enterprise Architecture Reference Framework classifies IT services according to collaboration scenarios, i.e. according to the expected degree of collaboration among ESS members in the use of the service. The ESS Enterprise Architecture Reference Framework identifies four possible collaboration scenarios:

- Autonomous: Services are designed and operated without coordination with other ESS members;
- Interoperable: ESS members have the autonomy to design and operate their own services, as long as they have the ability to exchange information and operate together effectively;
- Replicated: Services are duplicated: ESS members implement an instance of a generic service in their local environment;
- Shared: Services are common, shared and accessible to all the ESS members. There is a single instance that is shared and available to all. The table below outlines which services are foreseen for each business function.

The following table identifies, for each business function, the IT services foreseen and the corresponding collaboration scenario.

Business function	Supporting service	Collaboration scenario	Comments
Design data structures	Registry enabling the creation, management and retrieval of data structures	Shared	The centrally hosted registry will give access to the data structures to all authorized users
Design validation rules	Registry enabling the creation, management and retrieval of validation rules	Shared	The centrally hosted registry will give access to the validation rules to all authorized users
Validate data	Granular validation services executing previously defined validation rules	Shared/ Replicated/ Interoperable/ Autonomous	Member States can choose to which extent they want to use ESS validation services to validate their data prior to transmission to Eurostat (see section 6.3)
Accept data	No specific service foreseen	N/A	

Based on the inputs and outputs of each business function (see section 5.2), the following diagram shows the expected information flow between the different services. The diagram considers a scenario where two validation services are made available: one for validating the data structure and one for validating the content of the data.



6.0 Detailed description of business functions

6.1 Design data structure

Description

The first step in the target ESS validation process is the definition of the structure and format of the data files to be sent to Eurostat. This step should be conducted jointly by Eurostat and the Member States through consultations in each specific domain's Working Group. This business function is important for validation for two distinct reasons:

- It implicitly provides a first set of validation rules related to the expected structure of the data file.
- Since the availability of a common data structure is a precondition for the definition of content-related validation rules, it provides the foundation for the work that will be carried out as part of the *Design validation rules* business function.

The main output of this Business Function is a standard description of the expected data structure. The definition of which metadata standard should be used for data structures falls outside of the scope of this validation architecture. However, it should be noted that, in recent years, SDMX has played an increasing role in fulfilling this role. Around 40% of European statistical production processes are now describing their data structures using the SDMX formalism. The emergence of SDMX has enabled the creation of shared services in support of this Business Function (e.g. the Euro-SDMX Registry and the SDMX Global Registry).

Business Function Name		Design data structure
GSBPM reference		Design outputs (2.1)
Actor		Working Group
Supporting IT service	**Description**	Registry enabling the creation, management and retrieval of data structures
	Collaboration scenario	Shared
	Link to SPRA service	Design data collection instruments
Input	**Description**	N/A
	GSIM object	N/A
	Standard available	N/A

Output	**Description**	Structure of the datasets to be sent to Eurostat
	GSIM object	Data Structure
	Standard used	Metadata standards (e.g. SDMX) to be defined by the relevant ESS bodies

Relevant principles

As this business function creates a shared definition of the structural validation rules for the data transmitted by Member States to Eurostat, it contributes to fulfilling the recommendations expressed in principle 3, *Well-documented and appropriately communicated validation rules*.

6.2 Design validation rules

Description

In order to guarantee the transparency of the validation process, Eurostat and Member States jointly define at Working Group level the validation rules to be applied to the data sent to Eurostat. In order to support this work and facilitate communication between the different actors involved, a common standard for the description of validation rules should be developed and a shared registry of validation rules should be created.

In particular, Eurostat and Member States should jointly define who among them is responsible for applying a certain validation rule. They should also define the severity level associated to each rule. The possible severity levels and their implications are detailed in section 6.4.

Business Function Name		Design validation rules
GSBPM reference		Design processing & analysis (2.5)
Actor		Working Group
Supporting IT service	**Description**	Registry enabling the creation, management and retrieval of validation rules
	Collaboration scenario	Shared
	Link to SPRA service	Design data collection instruments

Input	**Description**	Structure of the datasets to be sent to Eurostat
	GSIM object	Data Structure
	Standard available	Metadata standards (e.g. SDMX) to be defined by the relevant ESS bodies
Output	**Description**	Validation rules
	GSIM object	Rule
	Standard used	ESS Standard for validation rules

Relevant principles

This business function contributes to fulfilling the recommendations expressed in principle 3, *Well-documented and appropriately communicated validation rules*.

In the definition of the validation rules, and in particular when assigning validation responsibilities or severity levels, two further principles should be taken into account:

- Principle 1, The sooner, the better: validation responsibilities should be assigned to the first actor in the production chain who has sufficient knowledge to properly verify the validity of the data.
- Principle 6, Good enough is the new perfect: the design of the validation rules should strike the right balance between the need for data quality and the additional burden validation may impose in the whole production chain. The number of validation rules and their severity level should be adjusted accordingly.

6.3 Validate data

Description

The *Validate Data* business function is subdivided into two steps:

- Member States, prior to transmitting their data to Eurostat, validate them by applying the validation rules for which they are responsible, as defined in the *Design validation rules* business function.
- Upon transmission of the data to Eurostat, Eurostat verifies that Member States have correctly applied the validation rules under their responsibility and performs the additional validation checks it is responsible for. The results of this validation procedure are sent back to Member States under the form of a standard validation report.

Business Function Name		Validate Data
GSBPM reference		Review & validate (5.3) Validate outputs (6.2)
Actor		Eurostat Member States
Supporting IT service	Description	Granular validation services executing previously defined validation rules
	Collaboration scenario	Shared/Replicated/Interoperable/Autonomous
	Link to SPRA service	Validate
Input 1	Description	Structure of the datasets to be sent to Eurostat
	GSIM object	Data Structure
	Standard available	Metadata standards (e.g. SDMX) to be defined by the relevant ESS bodies
Input 2	Description	Validation rules
	GSIM object	Rule
	Standard available	ESS Standard for validation rules
Input 3	Description	Data
	GSIM object	Data set
	Standard available	Standard formats (e.g. SDMX-ML) to be defined by the relevant ESS bodies
Output	Description	Validation report
	GSIM object	Process Metric
	Standard used	ESS Standard for validation reports

In order to support Member States in the validation of their data prior to transmission to Eurostat, shareable and reusable ESS validation services will be developed and made available. Each Member States will be able to freely choose, for each statistical production process, the extent to which it wants to use these services. We can thus delineate three basic scenarios for how, in the target to-be state, Member States could apply the agreed upon validation rules before transmitting the data to Eurostat.

It should be noted that the scenarios illustrated below represent somewhat idealised archetypes. In real-life situations, it is probable that Member States will create hybrid scenarios which will incorporate elements of two or more scenarios. Each Member State would be free to mix and match the three scenarios as it sees fit.

Scenario 1: Autonomous / Interoperable validation services

Under this scenario, Member States use their own autonomous or interoperable services to execute the validation rules prior to data transmission to Eurostat. These autonomous or interoperable services would however use the jointly agreed upon data structures and validation rules, which will be stored in centrally hosted registries. The translation of these validation rules into the autonomous or interoperable validation services would be the responsibility of each individual Member State.

Illustration of the scenario

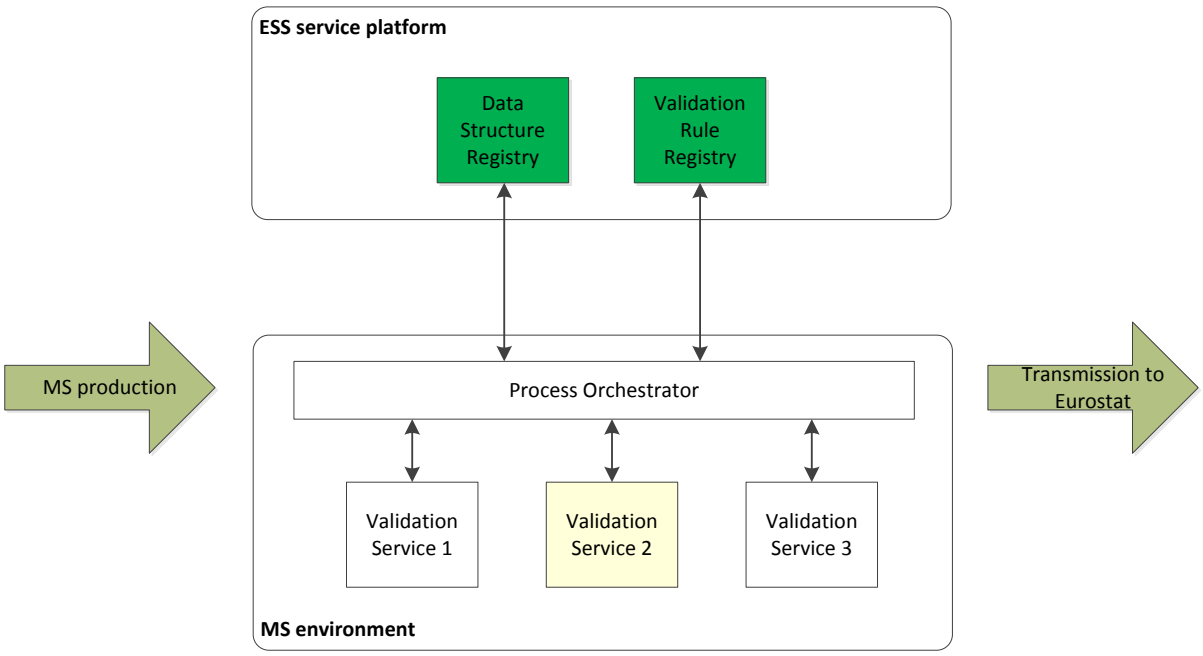
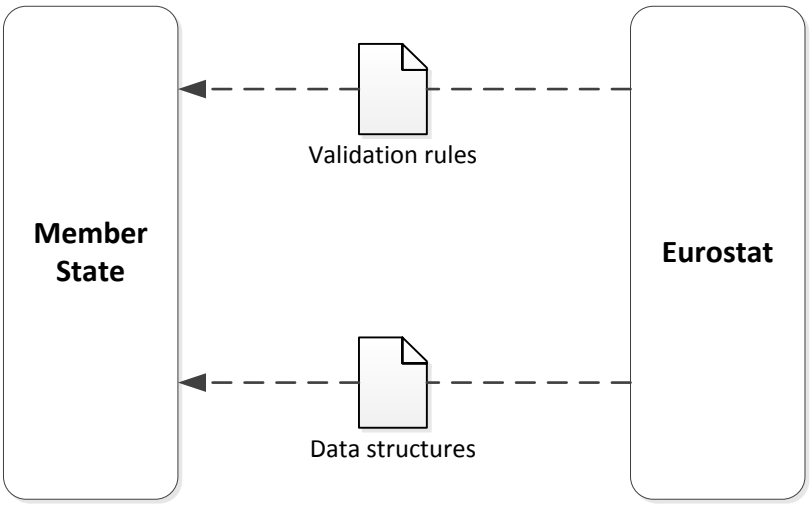


Illustration of the information flows in this scenario



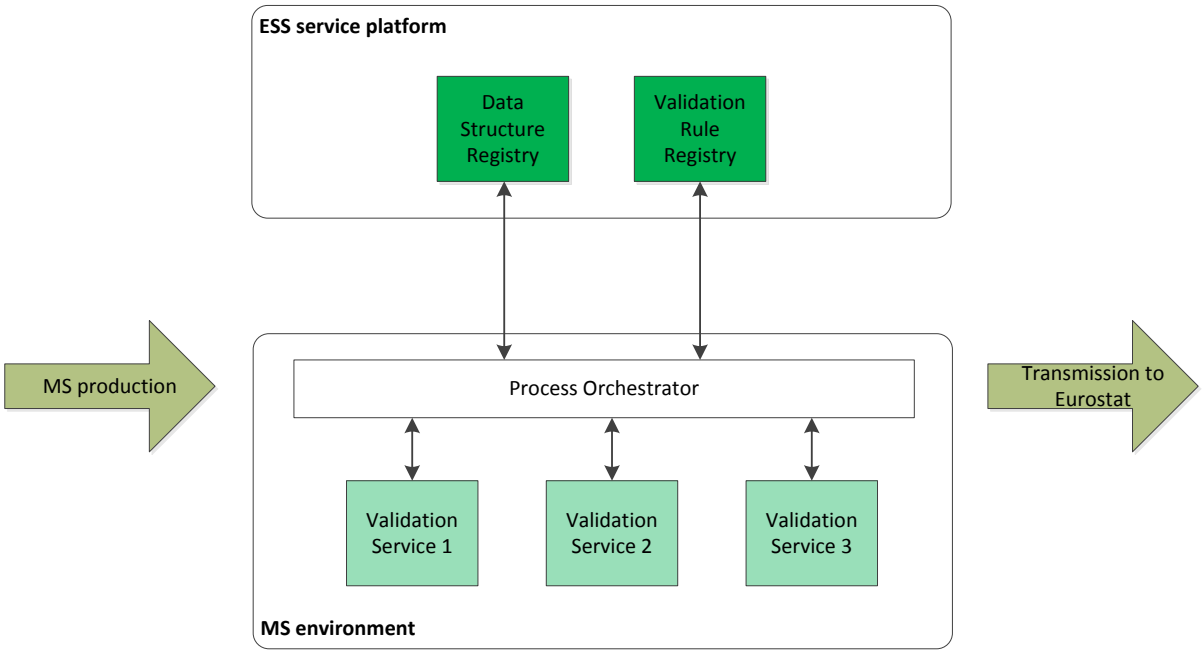
In this scenario, validation rules and data structures travel from Eurostat to Member States. However, as no potentially confidential information is exchanged between the Member State environment and centrally hosted services, this scenario poses no particular security concerns.

Scenario 2: Replicated/Shared validation services

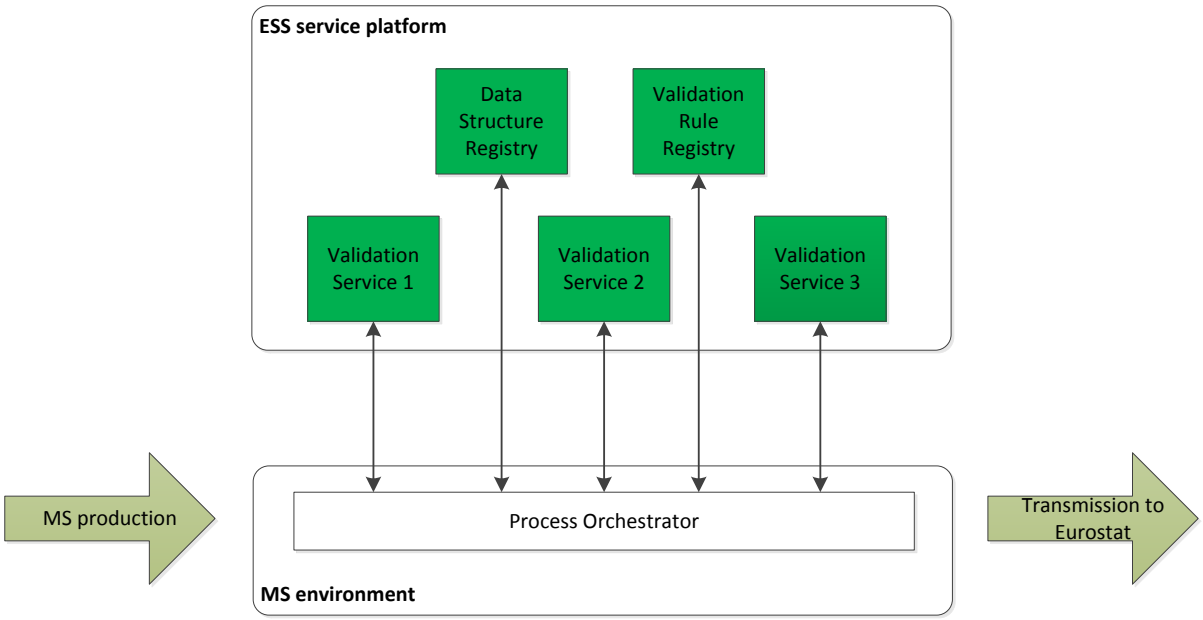
Under this scenario, in addition to the shared registries for data structures and validation rules, Member States use replicated and/or shared validation services in their validation process. Member States would remain responsible for the orchestration of the different services used in the validation process. This scenario can be equated with the Software as a Service (SaaS) paradigm in cloud computing.

For the sake of clarity, the diagrams below will represent two possible configurations: one where Member States use only replicated services and one where Member States use only shared services. It should however be stressed that Member States are free to use whichever combination of shared or replicated services they prefer.

Illustration of the scenario

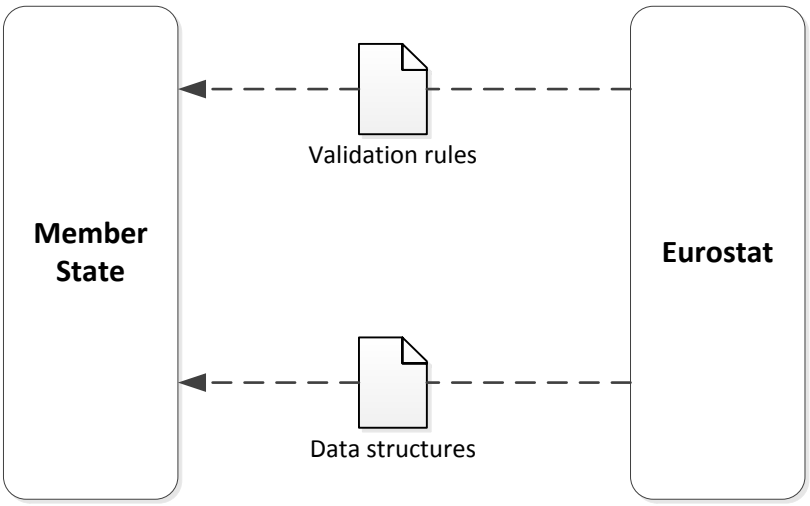


Scenario 2 if only replicated services are used

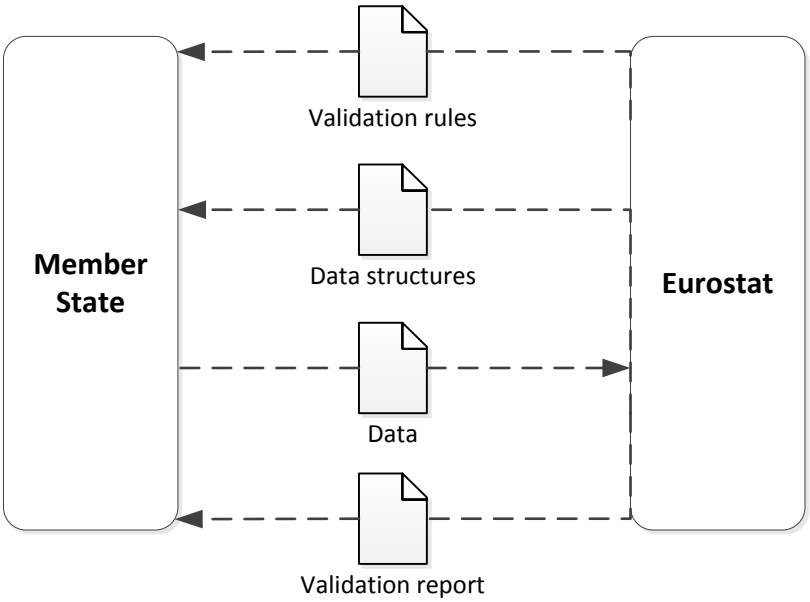


Scenario 2 if only shared services are used

Illustration of the information flows in this scenario



Information flows if only replicated services are used



Information flows if only shared services are used

In the case only replicated services are used, the information flows are similar to those for scenario 1. Consequently, there should be no major security concerns, as no confidential data crosses organisational boundaries. However, security concerns might arise in case a Member State wants to use an ESS validation service in the "shared" collaboration mode for confidential data. Appropriate arrangements for a secure transmission of data to and from the centrally hosted services will need to be taken in order to enable the use of this scenario in the case of confidential data. Member States should check the compatibility of this scenario with their national legal framework.

Scenario 3: Shared validation process

Under this scenario, Member States would delegate the validation of their data to a centrally predefined shared validation process. This shared validation process would take care of the orchestration of the various services needed and would return to the Member State a comprehensive validation report. This scenario can be equated with the Business Process as a Service (BPaaS) paradigm in cloud computing.

Illustration of the scenario

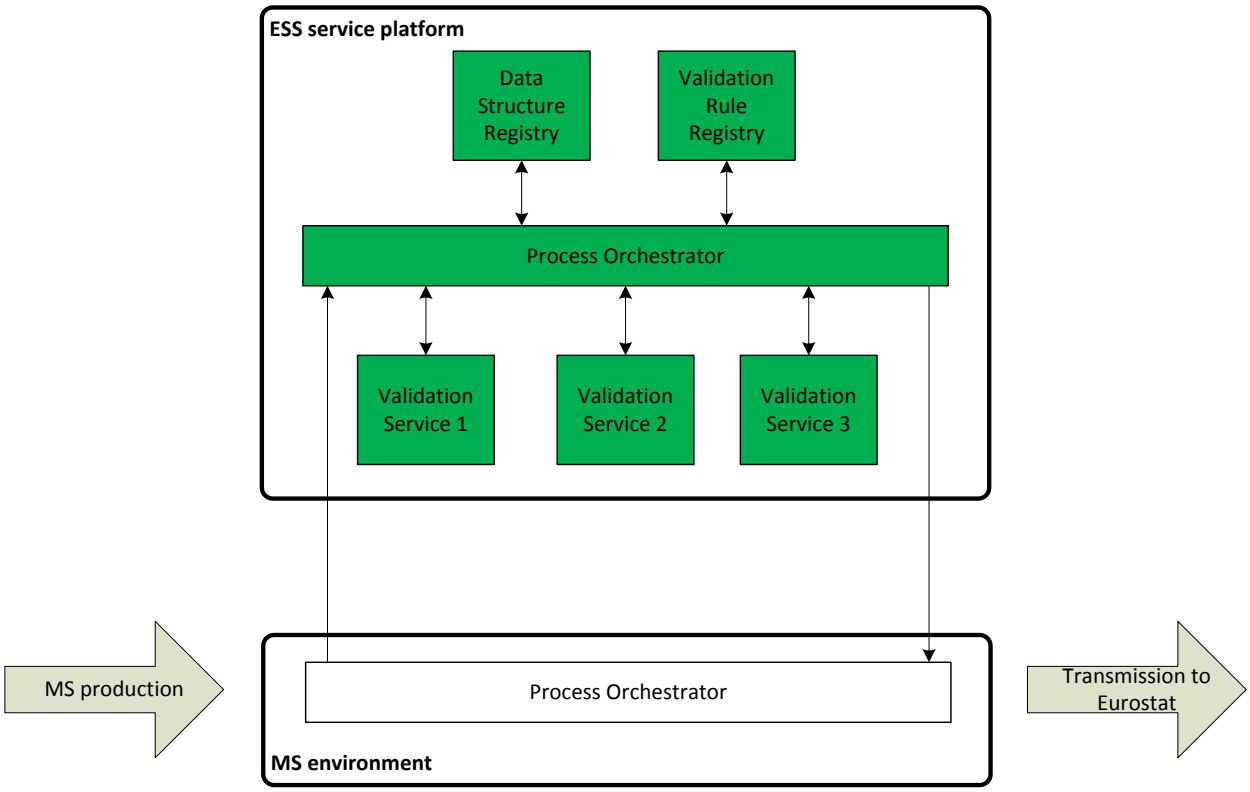
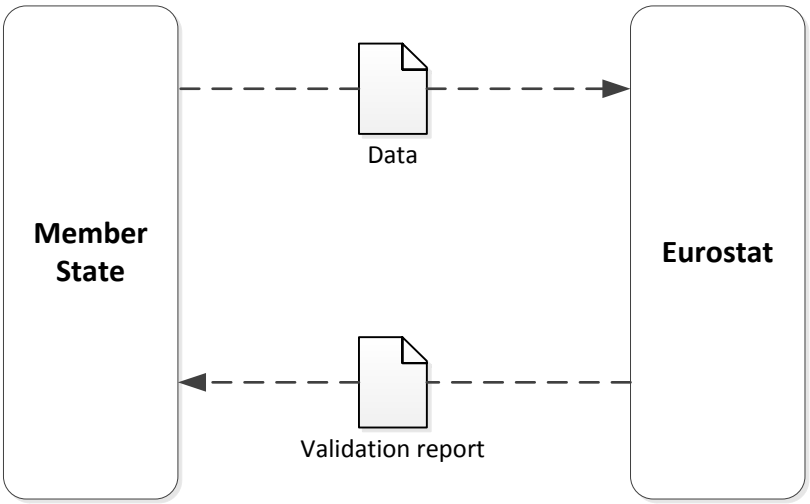


Illustration of the information flows in this scenario



Security concerns might arise in case a Member State wants to use this scenario for confidential data, as confidential data would cross organisational boundaries prior to official transmission to Eurostat. Appropriate arrangements for a secure transmission of data to and from the centrally hosted process will need to be taken in order to enable the use of this scenario in the case of confidential data. Member States should check the compatibility of this scenario with their national legal framework.

Relevant principles

The way the *Validate data* business function has been designed reflects principle 2, *Trust, but verify*. Eurostat always verifies that the validation rules for which Member States are responsible have been correctly applied.

This business function also implements principle 4, *Well-documented and appropriately communicated validation errors*, by ensuring that a standard validation report is sent back to Member States when they transmit data to Eurostat or when they use shareable and re-useable ESS services to perform validation.

6.4 Accept data

Description

The *Validate data* business function has as its main output a validation report outlining which rules are not satisfied by the data. On the basis of this validation report, Eurostat must judge whether the data can be accepted or not. If the data can be accepted depends on the severity level of the rules breached. The severity level of each rule is defined during the design of the rules themselves (see section 6.2) and can have three possible values:

- Error: errors are rules that must necessarily be satisfied for mathematical or logical reasons. Data that do not satisfy these rules will not be considered acceptable, except in exceptional circumstances (see section 6.4 for more details).
- Warning: warnings are rules that, when not satisfied, highlight suspicious values. Data that do not satisfy these rules are not automatically rejected. The data provider must however supply a justification/comment. The data are accepted if Eurostat deems the justification to be sufficient.
- Information: rules whose severity level is marked as "Information" highlight potentially suspicious values. Contrary to the case of warnings, when data do not satisfy these rules, a justification/comment by the data provider is not required for them to be accepted.

In exceptional cases, the data provider or data receiver may agree upon a one-time override of the severity of a given rule. This may be necessary in case the severity level of a given rule is considered to be too lax or too strict due to special circumstances. There are two possible kinds of overrides:

- Member States may request to treat an "error" as a "warning". This would mean that data that do not satisfy the rule in question can be accepted, provided that a suitable justification/comment is given. Eurostat must agree to this override request.
- Eurostat may request to treat an "information" as a "warning". This would mean that a suitable justification/comment is required in order to accept data that do not satisfy the rule in question. The concerned Member State(s) must agree to this override request.

It should be stressed that these overrides should only be used when extraordinary or unusual circumstances justify it. The occurrence of regular override requests for a given rule suggests that the default severity level for the rule in question should be changed. Working Groups should therefore take into account the statistics on the number of override requests when reviewing the severity of validation rules.

Many Eurostat data collections are based on a legal framework which specifies deadlines for the transmission of data. Different domains have established different procedures to judge whether these legal obligations have been respected by data providers: some domains only look at the date of the first transmission, while others take into account quality criteria to judge when the first transmission of valid data occurred. While each domain retains the possibility to set its own quality standards, Eurostat will instruct statistical domains to consider as a minimum quality requirement that data providers must provide data *without errors* before the legal transmission deadline in order to be deemed compliant.

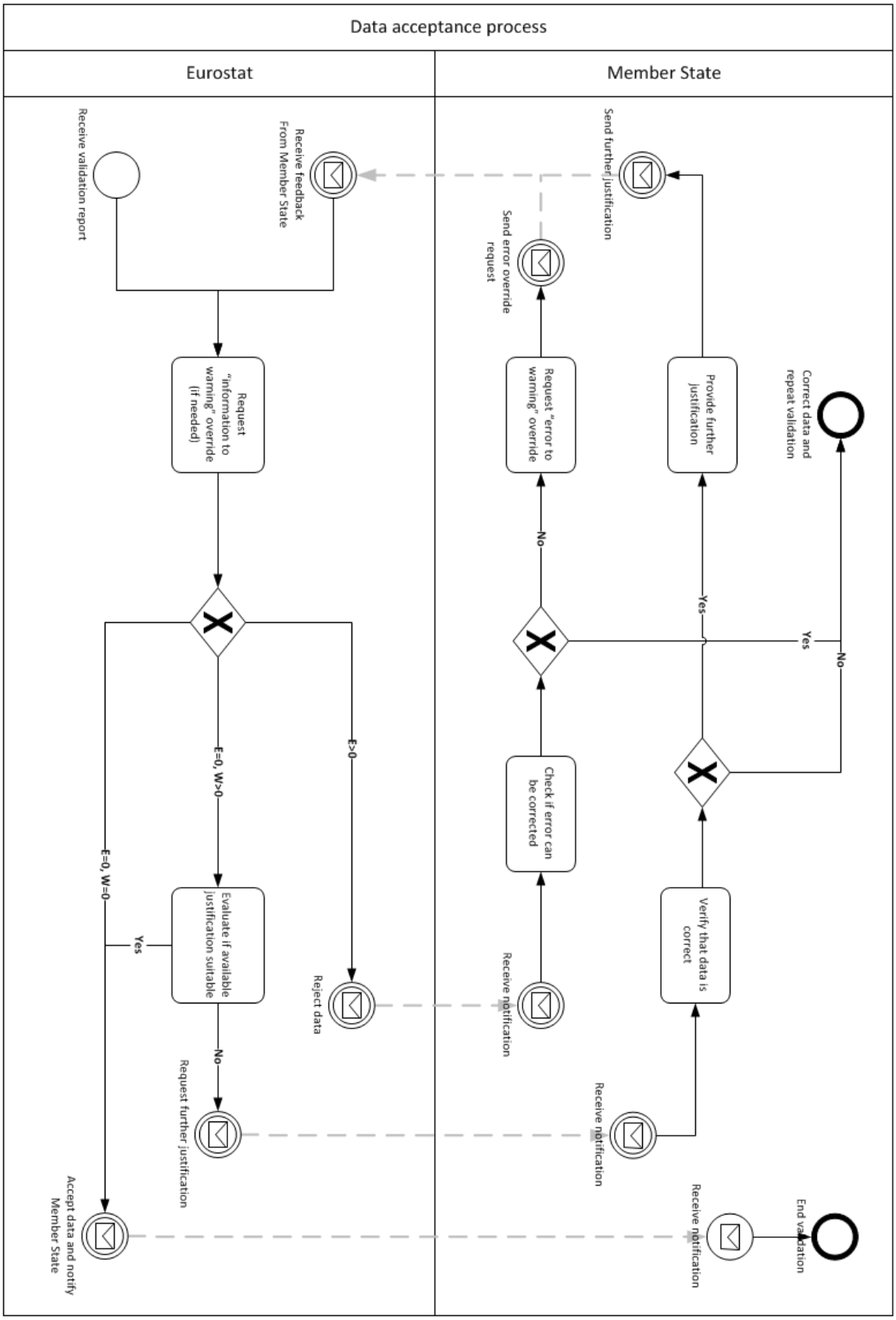
The diagram on page 24 illustrates the acceptance process. It should be noted that, given the importance of justifications/comments by data providers in the data acceptance process, Eurostat will ensure that Member States have the possibility to directly provide such comments alongside the data itself when transmitting the data to Eurostat.

Business Function Name		Accept data
GSBPM reference		Review & validate (5.3) Validate outputs (6.2)
Actor		Eurostat
Supporting IT service	**Description**	N/A
	Collaboration scenario	N/A
	Link to SPRA service	N/A
Input	**Description**	Validation report

	GSIM object	Process Metric
	Standard available	ESS Standard for validation reports
Output	**Description**	N/A
	GSIM object	N/A
	Standard used	N/A

Relevant principles

The *Accept data* business function implements principle 5, *Comply or explain*. Data sent to Eurostat must abide by the agreed upon validation rules. If they do not, Member States should be prepared to provide justification.



Legend:

E = Error

W = Warning

Annex A: Validation principles

PRINCIPLES

1. THE SOONER, THE BETTER

a. Statement

Validation processes must be designed to be able to correct errors as soon as possible, so that data editing can be performed at the stage where the knowledge is available to do this properly and efficiently.

b. Rationale

This principle is at the core of any statistical validation process. There may be many reasons underlying a validation error. Finding the cause and fixing it might well include investigating the correctness of data, software, methodologies or statistical processes as a whole. This can only be done by people sufficiently familiar with the statistical domain and the way the data was produced. Hence, the sooner errors are detected in a statistical production chain, the easier and more efficient it is to correct them.

c. Implications

For the ESS this means that validation of national data should take place at the NSI's who have the sole responsibility for the correctness of the national data. The NSI can only do so if the validation rules are well-defined and understood (see principle 3). If national data appears to be violating validation rules after data exchange, Eurostat should inform the NSI so that correction can be done at the right place. In cases where validation errors arise from rules involving multiple countries, data editing cannot be done by only one NSI. In those cases it is up to Eurostat, being responsible for European figures, to come up with the best possible solution.

2. TRUST, BUT VERIFY

a. Statement

When exchanging data between organisations, data producers should be trusted to have checked the data before exchange and data consumers should verify the data on the common rules agreed.

b. Rationale

Successful data exchange between organisations is a shared responsibility of data producers and data consumers. This cannot be done without a reasonable amount of trust and understanding of each other's duties and challenges. It is the duty of data producers to validate data in the scope of the local perspective before providing it to others. It is the task of the data consumer to validate data in the scope of its broader perspective and provide data producers with useful feedback.

c. Implications

For the ESS this means that member States have a duty to provide Eurostat with data which conform to the validation rules agreed upon. Eurostat, guaranteeing and monitoring the quality of European statistics, has a duty to check that Member States data abide by these same rules and provide them with timely feedback on conformance.

3. WELL-DOCUMENTED AND APPROPRIATELY COMMUNICATED VALIDATION RULES

a. Statement

Validation rules must be clearly and unambiguously defined and documented in order to achieve a common understanding and implementation among the different actors involved.

b. Rationale

This principle seeks (1) to facilitate the development of sound and efficient validation processes, (2) to formalise them and achieve their harmonised implementation and (3) to raise awareness of each participant's role in the validation process.

c. Implications

For the ESS two elements are needed to make this principle operational: a common and easy understandable validation language and an effective communication mechanism. This means that a universal validation language must be chosen by the ESS and that domain specialists (statistical working groups) must agree upon the validation rules for their respective domains.

4. WELL-DOCUMENTED AND APPROPRIATELY COMMUNICATED VALIDATION ERRORS

a. Statement

The error messages related to the validation rules need to be clearly and unambiguously defined and documented, so that they can be communicated appropriately to ensure a common understanding on the result of the validation process.

b. Rationale

This will ensure (1) that errors can be properly corrected, (2) their recurrence is minimised and (3) the risk of false negatives is reduced.

c. Implications

For the ESS this principle requires the definition of a standard ESS validation report structure that is expressive enough to explain the error, its type and severity at a minimum and clear and unambiguous enough to be easily understood by domain and data managers of the NSI's. A streamlined communication process between Eurostat and the NSI's is necessary to make this principle operational.

5. COMPLY OR EXPLAIN

a. Statement

Validation rules must be satisfied or reasonably well explained.

b. Rationale

There may be situations that even earlier agreed validation rules cannot be satisfied. In that case there should be a possibility to escape from them, but only with a well described and understandable explanation that is accepted by the data consumer.

c. Implications

For the ESS this means the validation architecture should provide for a mechanism to explain the exceptional case of non-conformance and to define criteria to decide when an explanation is sufficient. Too strict criteria might become unworkable, too relaxed will not gain the quality improvements necessary. We advise to put together a set of best practices for explanations based on the use of this principle in practice. Repeated occurrences of non-conformance require joint re-evaluation of earlier agreed validation rules.

6. GOOD ENOUGH IS THE NEW PERFECT

a. Statement

Validation rules should be fit-for-purpose: they should balance data consistency and accuracy requirements with timeliness and feasibility constraints.

b. Rationale

It is well known and accepted that perfect data is a myth: errors always exist. The responsibility of the statistician is to manage them so that the final outcome represents a good compromise between all dimensions of data quality.

c. Implications

For the ESS this means that for the design of domain-specific validation rules in the statistical working groups one should look for the right balance between:

- Number of errors to be detected: detecting too many errors risks slowing down the process and makes it inefficient; detecting too few creates the risk that important errors are left undetected.
- Level of severity: rules that are too strict could slow down the process or may lead to a high rate of false positives.
- Level of complexity: rules that are too complex could be source of inconsistencies and therefore of flagging false errors.
- Output orientation versus book-keeping: validation rules should have a clear purpose in the broader context of the statistical ouput to be created.

Annex B: Glossary

Term	Definition	Source
Business Architecture	A description of the structure and interaction between the business strategy, organization, functions, business processes, and information needs.	TOGAF 9.1
Business Function	Something an enterprise does, or needs to do, in order to achieve its objectives.	GSIM
Capability	An ability that an organization, person, or system possesses. Capabilities are typically expressed in general and high-level terms and typically require a combination of organization, people, processes, and technology to achieve.	TOGAF 9.1
Autonomous service	A service that is designed and operated without coordination with other ESS members.	ESS EARF
Interoperable service	A service designed and operated by individual ESS members, but which have the ability to exchange information and operate together effectively.	ESS EARF
Replicated service	A service that is duplicated: ESS members implement an instance of a common service in their local environment.	ESS EARF
Shared service	A service which is common, shared and accessible to all the ESS members. There is a single instance that is shared and available to all.	ESS EARF