Business Architecture
for ESS Validation

European Commission

eurostat   Your key to European statistics

Table of Contents

1.0 Introduction

1.1 Purpose

In January 2013, the ESSC launched the ESS.VIP Validation project in order to start modernising the way data validation is performed in the ESS. This modernisation will require the introduction of changes along different dimensions (e.g. methodology, information standards, IT solutions, business processes etc.). In order to manage the resulting complexity, it is imperative to have a clear overview of the relationships between these different layers and of how they combine to achieve the medium-term goals for validation in the ESS.

This Business Architecture document delivers this overview. 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 a change initiative and of the way it will impact current business processes: it identifies the changes an initiative aims to realize and translates them into a blueprint for concrete implementation. 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 shows how the current situation will need to be changed in order to achieve the target capabilities identified in Chapter 3. It outlines the to-be state for validation in the ESS and analyses the gaps between the current and to-be states.

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 ESS.VIP Validation project and on the changes it aims to produce in the way validation is performed 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 UNECE's Glossary on Statistical data editing, data validation is "an activity aimed at verifying whether the value of a data item comes from the given (finite or infinite) set of acceptable values." 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 involvement of Member States in all stages of implementation should ensure that the solutions outlined in this document can also be used by Member States to modernise validation processes in their national environments.
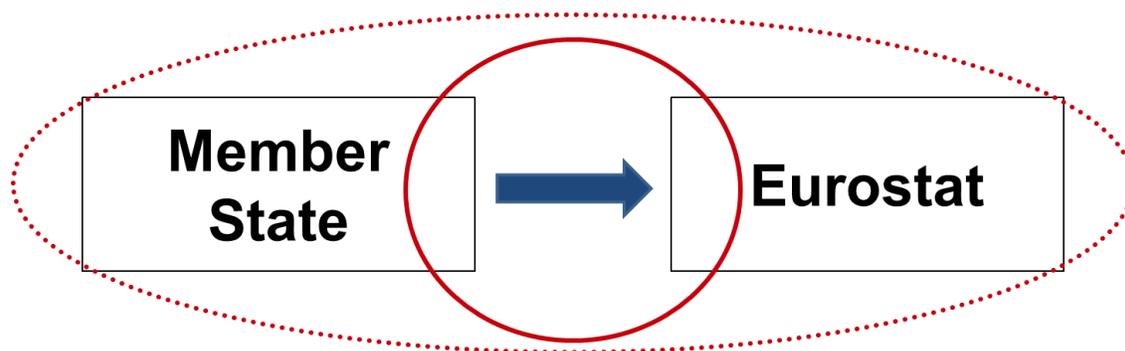
The continuous line represents the scope of this business architecture. Some solutions could however be reused outside this original scope, as represented by the dotted line.

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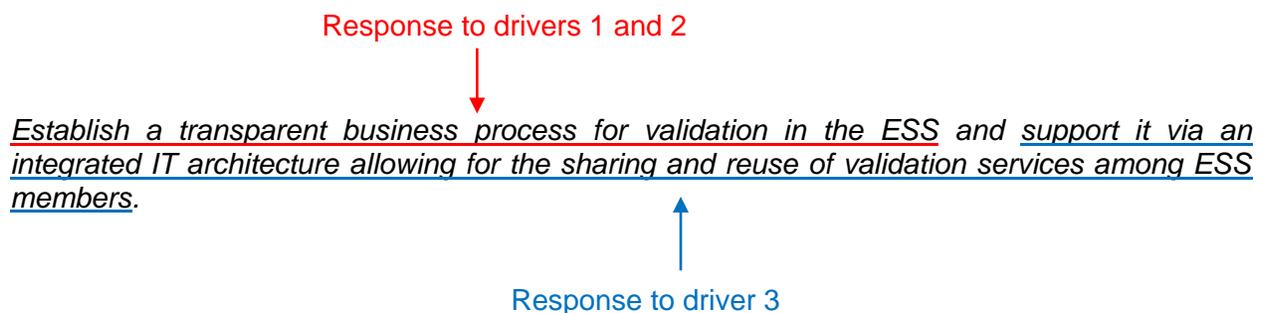
While they vary considerably between different domains, current validation practices exhibit shortcomings which could be corrected through strengthened collaboration. The main such shortcomings are listed below:

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

This Business Architecture document follows an approach based on business capabilities to describe the changes the modernisation of ESS validation will bring about. Business capabilities express what an organisation wants to be able to do in the future. Achieving them requires a combination of five dimensions: human resources, methodology, information standards, IT and processes/governance.

Capability = **HR** + **Methods** + **Processes/ Governance** + **IT** + **Standards**
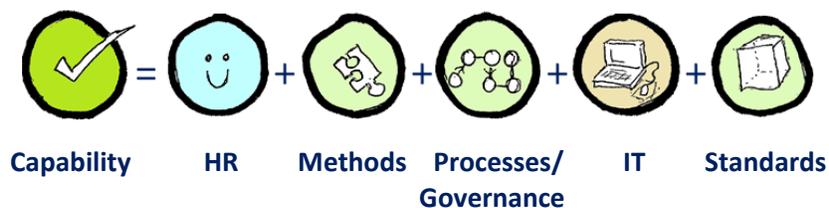
Our approach uses business capabilities as a simple yet powerful conceptual tool to show how the realisation of the business objectives will impact different aspects of an organisation. The first step is to determine the target business capabilities. These can be extracted from the previous section's vision statement. The table below formulates these two, validation-specific capabilities and maps them to the ESS EARF's Business Capability Model in order to correctly position them in the context of the Vision implementation portfolio.

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   From As-is to To-be

The following sections will be dedicated to outlining the as-is and to-be states for ESS validation. The description of the as-is and to-be states will rely on the identification of the different Business Functions involved in the ESS validation process.

According to the definition provided by GSIM, a Business Function is simply "something an enterprise does, or needs to do, in order to achieve its objectives." It is through Business Functions that capabilities are delivered. Business functions will be described in this document by specifying the involved actors, the GSBPM process step they refer to, their inputs and outputs in terms of GSIM information objects and the IT services required to support them.

The concept of *collaboration scenarios* will also be used extensively in this section. Collaboration scenarios describe the expected degree of collaboration among ESS members in the use of IT services. 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.

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. Information objects are also marked differently according to whether or not they are supported by a standard.

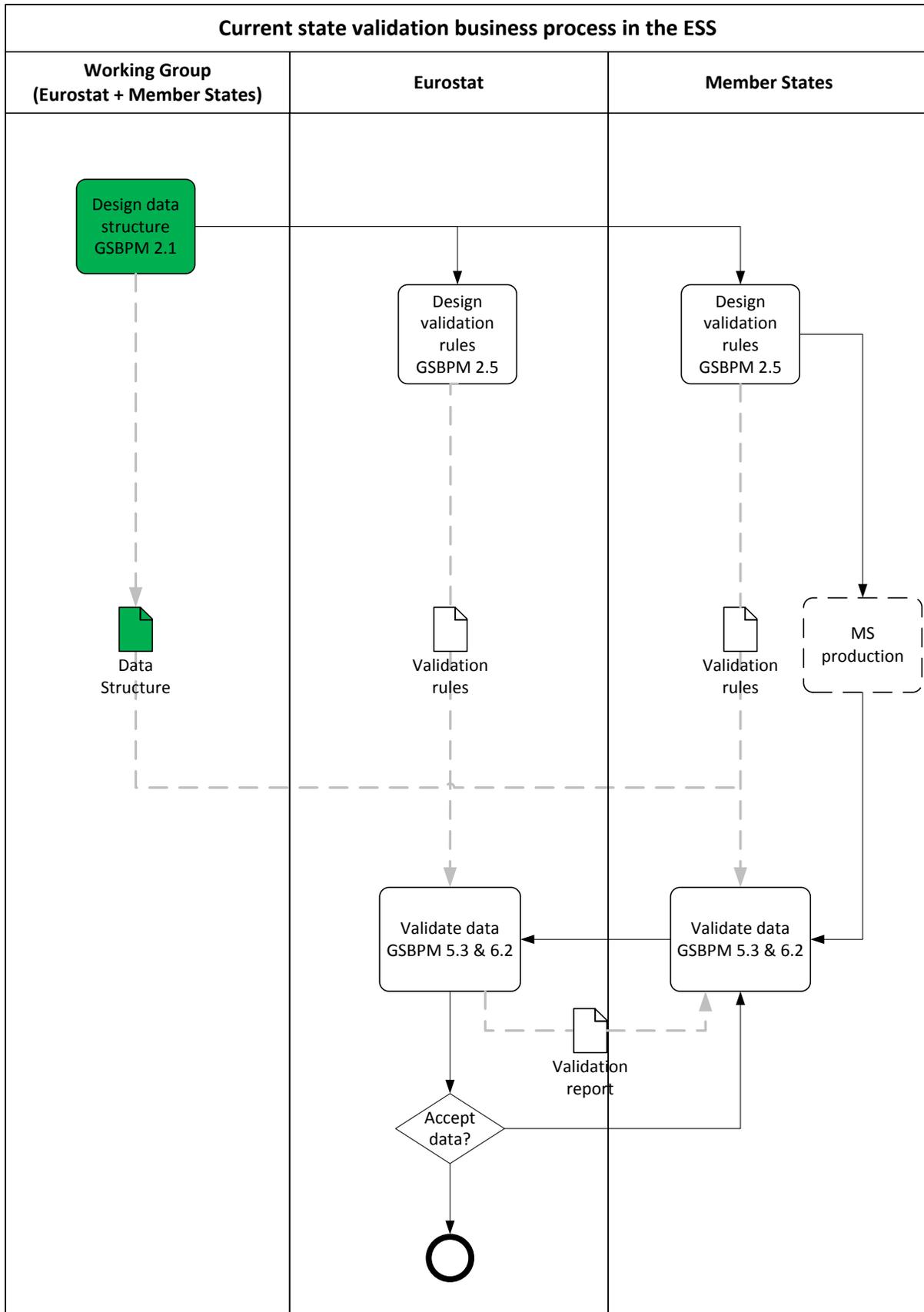|		
Business Function (green)	Business Function supported by shared services	**Information Object** (green) — Information Object supported by a standard
Business Function (light green)	Business Function supported by replicated services	**Information Object** (white) — Information Object not supported by a standard
Business Function (yellow)	Business Function supported by interoperable services	
Business Function (white)	Business Function supported by autonomous services	
Business Function (hatched)	Business Function supported by services in multiple collaboration scenarios	
Business Function (dashed)	Out-of-scope Business Function	

4.1 Current State Business Architecture

As was mentioned in section 2, the process currently in place for the validation of European statistics varies from domain to domain, so much so that it could be said that there is not only one business architecture for validation in the ESS. The current state business architecture presented in this section should therefore be viewed as a sort of archetype, i.e. a typical pattern to which most domains adhere.

The business process diagram in the next page illustrates this archetypal business architecture. In it, Eurostat and Member States define jointly at the Working Group level the structure of the data to be sent. However, the design and execution of the validation rules to be applied to the data occur in parallel with little coordination. Each of the Business Functions mentioned in the business process diagram is described in more detail below.

Design Data Structure

The first step in the current ESS validation process is the definition of the format of the data files to be sent to Eurostat. This Business Function is important for validation as it implicitly provides a first set of validation rules related to the expected structure of the data file. This first step is usually conducted jointly by Eurostat and the Member States through consultations in each specific domain's Working Group. The main output of this Business Function is a document describing the expected *data structure*. In recent years, SDMX has played an increasing role in standardising this step. 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
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	SDMX

Current state validation business process in the ESS

Working Group (Eurostat + Member States)	Eurostat	Member States

Design data structure GSBPM 2.1

Design validation rules GSBPM 2.5 (Eurostat)

Design validation rules GSBPM 2.5 (Member States)

Data Structure

Validation rules (Eurostat)

Validation rules (Member States)

MS production

Validate data GSBPM 5.3 & 6.2 (Eurostat)

Validate data GSBPM 5.3 & 6.2 (Member States)

Validation report

Accept data?

Design Validation Rules

The second Business Function to come into play is the definition of the rules to be applied to validate the content of the data to be sent to Eurostat. Only in a fraction of domains is this step performed jointly by Member States and Eurostat at the Working Group level: in most cases, Member States and Eurostat define the validation rules independently and store them in their own documentation systems. The main outputs of this Business Function are the validation rules themselves. At ESS level, there is currently no standard syntax for validation rules and no shared services to support this Business Function are available.

Business Function Name		Design validation rules
GSBPM reference		Design processing & analysis (2.5)
Actor		Eurostat Member States
Supporting IT service	Description	Validation rule documentation system
	Collaboration scenario	Autonomous
Input	Description	N/A
	GSIM object	N/A
	Standard available	N/A
Output	Description	Validation rules
	GSIM object	Rule
	Standard used	N/A

Member State production

After the design of the data structure and of the validation rules is completed, Member States produce the required datasets. As already explained in chapter 2, Member State-internal production processes are considered to be out of the scope of this business architecture. This business function will therefore not be described in more detail.

Validate Data

The data sent to Eurostat are validated by both Member States and Eurostat using the previously defined data structures and validation rules. Some IT services are offered by Eurostat to support this function in Member States (e.g. SDMX Converter, EDIT, Webforms). These are usually deployed as either replicated or shared services. A survey conducted by the ValiDat Foundation ESSnet revealed however that in most statistical production processes Member States use their own tools. The main output of data validation is a validation report. Currently, no standardised validation report structure exists in the ESS.

Business Function Name		Validate Data
GSBPM reference		Review & validate (5.3)
		Validate outputs (6.2)
Actor		Eurostat
		Member States
Supporting IT service	Description	Validation services executing the previously defined validation rules
	Collaboration scenario	Autonomous
Input 1	Description	Structure of the datasets to be sent to Eurostat
	GSIM object	Data Structure
	Standard available	SDMX
Input 2	Description	Validation rules
	GSIM object	Rule
	Standard available	N/A
Output	Description	Validation report
	GSIM object	Process Metric
	Standard used	N/A

4.2   Target State Business Architecture

Starting from the target capabilities listed in chapter 3, this section will highlight the main changes that need to be undertaken in order to realise them. The outcome of this analysis will be an outline of the target to-be state architecture.

Capability 1: Transparency of the validation procedures

In the as-is business architecture, the main obstacle to the transparency of the validation procedures applied to the data sent by Member States to Eurostat is the lack of coordination on the design of the validation rules. As the previous section showed, Member States and Eurostat often perform this step in parallel and do not communicate the respective validation rules to each other.

In order to guarantee transparency, the "Design Validation Rules" Business Function must therefore evolve from its current condition. Working Groups should take over the responsibility to design validation rules and to assign validation responsibilities. 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. The table below summarises the changes needed.

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
Input	Description	N/A
	GSIM object	N/A
	Standard available	N/A
Output	Description	Validation rules
	GSIM object	Rule
	Standard used	ESS Standard for validation rules

Capability 2: Sharing and reuse of validation services

The wider sharing and reuse of validation services in the ESS is currently hindered by three major factors:
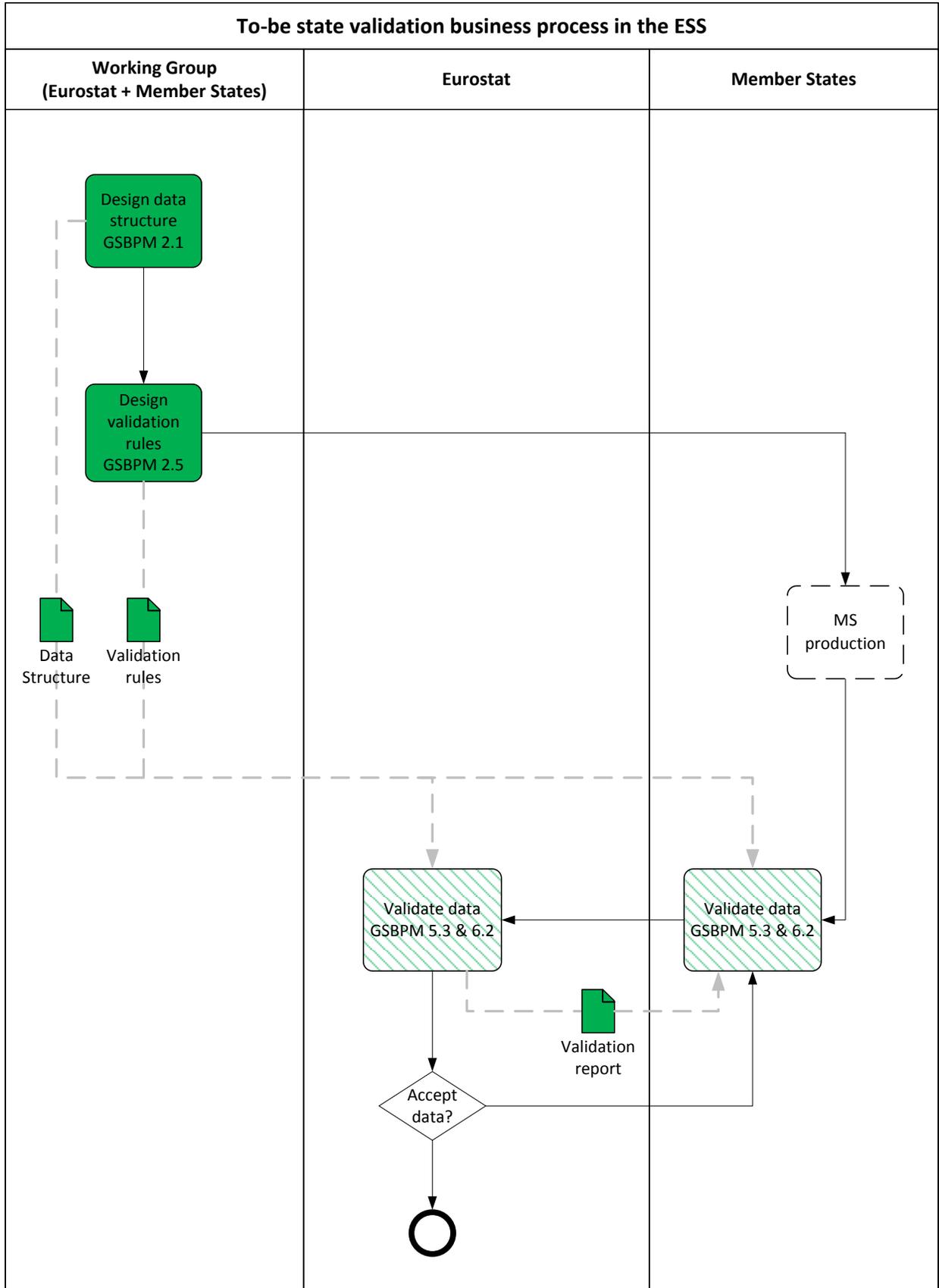
- The first one is the lack of ESS standards for the main inputs and outputs of the "Validate Data" Business Function. In the target to-be state, validation services should be based on a standard description of validation rules and of validation reports.

- The second one is the lack of agreement on the correct granularity and scope of validation services. In the target to-be state the validation process should be supported by a portfolio of reusable ESS validation services, each responsible for a specific subset of validation (e.g. structural validation, outlier detection, etc…).

- The third one is the absence of a common IT infrastructure in the ESS to support sharing and reuse. In the target to be state, an ESS infrastructure to enable a plug-and-play approach to validation should be set up.

Taking into account all the needs outlines above, the "Validate Data" Business Function for the target to-be state can be summarised as follows.

Business Function Name	Validate Data
GSBPM reference	Review & validate (5.3) Validate outputs (6.2)

|		Eurostat
Actor		Member States
Supporting IT service	**Description**	Granular validation services executing the previously defined validation rules
	Collaboration scenario	Shared/Replicated/Interoperable/Autonomous
Input 1	**Description**	Structure of the datasets to be sent to Eurostat
	GSIM object	Data Structure
	Standard available	SDMX
Input 2	**Description**	Validation rules
	GSIM object	Rule
	Standard available	ESS Standard for validation rules
Output	**Description**	Validation report
	GSIM object	Process Metric
	Standard used	ESS Standard for validation reports

The target to-be state resulting from the two changes outlined above is illustrated by the business process diagram on the next page.

To-be state validation business process in the ESS

Working Group (Eurostat + Member States)	Eurostat	Member States

Design data structure
GSBPM 2.1

Design validation rules
GSBPM 2.5

Data Structure

Validation rules

MS production

Validate data
GSBPM 5.3 & 6.2

Validate data
GSBPM 5.3 & 6.2

Validation report

Accept data?

13

As long as the jointly agreed upon validation rules are applied, each Member States should be able to freely choose, for each statistical production process, the extent to which it wants to benefit from the availability of reusable ESS validation services. We can thus delineate three basic scenarios for how Member States could execute validation rules in the target to-be state.

Scenario 1: Autonomous validation services

Under this scenario, Member States use their own autonomous services to execute the validation rules prior to data transmission to Eurostat. These autonomous services would however use the jointly agreed upon data structures and validation rules, which would be stored in centrally hosted registries. The translation of these validation rules into the autonomous validation services would be the responsibility of each individual Member State.
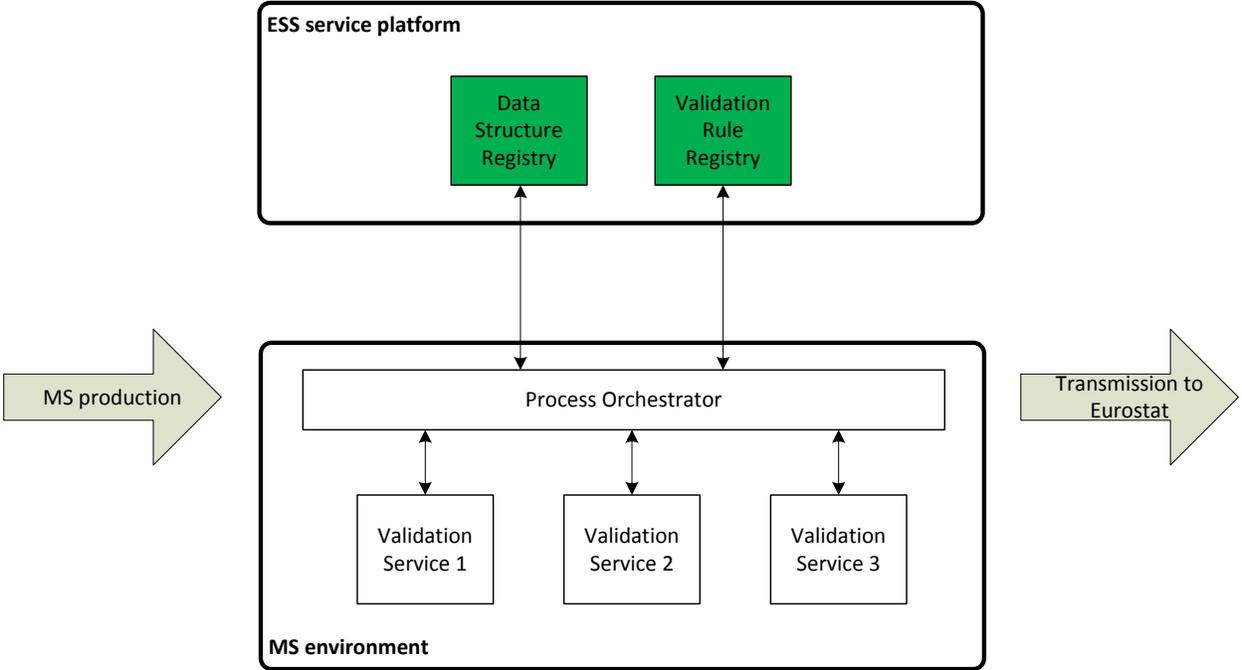
Scenario 2: Replicated/Shared validation services

Under this scenario, in addition to the shared registries for data structures and validation rules, Member States use some replicated and/or shared validation services in their validation process. Member States would be free to select the combination of autonomous/ interoperable/ replicated and shared services they find most suitable. 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.
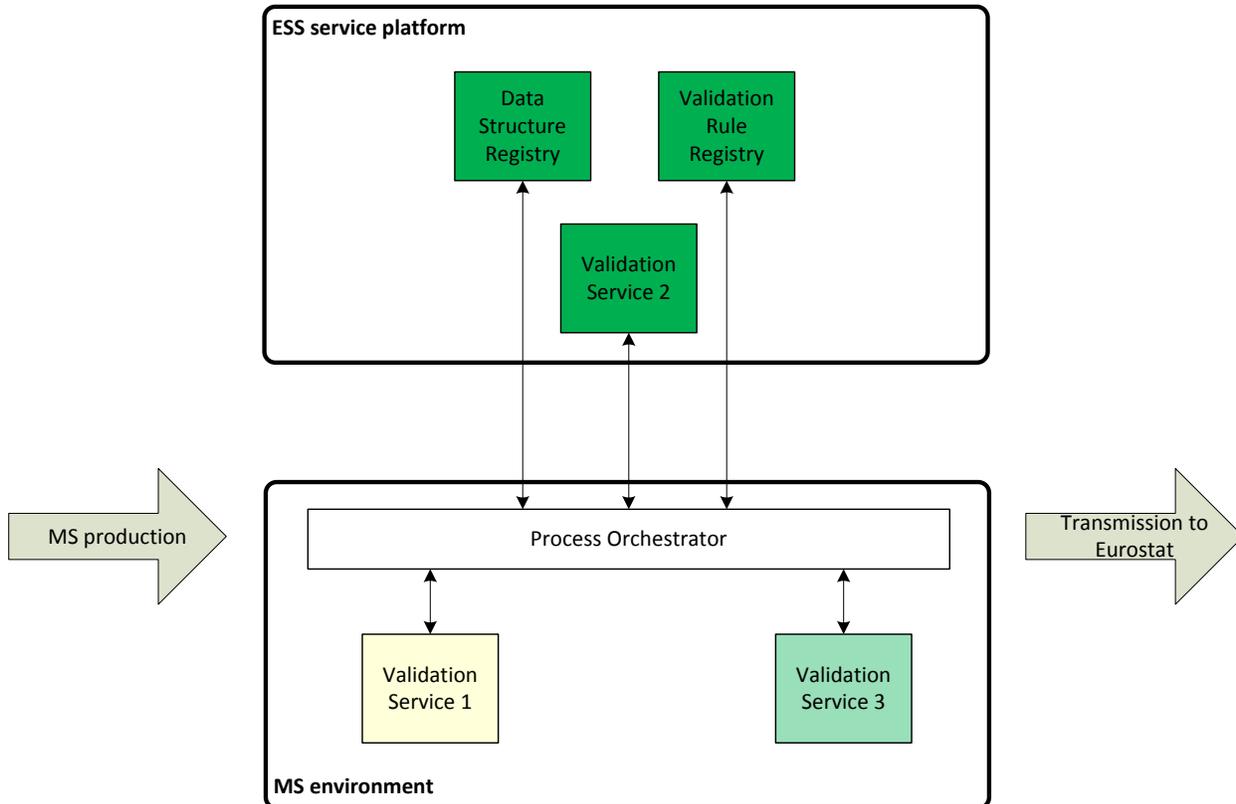
Scenario 3: Shared validation process

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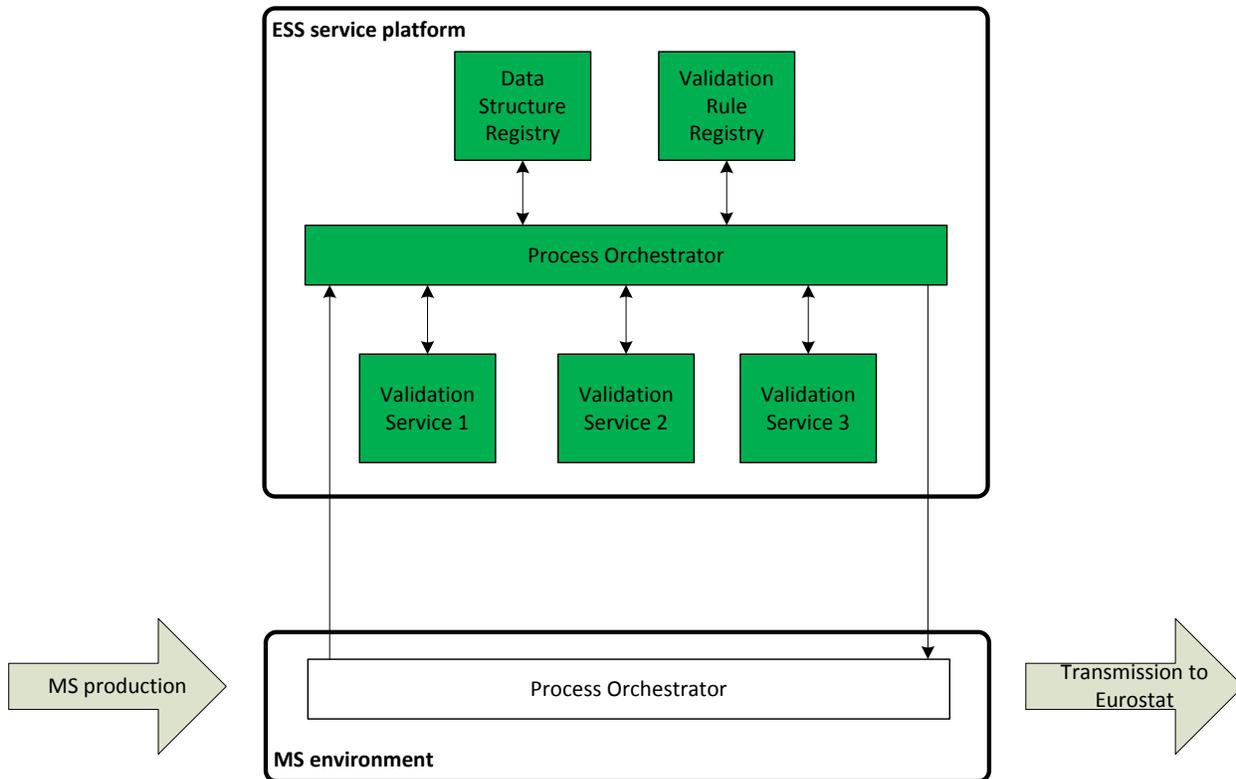
The scenarios illustrated above represent somewhat idealised situations. In real-life situations, it is probable that Member States will create hybrid scenarios which will incorporate elements of two or more scenarios. Each Member State would be free to mix and match the three scenarios as it sees fit.

4.3   Gap analysis

In this section, the analysis will concentrate on the identification of the capability gaps between the current state and the target state, as described in the previous paragraphs. The analysis will also determine which elements are needed to fill the identified gaps.

Capability 1: Transparency of the validation procedures

The changes required to achieve this capability concern the "Design validation rules" Business Function. Below is a side-by-side comparison of this Business Function in the current and to-be states, with the main foreseen changes written in red.

			Current	To-be
Business Function Name			Design validation rules	Design validation rules
GSBPM reference			Design processing & analysis (2.5)	Design processing & analysis (2.5)
Actor			Eurostat Member States	Working Group
Supporting IT service		**Description**	Validation rule documentation system	Registry enabling the creation, management and retrieval of validation rules
		Collaboration scenario	Autonomous	Shared
Input		**Description**	N/A	N/A
		GSIM object	N/A	N/A
		Standard available	N/A	N/A
Output		**Description**	Validation rules	Validation rules
		GSIM object	Rule	Rule
		Standard used	N/A	ESS Standard for validation rules

Realising the changes highlighted in the comparison table above implies improvements along all five of the capability dimensions introduced in Chapter 3. The table below contains the main deliverables which will need to be provided to adequately cover each dimension.

Capability	Dimension	Needed deliverables
Capability 1	HR	Training and communication on new role of Working Groups and on new standard language for validation rules
	Methodology	Methodological handbook providing common definitions for concepts important for validation (validation rules, levels, metrics) Working Group templates to be used by Working Groups to define validation rules
	Processes/ Governance	Common ESS validation policy focusing on the attribution of validation responsibilities
	IT	Validation Rule Registry and GUI
	Standards	Standard language for validation rules

Capability 2: Sharing and reuse of validation services

The same procedure used to highlight the gaps for Capability 1 can be applied to Capability 2. Here the only concerned Business Function is "Validate Data". The comparison table in this case is the following:

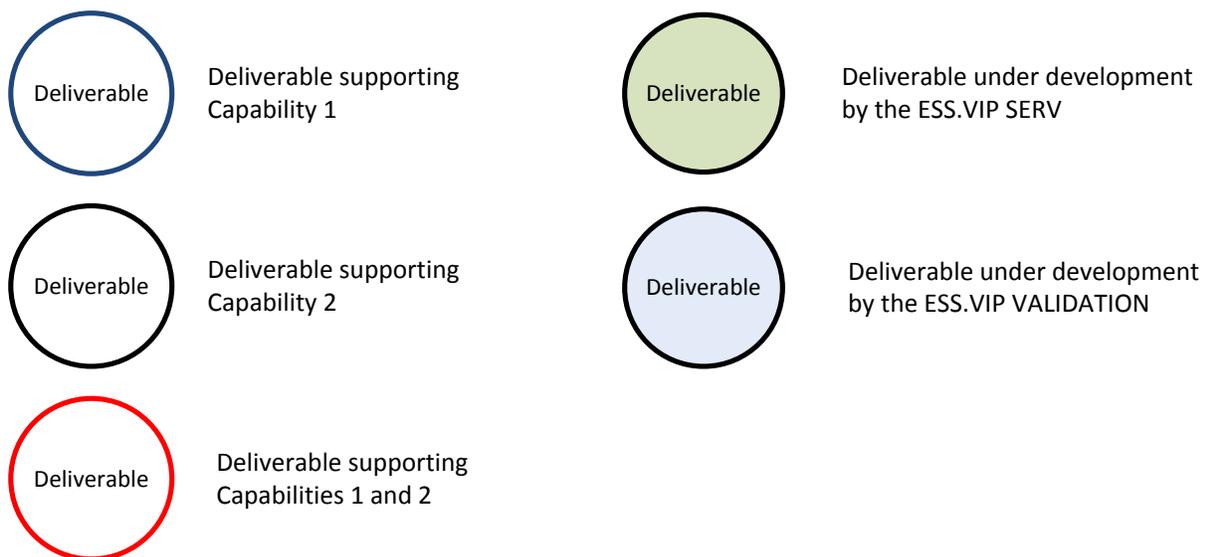|		Current	To-be
Business Function Name		Validate Data	Validate Data
GSBPM reference		Review & validate (5.3) Validate outputs (6.2)	Review & validate (5.3) Validate outputs (6.2)
Actor		Eurostat Member States	Eurostat Member States
Supporting IT service	**Description**	Validation services executing the previously defined validation rules	Granular validation services executing the previously defined validation rules
	Collaboration scenario	Autonomous	Shared/Replicated/Interoperable/Autonomous
Input 1	**Description**	Structure of the datasets to be sent to Eurostat	Structure of the datasets to be sent to Eurostat
	GSIM object	Data Structure	Data Structure
	Standard available	SDMX	SDMX
Input 2	**Description**	Validation rules	Validation rules
	GSIM object	Rule	Rule
	Standard available	N/A	ESS Standard for validation rules
Output	**Description**	Validation report	Validation report
	GSIM object	Process Metric	Process Metric
	Standard used	N/A	ESS Standard for validation reports

The deliverables needed to bridge the identified gaps can once again be classified according to the five dimensions of Capability 2, as in the table below.
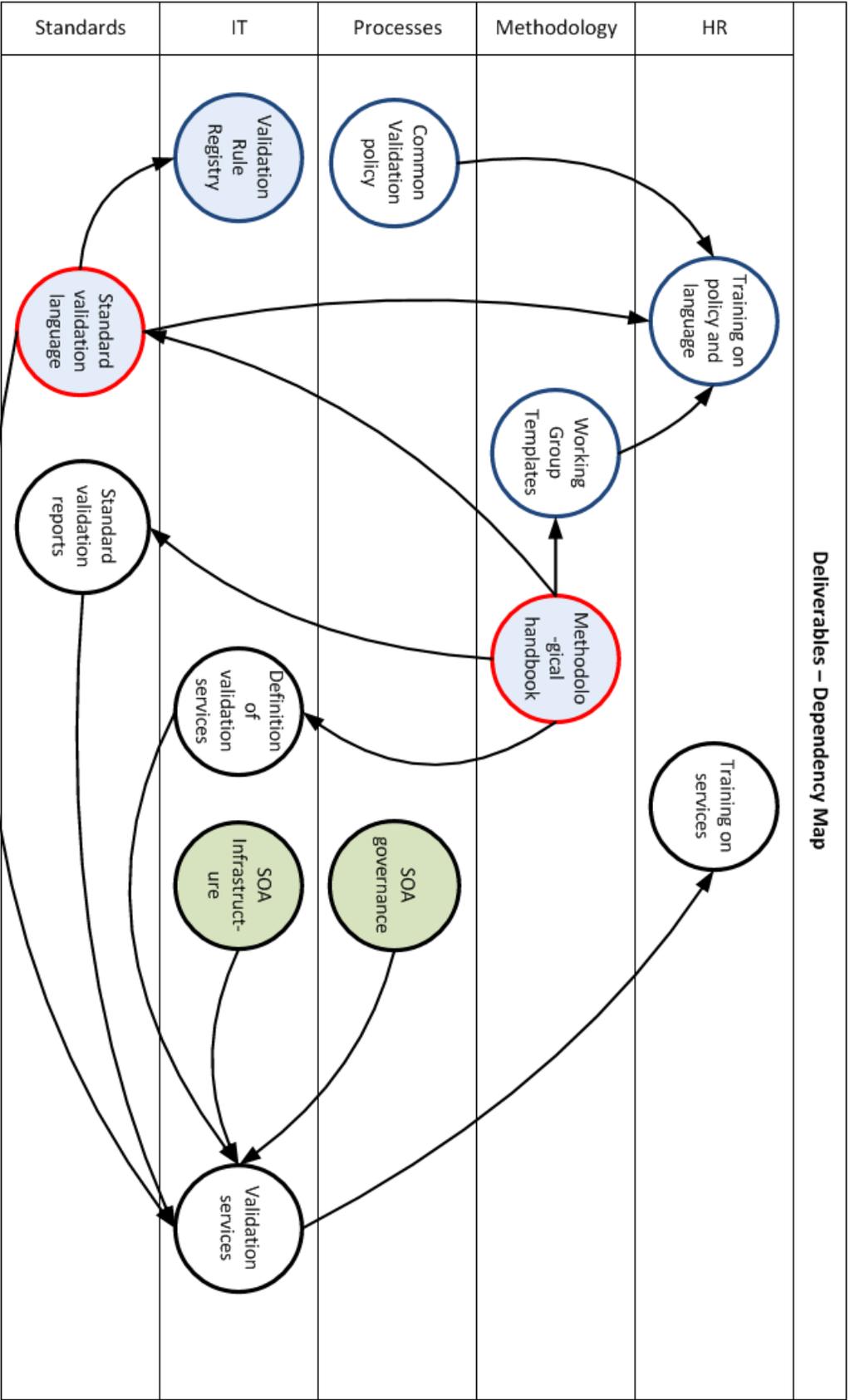
Capability	Dimension	Needed deliverables
Capability 2	HR	Training and communication on validation architecture and validation services
	Methodology	Methodological handbook providing common definitions for concepts important for validation (validation rules, levels, metrics)
	Processes/ Governance	SOA Governance
	IT	SOA Principles and Infrastructure Definition of needed validation services conforming to the architecture Development of needed validation services
	Standards	Standard language for validation rules Standard format for validation reports

4.4   Roadmap

In the preceding section, the main deliverables needed to achieve the two target capabilities for ESS Validation were listed. While some deliverables contribute to a single capability, others contribute to both.

The preceding analysis did not highlight the dependencies between the different deliverables. The dependency map on the following page illustrates these dependencies and shows which deliverables are currently under development. While a precise roadmap for all the deliverables has not yet been established, the dependency map shows the sequence that will need to be followed to fully achieve the target capabilities.

Deliverable — Deliverable supporting Capability 1

Deliverable — Deliverable under development by the ESS.VIP SERV

Deliverable — Deliverable supporting Capability 2

Deliverable — Deliverable under development by the ESS.VIP VALIDATION

Deliverable — Deliverable supporting Capabilities 1 and 2

Deliverables – Dependency Map

Standards	IT	Processes	Methodology	HR

- Validation Rule Registry
- Common Validation policy
- Standard validation language
- Training on policy and language
- Standard validation reports
- Working Group Templates
- Definition of validation services
- Methodolo-gical handbook
- SOA Infrastruct-ure
- SOA governance
- Training on services
- Validation services

5.0 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	Services that are are designed and operated without coordination with other ESS members.	ESS EARF
Interoperable service	Services designed and operated by individual ESS members, but which have the ability to exchange information and operate together effectively.	ESS EARF
Replicated service	Services that are duplicated: ESS members implement an instance of a common service in their local environment.	ESS EARF
Shared service	Services which are common, shared and accessible to all the ESS members. There is a single instance that is shared and available to all.	ESS EARF