

ESS EA Statistical Production Reference Architecture

29 October 2017

Table of Contents

[1.0 Introduction 2](#_Toc525054982)

[1.1 GSBPM phases covered 3](#_Toc525054983)

[2.0 Use cases 3](#_Toc525054984)

[3.0 Key concepts in the SPRA 5](#_Toc525054985)

[3.1 Service Oriented Architecture 5](#_Toc525054986)

[3.2 Application functions, application services and application components 5](#_Toc525054987)

[3.3 Consolidation approaches and scenarios 6](#_Toc525054988)

[3.4 Service orchestration 12](#_Toc525054989)

[3.5 Statistical Data Management 12](#_Toc525054990)

[3.6 Metadata Management 14](#_Toc525054991)

[4.0 Building blocks of the SPRA 15](#_Toc525054992)

[4.1 Description 15](#_Toc525054993)

[4.2 User story illustrating statistical production and the role of the building blocks 20](#_Toc525054994)

[4.3 Actual use and templates 21](#_Toc525054995)

[5.0 Services in the SPRA: Data Collection 22](#_Toc525054996)

[5.1 Context 22](#_Toc525054997)

[5.2 ESS relevant services 22](#_Toc525054998)

[5.3 Design Principles 25](#_Toc525054999)

[6.0 Services in the SPRA: Process & Analyse 26](#_Toc525055000)

[6.1 Context 26](#_Toc525055001)

[6.2 ESS relevant service 26](#_Toc525055002)

[6.3 Design Principles 30](#_Toc525055003)

[7.0 Services in the SPRA: Dissemination 32](#_Toc525055004)

[7.1 Context 32](#_Toc525055005)

[7.2 ESS relevant services 33](#_Toc525055006)

[7.3 Design Principles 35](#_Toc525055007)

[Attachments 38](#_Toc525055008)

[Implementations of Building Blocks 38](#_Toc525055009)

[Illustration of consolidation approaches 40](#_Toc525055010)

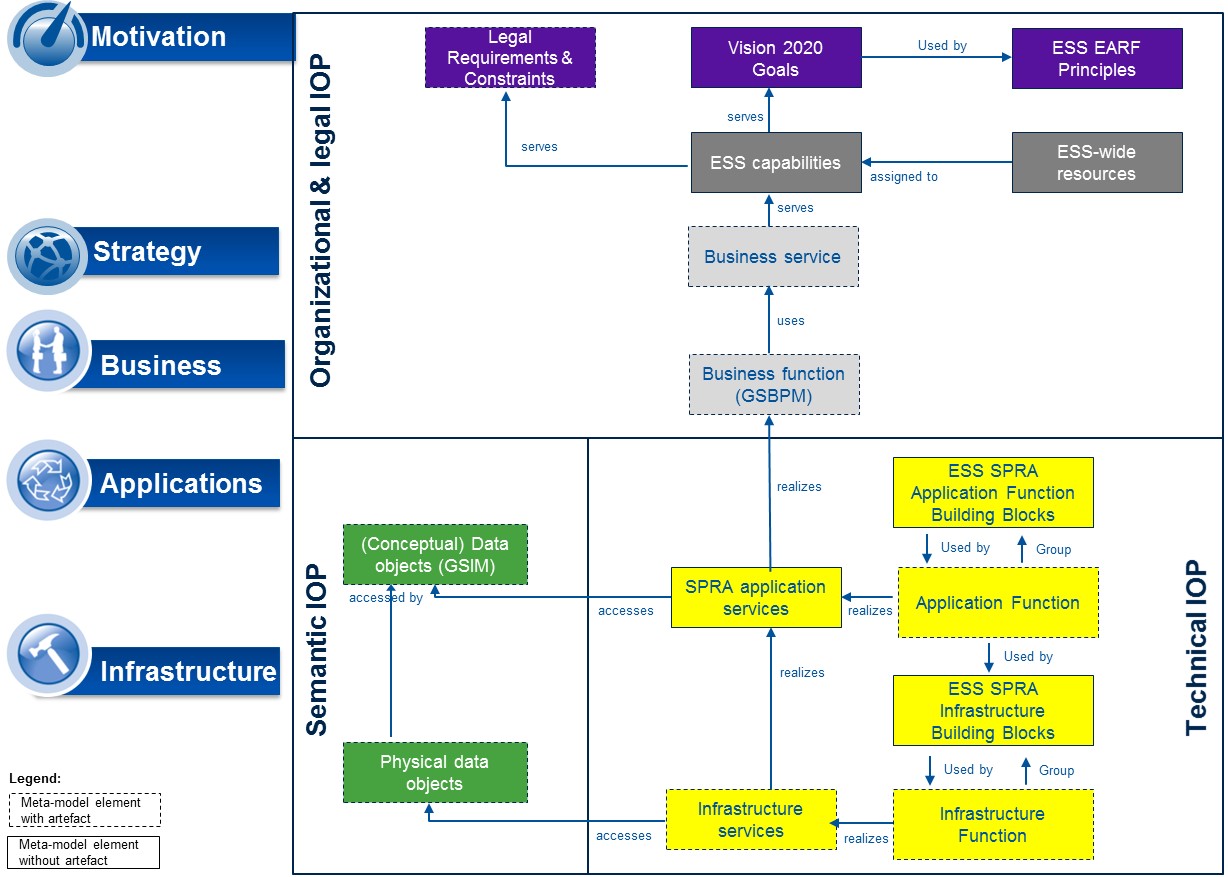
[Consolidation approaches and their impact on interoperability requirements 43](#_Toc525055011)

1. Introduction

The ESS Statistical Production Reference Architecture (SPRA) provides a modelling of the **to-be state architecture of ESS information systems** based on various key artefacts of the ESS EA Reference Framework.

The SPRA uses the guidance from the ESS EARF to model aspects of statistical production believed to be crucial for realizing the ESS modernisation and to facilitate the sharing of solutions and services within the ESS.

The role of the SPRA in the EARF is depicted in the figure below.



Scope of the SPRA

1. The SPRA and the ESS EARF Metamodel

The SPRA consists of a number of views modelled using the Arechimate language in Archi, and is available in the distribution of the ESS EARF Archi model.

The SPRA should be used to support the modernization of statistical production and to facilitate sharing and reuse. It has taken guidance from a number of sources, indlucing the CSPA work undertaken by the UNECE as well as the EIRA (European Interoperability Reference Archictecture) develop by the European Commission.

The current document provides a description of the architecture models and elements and explains how the SPRA can be put to use.

In particular, the SPRA models:

* How ESS members can collaborate to solve aspects of statistical production using three different modes. Either through **shared** services, **replicated** solutions or services or **coordinated** solutions and services.
* Key services of importance to statistical production, which are candidate units of either **sharing** services or **replicating** or **coordinating** underlying software solutions.
* A set of key Building blocks, which form bundles of functionality supporting statistical production. Building blocks are seen as homogeneous sets of of functionalities at the level of applications or infrastructure candidate for reuse.
  1. GSBPM phases covered

The current version of the SPRA provides the architecture models and principles for those aspects of the ESS EARF which relate to the core GSBPM phases of **Collect, Process & Analyse, and Disseminate**. In addition, the document also addresses Design or Build phases where these inherently relate to core ones. In future versions, additional GSBPM phases may be covered, for example those associated with planning and design to obtain an even more complete, end-to-end perspective on the statistical production process.

1. Use cases

The primary audience of this document are project managers and architects engaged in developing Information System Architectures with the objective of modernising one or more phases of the statistical production chain and building the required IT systems to support it.

The current SPRA provides the following viewpoints:

* It models a number of Application **Services** needed to support the sub-processes in the GSBPM model
* It defines a number of **principles** that are **specific to the GSBPM phase** modeled
* It describes possible consolidation **scenarios** for how the architecture could be realized by the ESS members

The architecture described in this document has been implemented in respect of NSI specific systems, investment programmes and the principle of **subsidiarity**. Therefore, the SPRA:

* Enables a gradual or selective (“opt-in, opt-out”) adoption of (selected components of) the to-be state architecture by the ESS members;
* Enables gradual investments, broken down into value-creating elements;
* Enables ESS members to utilize investments by providing different scenarios for how individual NSIs will implement the target architecture enabling for better interoperability;
* Maximizes the sharing of investments in IT solutions and still makes it sufficiently flexible to suit different investment cycles with ESS members;

To ensure the architecture corresponds to what is considered state-of-the-art in terms of a flexible, sustainable, cost- efficient and effective architectures, a number of **requirements** have been formulated to the architecture itself. The next table summarizes these drivers influencing the SPRA and highlights their impact on architecture design.[[1]](#footnote-2)

|  |  |
| --- | --- |
| Driver | Impact on SPRA |
| *Need to re-use Services in the ESS* | * The SPRA must enable a Service-based model where Services can be provided and/or called by different ESS members; this regardless of who is providing the Services (central provisioning through DG ESTAT, Centre of Excellence-based provisioning, provisioning by a single Member State, etc.). * The idea of re-use of Services requires the architecture to implement SOA thinking, outlining Services with different functionalities to be made available in various statistical production processes through a generalized Service interface. |
| *Need to reuse methods and algorithms for statistical processing in the ESS* | * The SPRA must support a way of sharing methods and algorithm for statistical processing. This may be by implementing services or exchange software. |
| *Need to re-use data in the ESS* | * The SPRA must enable efficient & effective re-use of data through consolidating the management and storage of data across statistical domains and possibly also Member States where relevant. |
| *Need to re-use Metadata in the ESS* | * The SPRA must support a gradual transition towards unified management of Metadata across the life cycle of statistics by integrating the Metadata through an ESS-wide (logically interrelated) Metadata Management system. * The SPRA shall also help realize the systematic use of active Metadata in statistical production. |
| *Need to orchestrate processes across the ESS* | * The SPRA must illustrate how various (new) business processes will be orchestrated to form a coherent statistical production chain across statistical domains and possibly also Member States where relevant. * Introducing concepts of process management is vital to supporting a greater industrialization of statistical production supported by IT and separating design from statistical production. |
| *Need to encourage conformance with standards* | * The SPRA must encourage conformance with (business process, technical, semantic) standards to enable more effective collaboration and data exchanges across ESS members. |
| *Need to embrace large amounts of (new types of) data* | * The SPRA must reflect the envisioned growth in volume, velocity and variety of data sources with which ESS partners will be confronted and on which they seek leveraging upon. |

1. ESS Statistical Production Reference Architecture: Requirements to the architecture
2. Key concepts in the SPRA

The SPRA consists of the current document, as well as a number of concepts and views modelled in the ESS EARF Archi model. The Archi model contains the following concepts:

* All phases and sub-phases of the GSBPM
* All services described in this document
* All building blocks defined in this document

These concepts can be used as a library for more specific modelling and specialisation but are also used to model a number of views of the SPRA:

* SPRA – GSBPM: mapping the defined services to the GSBPM sub-phases.
* ESS SPRA Building Blocks: providing an overview of the proposed building blocks.
  1. Service Oriented Architecture

The SPRA introduces Services in the sense of a Service-Oriented Architecture (SOA). SOA is an approach applicable to both business & IT.

* In the business sense, SOA is about focusing on which Services the business requires in order to fulfil its mission, and how the organization of Services can be improved.
* In the IT sense, SOA is about the integration of various technologies through a common Services framework, standards and protocols.

SOA permits to extract functionalities of existing systems (whether legacy or not) and to expose them as Services. Of course also completely new Services can be created, either relying on existing IT solutions, or by newly created solutions, or by a mixture of both.

The identification of Services in the SPRA has been achieved by:

* Examining the activities of the GSBPM phases
* Analysing ESS-specific literature and documentation relevant to the GSBPM phase
  1. Application functions, application services and application components

The SPRA models what Archimate refers to as “application services”.[[2]](#footnote-3)

An **application service** is defined as a service that exposes automated behavior.

An application service exposes the functionality of components to their environment. This functionality is accessed through one or more application interfaces. An application service is realized by one or more application functions that are performed by the component. It may require, use, and produce data objects.

An application service should be meaningful from the point of view of the environment; it should provide a unit of functionality that is, in itself, useful to its users. It has a purpose, which states this utility to the environment. This means, for example, that if this environment includes business processes, application services should have business relevance.

A purpose may be associated with an application service. An application service may be used by business processes, business functions, business interactions, or application functions. An application function may realize an application service. An application interface may be assigned to an application service. An application service may access data objects. The name of an application service should preferably be a verb ending with ­“‑ing”; e.g., “transaction processing”. Also, a name explicitly containing the word “service” may be used.

* 1. Consolidation approaches and scenarios

In order to enhance re-use, transparency and interoperabiliy, consolidation scenarios should be jointly modelled and shared across Member States. Architecture detail and breadth depends on a scenario as they differ from an interoperability perspective.

Interoperability requires agreements on and definition of reference architectures to be adopted. Conceptually and concretely, ESS Members’ architectures should specialize the reference architectures into architectures specific to their organisations. As an example, an ESBRs reference architecture should be used and should be specified and implemented in each ESS member , with differing solution architecture catering to specific needs and conditions within the member. The link between reference architecture and implementation details (specialization) should always be explicit (modelled), so that any changes to reference architectures could be easier understood.

The ESS SPRA defines four possible approaches for implementing functionality (application services in ArchiMate) with respect to the aspect of sharing and reuse:

|  |  |  |
| --- | --- | --- |
| Consolidation approach | Short description | Existing Examples |
| Autonomous | The application service is implemented as an individual solution of the ESS partner without coordination with others. Optimization takes places at the level of each ESS partner. | Financial management or Human Resource Management (may be shared at the national level but not within the ESS) |
| Coordinated | The application service is implemented as an individual software solution of the ESS partner with a coordinated exchange of data through agreed standards. | ESBRS |
| Replicated | The application service is implemented as shared software, which supports the execution of similar business processes through replication of software (possibly with differing configurations). | EDIT SDMX CONVERTER SDMX-RI WEB SERVICE The exchange of SAS or R scripts |
| Shared | The application service is implemented as a shared service, which executes identical business processes through a single solution instance. | EDAMIS |

The four approaches carry with them different interoperability requirements. In the appendix is an analysis of the four approaches mapped against the four layers of interoperability identified in EIRA.

The four approaches are described and modelled in greater detail in the following paragraphs.

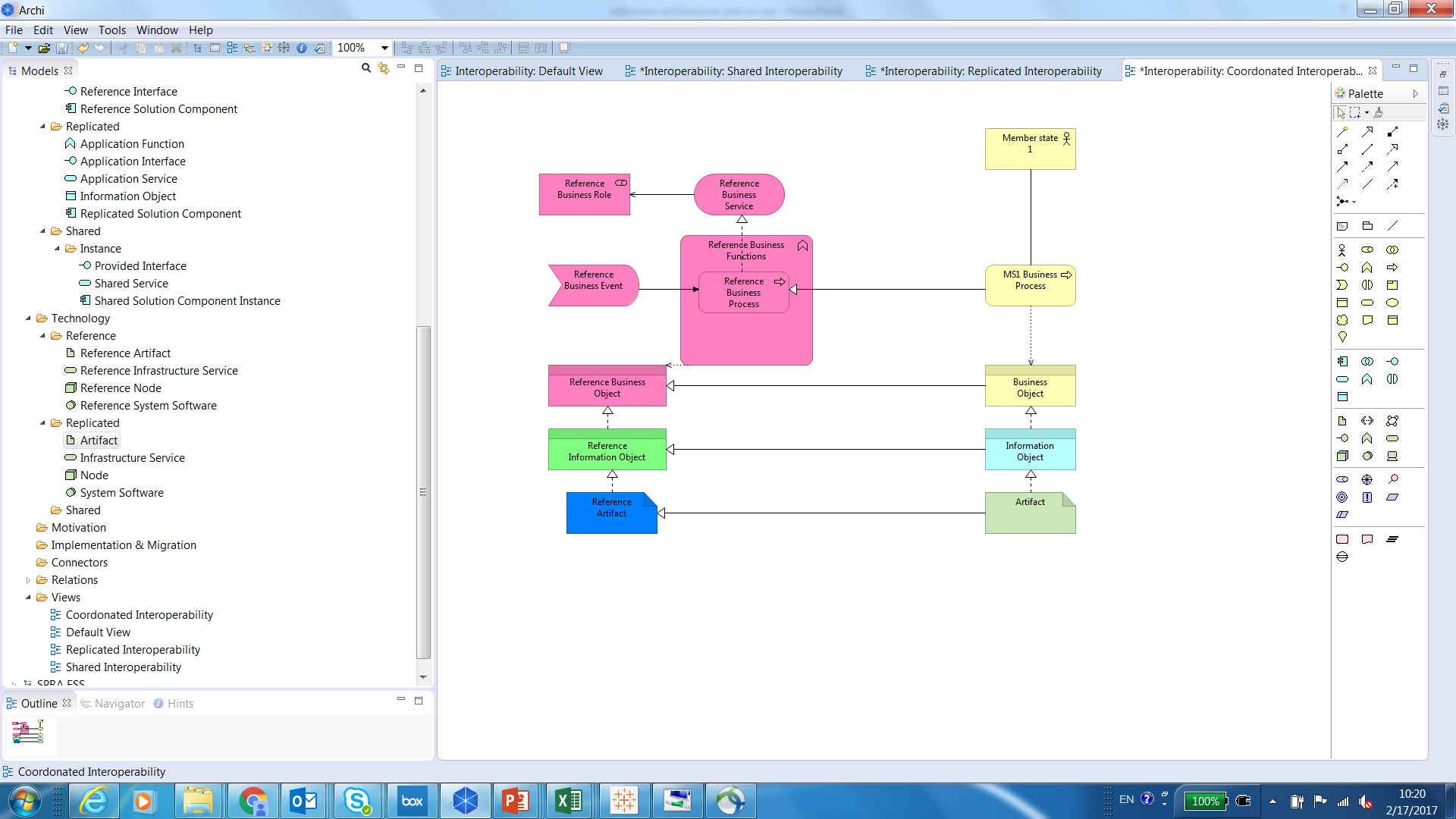
The Archimate models below show, how to model each scenario (except autonomous) in relation to a reference architecture. The models recognize the different level of detail required in each scenario across business, information, application and infrastructure architectures. In essence, each scenario means that ESS members need to agree on different aspects and therefore puts different requirements on the reference architecture**Autonomous:** The application service is implemented as an individual solution of the ESS partner without coordination with others. Optimization takes places at the level of each ESS partner.

* + - From the perspective of interoperability, the responsibility for the required capability remains with the ESS partner and there are multiple and different business processes in place.
    - The approach allows ESS partners to achieve synergies and efficiencies by engaging in knowledge exchanges e.g. on business process modelling, requirements definitions or data definitions.
    - From an architecture perspective, this scenario should be modelled within organization architectures of the ESS members and should appear as black-boxes in the ESS co-created architectures.

**Coordinated:** The application service is implemented as an individual software solution of the ESS partner which nevertheless effectively exchanges information and operates together with others in a coordinated way. This requires stability of standards as when they change, the downstream impact is significant.

* + - From the perspective of interoperability, the responsibility for the required capability remains with the ESS partner.
    - There are coordinated processes and interoperability agreements such as data exchange agreements in place. In this consolidation scenario ESS partners can leverage (statistical domain specific) industry standards.
    - In order to effectively exchange information, messaging middleware manages syntactic interoperability, the application manages semantic interoperability and common transport protocols are being used.
    - From the perspective of architecture, standards for statistical methodology (business processes) are jointly defined and agreed on in this scenario. This means that when agreed they should be modelled in the business layer in Archimate. The modelling should also include which standards to apply for Business objects, Information objects and Artifacts.
    - The following entities should therefore be modelled for all application services in a coordinated scenario, in order for the architecture to be useful:
      * + Business Layer: Modelling of the agreed statistical methodology
        + Application Layer: Modelling of Information Objects and interfaces
        + Technology Layer: Modelling of artifacts representing file formats, standards applied to files, database structures, scripts (e.g. VTL)
        + SLA: Modelling of SLA’s are optional and if defined should be concerned with SLA’s for releasing updates to the relevant standards (frequency of new releases, co-existence of multiple versions, etc.)

In the coordinated scenario, the following concepts and elements at ESS level for application service. As it can be seen on the figure below, the business process and information objects should be agreed and aligned across all member states that participate in this scenario for a given domain.



MS architecture

Reference architecture

1. Illustration of Coordinated scenario

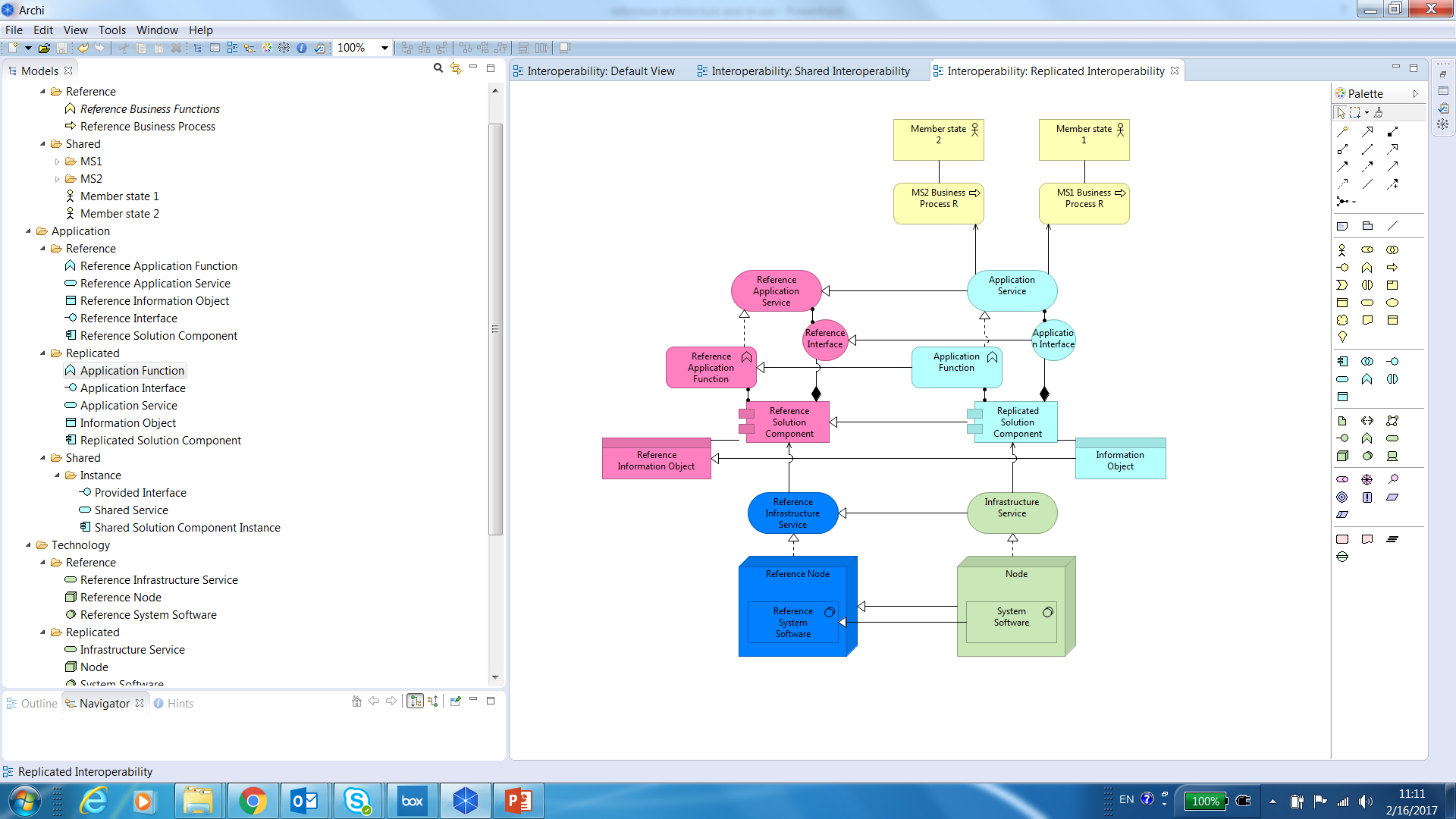
**Replicated:** The application service is implemented as shared software, which supports the execution of similar business processes through replication (possibly with differing configurations) by the ESS partner.

* From the perspective of interoperability, the business processes in place are similar and the solution has been defined in a coordinated way so that data models, structures and format are common.
* The software replicated requires specific infrastructure elements to be in place (such as an OS or a DBMS). A common technological infrastructure can be used and solutions and requirements for end-user technology are common (or closely identical).
* Software maintenance contracts between providers and users of the solution should be maintained to ensure proper management and use of the functionality.
* From the architecture perspective, a set of application services is shared physically between Member States in this scenario.
* Business processes/functions will be similar in order for the software to be useful, but need be not jointly defined/same. This implies that the business layer will differ and therefore has less importance for this scenario.
* Application services and underlying applications and supported infrastructure entities are defined and agreed upon in detail as the core of this scenario is to be able to re-use software in the different environments of the ESS members.

In a replicated scenario,the following entities should therefore be defined at ESS level for all application services::

* + - * + Business Layer: high level of business process definition to highlight the similarities of the business processes
        + Application Layer: Agreed and defined application layer in detail. Replication components provider provides solution following the defined architecture without significant deviations. Underlying Solution Components are defined to the level of identification of service groups
        + Technology Layer: Defined to the level of detail describing underlying infrastructure building blocks such as databases, operating systems, web servers etc.
        + SLA: Separately defined SLAs on the provided software such as maintenance schedules and support agreements.

The illustration below shows that Member States will be implementing different business processes based on the same (or similar) software. Each Member State will have their own processes in this domain, however the major differences should be understood as they provide requirements for the application. The focus of this scenario is on re-usability and installability of the application. Therefore application and technology layers should be modelled and communicated across Member States.



MS architecture

Reference architecture

1. Illustration of Replicated scenario

**Shared:** The application service is implemented as a shared service, which executes identical business processes/service through a single solution instance, possibly but not necessarily utilizing information that is shared across all ESS participants.

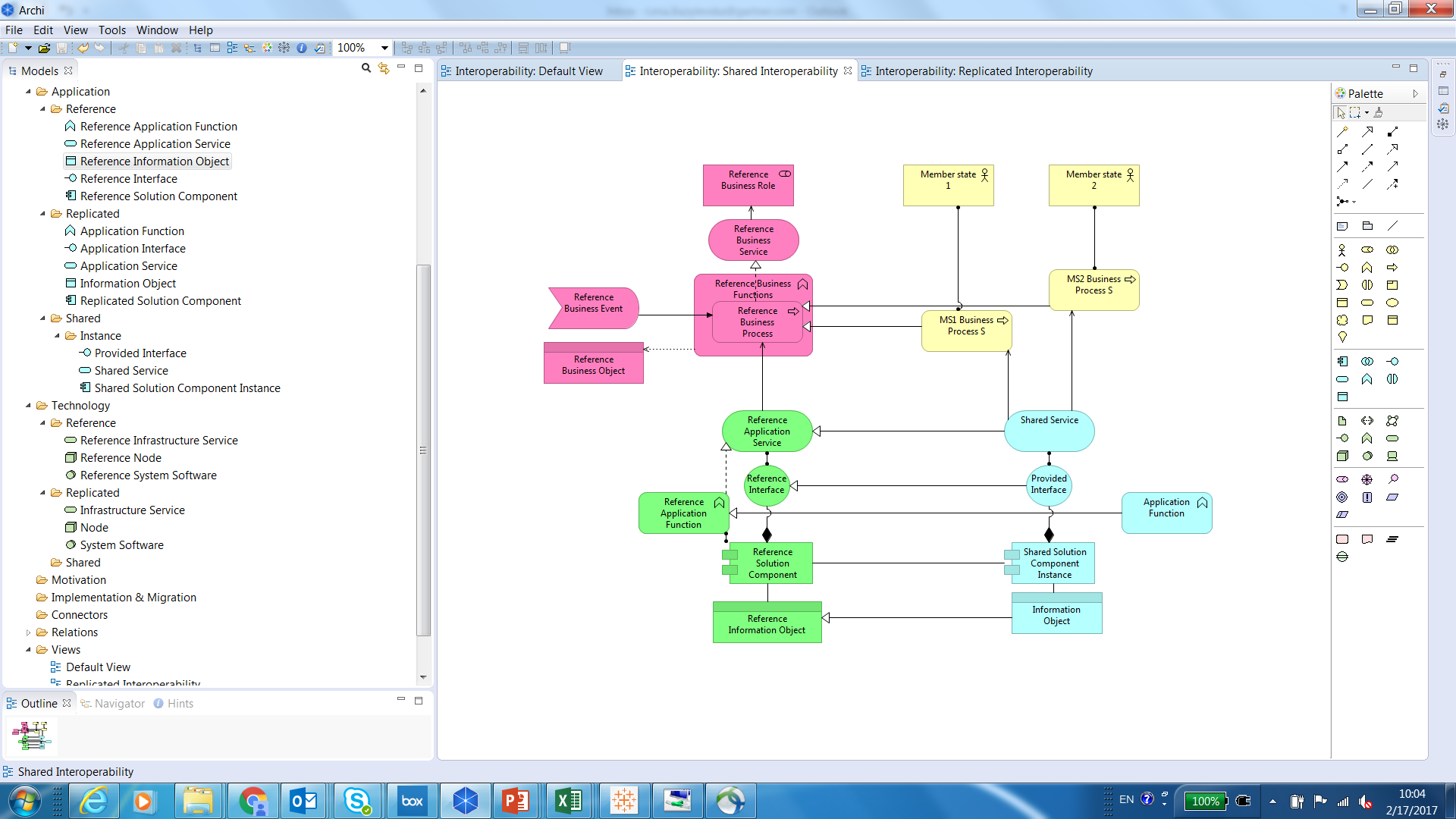
* + - From the perspective of interoperability, this is the consolidation approach that ensures interoperability in the long run. It relies on jointly defined business processes, service contracts between solution provider and users, and single sourcing. The solution ensures semantic, syntactic and technical interoperability.
    - From the architecture perspective, a set of application services is shared physically between Member States in this scenario. Business processes/functions, application services and interfaces are jointly defined and therefore modelled in detail. Each service has a defined set of SLAs.

In a shared scenario perspective, the following entities should therefore be defined at ESS level for all application services:

* + - * + Business Layer: Jointly defined and documented in detail in the industry reference architecture business entities. Member States using this scenario implement business process (organizational architecture) without significant deviations
        + Application Layer: Agreed and defined services and interfaces documented in Industry reference architecture. Shared service provider provides services and interfaces following Industry architecture without significant deviations. Underlying Solution Components are defined to the level of identification of service groups
        + Technology Layer: relying on common standards, but not particularly documented in the reference architecture
        + SLAs: Separately defined SLAs on the provided services defining operational SLA’s: recovery time objective, recovery point objective, availability. It also requires SLA’s on maintenance schedules and support.

In the scenario below, agreements should be made about business processes, interfaces, what a shared service will do (function) and which information objects will be exchanged. All deviations in business processes should be understood, modelled and mapped to the reference architecture of that domain. Technology layer is modelled only as a part of solution architecture of the hosting party.

Reference  
architecture



MS architecture

1. Illustration of Shared scenario

**Summary of scenarios**

In the Annex “Illustration of consolidation approaches” are three examples on how the shared, replicated, and coordinated scenarios should be interpreted in three specific cases: validation, SDMX converter, and for Metadata Management.

The different scenarios carry with them different potentials for synergies and poses different requirements on interoperability. The shared scenario carry a potential for significant synergies but creates direct dependencies between business processes of ESS members. The replicated scenario does not require a specific dependency between business process executions but require common infrastructure components or required adaptors such as containers or data base connectors. The coordinated scenario ensures interoperability but offers a more costly option, which also makes it more challenging to cater for change

The EIRA distinguishes four layers of interoperability, which should be considered when deciding the most suitable approach to consolidation.

It should be noted that the SPRA supports the co-existence of more than one scenario for consolidation such that a coordinated approach may be decided among a majority of ESS members while a minority decides to develop and use a shared service or replicate a software solution.

* 1. Service orchestration

In a specific statistical production process, a number of Services will be orchestrated using a Process Orchestrator, which will manage the execution of the statistical production process. A statistical production process will thus consist of a number of (GSBPM) phases and steps where Services are invoked together with datasets managed by the data management component (e.g. a data storage) as well as Metadata (parameters), all managed by the Process Orchestrator. The output of the Service invoked will then be stored in the data management component and/or in the Process Orchestrator and used as input for the next Service or used as the basis for manual inspection and manipulation.

The use of a Process Orchestrator using a standard notation such as BPMN and using shared Services will ensure an easy replication of statistical production across domains and geographies and therefore minimize the cost of adjusting or expanding statistical production.

The technological implementation of process orchestration may take multiple forms using ETL platforms, BPMN suites or data analytics workbenches, but agreeing on a shared format for describing and exchanging statistical production process offer a opportunity for efficient and coordinated change of a specific production.

* 1. Statistical Data Management

A key element in any architecture dealing with statistical production is the approach to the management of data. The SPRA models data management taking into account that requirements for data management are evolving rapidly with the increased use of administrative data sets and transactional data sets (big data).

From one perspective, the ESS is engaged in exchanging statistical data sets to be used in consolidated European statistics as well as in improving production of national statistics (e.g. ESBRs). Therefore interoperability at the level of data is key and a long-standing discipline within the ESS. The expansion of reuse and sharing beyond data also requires a coordinated way of managing data. Data virtualisation or data abstraction is key to the success of this. SDMX can be the preferred standard for this data abstraction but may require supplementing approaches to support micro data and more resource-efficient data formats for big data.

There are a number of new needs for data management, which leads to the emergence of new architectural approaches to data management. One of the important emerging trends is logical data warehouse. The logical data warehouse is a new data management architecture for analytics combining the strengths of traditional repository warehouses with alternative data management and access strategies. This approach is able to meet the current Member States demands for flexibility and new data sources provided in different formats and with varying frequency (e.g. streaming data, unstructural data).

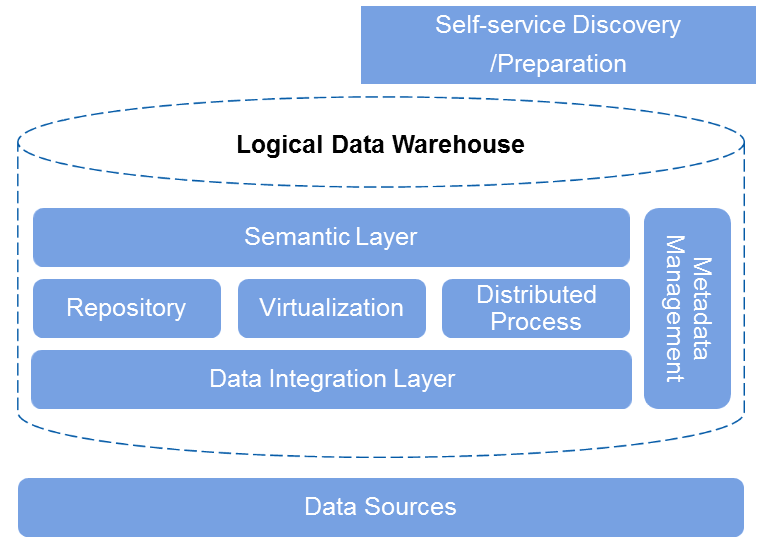


Figure 1 Logical data warehouse components

The logical data warehouse and supporting infrastructure consists of a number of key components:

* **Data integration:** Batch, ETL, Data Replication, Messaging, Federation, and other techniques.
* **Repository, virtualization and distributed process:** Three styles complement one another and provide an end-to-end information infrastructure platform for analytics in the big data era.
* **Semantic layer:** This maps disparate data into unified schemas to obtain semantic consistency and improve reuse.
* **Metadata management:** This provides traceability for the LDW information supply chain in order to support semantic consistency, impact analysis and regulatory compliance.
* **Analytic tools:** Consume data, provide a business-centric view and perform advanced analytics.
* **Self-Service and Data Preparation:** an iterative agile process for exploring, combining, cleaning and transforming raw data into curated datasets for data science, data discovery, and BI and analytics.

This trend is reflected in the SPRA building blocks below. Data integration function is supported by ESS Data Exchange, Data loading and Data processing building block. Repository function is supported by Data Staging Storage and Primary Data Storage. Virtualization and Metadata management can be mapped consequently one to one to Data Vistualisation and Metadata management. Distributed process inherent as a part of more advanced Data processing. Semantic Layer is distributed across Metadata management and Data Virtualisation. Self-service data discover and preparation is an addition to logical datawarehouse that enables to improve exploration of data across statistical domains, which can be mapped to Data exploration and Statistical analysis building blocks.

The other important trend is the emergence of different storages that have different characteristics. Minimally there is a need for data staging storage (that facilitates confidential and non-confidential requirements) and non-production data storage in addition to primary and dissemination data storages. Data staging storage facilitates data that is coming from the ourside of Member States that has to be checked and evaluated from quality and security point of view before moving it into production storage. Non-production storage is required for storing all data that is accessed rarely and the main characteristics of which is to provide storage for different data types with varied volumes and quality.

* 1. Metadata Management

The SPRA distinguishes three types of metadata as these will be managed differently and have different requirements for interoperability.

The distinction between **passive** and **active** metadata is an important driver for defining the SPRA. Traditionally metadata are passive,meaning that they are used for documentation purposes and human interpretation. In order to support computation and automation of processes, “Activating” the metadata is key. **Active metadata** are metadata stored and organized in a way that enables operational and copmutational use, manual or automated, in multiple processes.

The SPRA is assuming an extensive use of active metadata, distinguishing three types of metadata[[3]](#footnote-4) as they are supported quite differently in the architecture:

* **Process metadata** are metadata that describe the expected or actual outcome of one or more processes using evaluable and operational metrics
* **Reference metadata** are metadata that describe used concepts, used methods and quality measures for the statistical data
* **Structural metadata** are metadata that help the user find, identify, access and utilize the data
* **Quality metadata** are any kind of metadata that contributes to the description or interpretation of the quality of data.

The SPRA supports the extensive use of active metadata. The approach taken to data management and data abstraction is to simplify as much as possible the need for technical metadata in order to support interoperability.**Technical metadata** are metadata that describe or define the physical storage or location of data.

Conceptually speaking, the *Process Metadata* is managed by the *Process Orchestrator*, which will store information about the different production processes as well as information about each execution of the process. The Metadata managed by the Process Orchestrator will be: links to data sources (*statistical micro data*), Services invoked, parameters for Services invoked, as well as links to *statistical macrodata*, or any other *statistical output*.

The reference metadata will be managed by the Metadata Management building block.

The structural Metadata will be partly managed by the primary data storage, partly managed through the Metadata Management System and accessible to the Process Orchestrator through Service interfaces. The structural Metadata managed by the primary data storage will be the Metadata describing the specific data sets, whereas the structural Metadata managed by the Metadata System will be generic data such as structured code lists.

Quality Metadata will be managed and available in the quality assessment building block and will base itself on process-, reference- and structural metadata available in the metadata management building block.

Most Metadata should be available through a Metadata Management building block, which logically, but not necessarily physically unifies access to the Metadata.

The design of statistical production processes will take place using the same functionality provided by the process designer and services.

1. Building blocks of the SPRA
   1. Description

The ESS EARF building blocks group related application and infrastructure functions for supporting the production of official statistics.

Following TOGAF, the boundary and specifications of a Building Block (BB) should be loosely coupled to its implementation, i.e. it should be possible to realize the BB in several different ways. An appropriate choice of BBs can lead to improvements in system integration, interoperability and flexibility in the creation of new information systems.

In the ESS EARF, the BBs are kept at a **high level of aggregation**. They represent a way to structure potential candidates for reuse across statistical subject matter domains, phases of the statistical business process and members of the ESS. It is a grouping of functions within a statistical organization. They help structuring artifacts (e.g. functions, solutions, specifications) and make them findable in registries and repositories.

There are two types of Building Blocks required within the ESS context: a set of Building Blocks that represent the statistical production architecture (industry reference architecture) required to conduct statistical business and a set of Building Blocks that are required to collaboration between Member States in order to achieve ESS goals.

* **Industry reference architecture**: One set represents a high-level industry reference architecture, the purpose of which is to structure (potentially re-usable) components of **(IT) capabilities within a statistical organization.** The purpose of this group of building blocks is to identify/describe shareable/replicated/coordinated solutions across Member states;
* **ESS Collaboration:** The other set represents building blocks that are centrally located to facilitate collaboration within ESS. The purpose of these building blocks is to have a set of collaboration platforms for Member states.

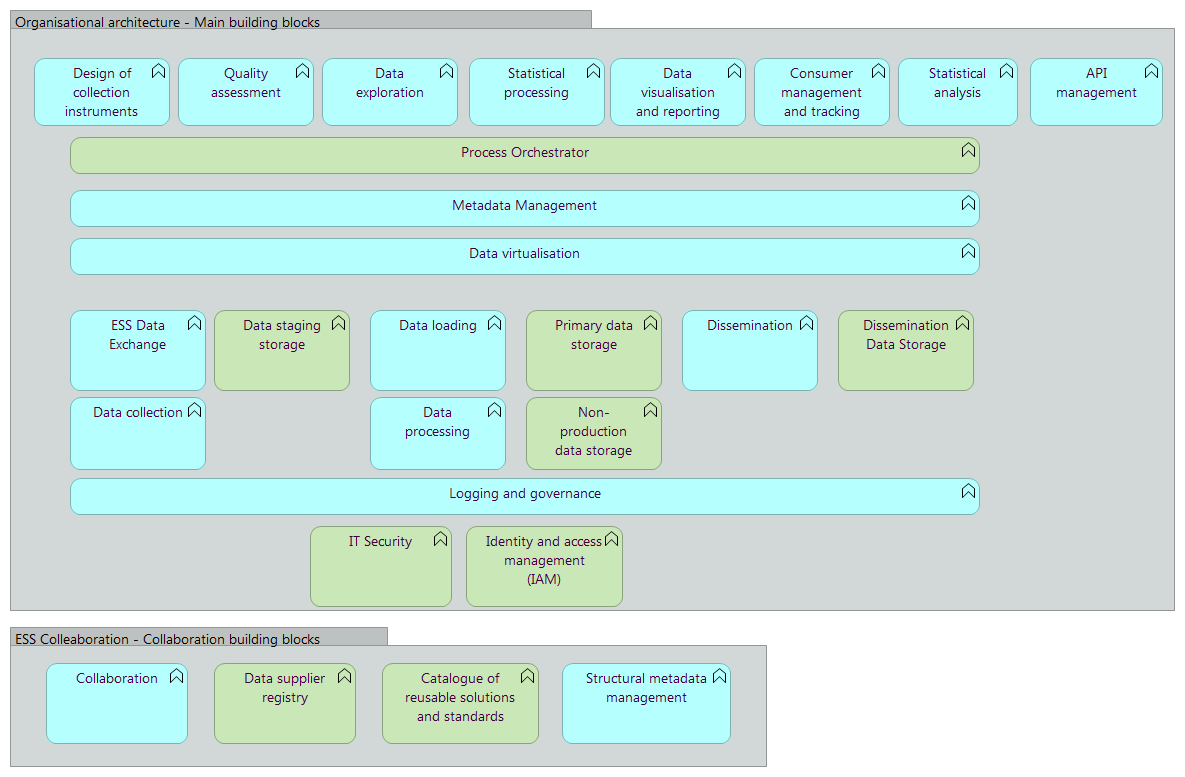


Figure 2 Industry Reference Building Blocks

Industry Reference Building Blocks will structure repositories and registries describing re-usable interoperability scenarios for different services. Different scenarios have different specification needs; those specifications are grouped according to building blocks to make them easily findable for Member States.

Example of potential structure: Validation architecture

* Model/Architecture/Statistical processing/Validation
  + /Overall
  + /Shared
  + /Replicated
  + /Coordinated
* Views/Architecture/Statistical processing/Validation
  + /Overall
  + /Shared
  + /Replicated
  + /Coordinated

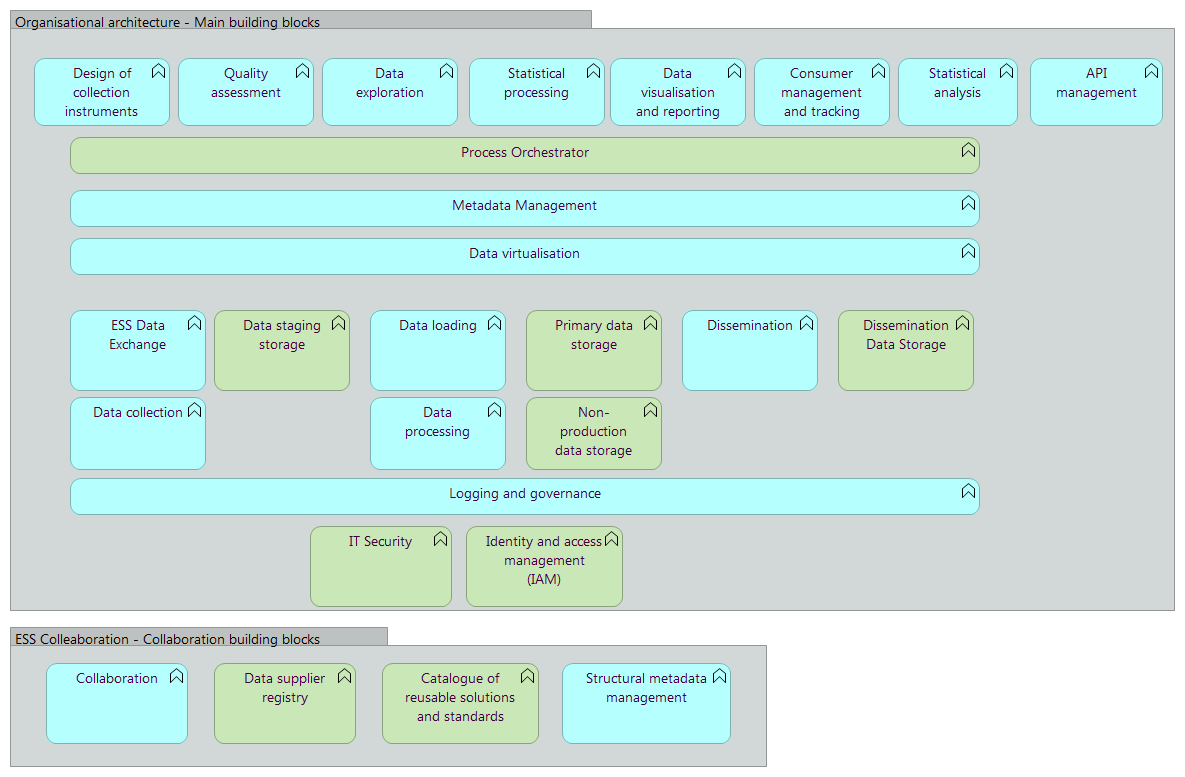


Figure 3 ESS Collaboration Building Blocks

From architecture perspective ESS collaboration Building Blocks will be realized centrally and will be always shared and not directly related to the core statistical processing. Therefore, architectures behind them should be described in a separate centralized (sub-) repository for the purpose of specification below lists and describes the ESS EARF BBs that constitute the major elements in the envisioned ESS IT landscape. Following the classification used in Archimate language, the ESS EARF classifies BBs into two categories:

* **Application Functions:** An application function describes the internal behavior of an application component (in other words a block of functionality). An application function abstracts from the way it is implemented. Only the necessary behavior is specified. An application function may realize one or more application services, for example SPRA services.
* **Infrastructure Functions:** An infrastructure function describes the specific behavior of an infrastructural component such as a database management system. An infrastructure function abstracts from the way it is implemented. Only the necessary behavior is specified. An infrastructure function may realize infrastructure services.

The former can be seen in the figures above in blue and the latter is colored in green.

|  |  |
| --- | --- |
| **ESS EARF BBs** | **Characteristics** |
| **Organisatonal architecture** | |
| **Design of collection instruments** | Building Block that supports the design of collection instruments so they can be produced faster and in a more coherent way across ESS members and statistical business processes. Ways to achieve this are the use of harmonized expression languages to build the surveys and harmonized structures for capturing data. |
| **Data collection** | Building Block that allows the collection of data and Metadata and loads the data in the Primary Data Storage. |
| **Statistical processing** | Building block, which bundles different statistical processing, transforming incoming data into output statistics. Examples of such Services would be “classification & coding”, “data aggregation”, etc. The Services should by design collect and track the Metadata required for quality monitoring purposes and trigger corrective/preventive action to address non-compliance with quality requirements where relevant. |
| **Statistical analysis** | The building block bundles various services for statistical analysis services. Examples of such Services would be “validation of data”, “output finalization”, etc. The Services should by design collect and track the Metadata required for quality monitoring purposes and trigger corrective/preventive action to address non-compliance with quality requirements where relevant. |
| **Data exploration** | Scalable platform to respond to new statistical needs and help explore potentially very large data sets, with possibly strong privacy requirements. The platform provides sophisticated features to aggregate data and compute indicators, perform statistical operations such as quality assurance and output editing as well as visualize data. |
| **Dissemination** | ESS-wide platform that enables the dissemination of statistical outputs. The platform’s set up and composition will need to reflect the policy & political priorities the ESS participants have agreed upon. |
| **Consumer management and tracking** | To support a consistent surveying of statistical data consumers to monitor user satification and query needs and analyze behaviour and feedback. |
| **Dissemination Data Storage** | Data storage where the data and statistics for dissemination are stored. Like the Primary Data Storage, this Building Block includes trusted recovery providing recovery facilities such as restoring from backups in ways that do not compromise security protection. |
| **Data Staging Sotrage** | Interim data storage with data collected from external data sources. Data storage where the collected data are stored. The Building Block supports large amounts of data and the flexible addition of new data sources. |
| **Primary Data Storage** | Data storage where the primary data from production are stored. The Building Block supports large amounts of data and the flexible addition of new data sources. It can be split into micro (unit level) data storage and macro (aggregated) data storage as different information (governance) models may apply. The Building Block includes trusted recovery providing recovery facilities such as restoring from backups in ways that do not compromise security protection. |
| **Process Orchestrator** | As a key element in the ESS, this Building Block will act as BPM implementation and orchestrate the statistical processing Services into production processes and support the design and management of these processes.The Process Orchestrator can also include select security Services such as audit (the ability to provide forensic data attesting that the systems have been used in accordance with relevant security policies) and access control for subjects (such as processes) and objects (such as files, including control of object creation and deletion). |
| **IT Security** | Secure hosting and network infrastructure providing for secure networking hardware, software and network services as well as additional computing capacity. |
| **Identity and access management (IAM)** | IAM copes with authentication (i.e. the substantiation of the identity of a person or entity) and authorization (i.e. the definition and enforcement of permitted capabilities for a person or entity whose identity has been established) in order to prevent resources from unauthorized and unintended use. It typically also includes system entry control warning unauthorized users, tracking unsuccessful login attempts and locking sessions when required; as well as non-repudiation proving that a user carried out an action, or sent or received data at a particular time.The Building Block shall implement trust-based identity and access management relying on legal, organizational, semantic and technical interoperability agreements between ESS participants. Such agreements include joint definition of access levels and role-based access criteria as well as harmonized security requirements. |
| **Logging and governance** | This Building Block is required to support juridical, governance and audit use cases. Platform that can be used to log and govern data across all systems and processes to enable traceability and governability.All systems usages should be logged for audit purposes and audit should be easily facilitated in case a need emerges from legal or data quality perspectives |
| **Quality assessment** | Quality assessment supports analytics of quality indicators. This should be integrated with Metadata Management. |
| **ESS Data exchange** | Platform that can be used to report collected data to either DG ESTAT or other international institutions. The platform facilitates sending and receiving sets of data among ESS members, providing functionalities such as Service message routing, message transformations and data en/decryption. ESS participants should still implement member-specific integration platforms to support internal Service integration. |
| **Data Virtualisation** | Creates virtualized and integrated views of data that may be stored in memory if needed (rather than executing data movement and physically storing integrated views in a target data structure). It provides a layer of abstraction above the physical implementation of data. It is based on the execution of distributed queries against multiple data sources, federation of query results into virtual views, and consumption of these views by applications, query/reporting tools or other infrastructure components. |
| **Data visualisation en reporting** | This Building Block supports the building of reports and visualisations that can be used in ESS Member to support various aspects of statistical production. Functionally this Building Block will also be used to support dissemination. |
| **API management** | A platform to convert data into different formats and provide an access to different end-points (e.g. sparql, REST API, webbrowser) |
| **Data loading** | Building Block that allows the collection of data and metadata and loads the data in the Primary Data Storage. |
| **Big Data processing** | Building Block that supports the proccessing of Big Data and loads the resulting data in the Primary Data Storage. This includes the sampling of very large data sets |
| **Non-production data storage** | Data storage where non-production and copy of production data is stored for ad-hoc and big data analysis. The Building Block supports large amounts of data and the flexible addition of new data sources. It can be split into micro (unit level) data storage and macro (aggregated) data storage as different information (governance) models may apply. |
| **ESS collaboration** | |
| **Collaboration** | This is an ESS platform that enables a collaborative way of working among ESS statisticians across the various phases of the statistical production process. It supports specialized (closed) communities and structured collaboration on developing (new) statistical outputs and processes. |
| **Catalogue of reusable solutions and standards** | The catalogue contains references to various artifacts and Services to support standardization and efficient sharing and reuse of process, information and Services at ESS Level. It contains among others references to methodological guidelines, resources (centre of excellence, helpdesk, …) and IT Services. |
| **Data supplier registry** | This Building Block contains key information on potential data providers as well as the Service Level Agreements with them. |

Table 1 ESS EARF BBs

The Metadata Management Building Block is broken down in threy key building block elements to reflect the computationally diverse functionality required to manage the different types of metadata.

|  |  |
| --- | --- |
| **Metadata management BBs** | **Characteristics** |
| **Metadata management** | Building block that manages the entire life cycle of structural, reference, and process Metadata and logically connects these data. This will form a key Building Block and prerequisite for standardizing and industrializing statistical production. The building block ensures the availability and integrity of Metadata, as well as the suitability of Metadata for consistent quality monitoring in the ESS. This also includes data lineage tracking, quality management and data security assurance. |
| **Structural metadata management** | Management of data describing the data set as well as reference data such as classifications, code lists, etc.  It manages the entire life cycle of structural Metadata that are shared between ESS members such as code lists, reference tables etc.This will include both ESS-agreed metadata as well as link to other international meta data standards. |
| **Descriptive Metadata management** | Management of information facilitating the interpretation of the data sets, how they were produced, etc. |
| **Process metadata management** | The management of the process metadata, i.e. the metadata describing the steps, services, and parameters used to produce the data set. This building block will be tightly connected to the process orchestrator. |

Table 2 ESS EARF Metadata management BBs

* 1. User story illustrating statistical production and the role of the building blocks

###### The user story serves to describe how the various building blocks are mobilised to provide an end-to-end solution to the user

###### User Story: Development and production of national level statistics from multiple data sources

The user story describes a statistical production process which combines data from various data sources. This includes managing data of differing quality and using different means of access.

**Storyline:**

Marco, a Member State statistician, is responsible for producing national level statistics. For this, he regularly receives administrative data sets from national authorities. He is also provided access to a database at the national bank as well as receives a data set in .CSV from a private data provider.

The data sets submitted by national authorities are validated, transformed, and loaded into the **Data staging storage**. Marco triggers this upload in the **Process Orchestrator** using a pre-defined workflow orchestrating a number of services from **Statistical Processing**. The workflow validates the .csv file and transfers it to **Primary data storage** as-is. As Marco has recently modified the statistics by also making use of this data set, he has registered the data set in the **Metadata Management**, allowing his services from **Statistical processing** to interpret the file correctly as well as providing easy access to the file by others.

When the data sets and the .csv file are in the **Primary data storage**, Marco executes a second workflow in the **Process orchestrator** orchestrating another set of **Statistical processing** services, which uses the two data sources as well as opens a query through the **API management** in the national bank database to produce the output data set, which gives an index for all individual regions as well as aggregate national numbers.

Once visually inspected using the **Data visualisation and reporting** building block, Marco executes a third workflow using the **Process orchestrator** and services from **Statistical processing**, which transfers the data to the **Dissemination data storage**. In addition to the index, the Member State data sets are transferred to be used by data scientists in the national bank as well as in the public domain. The process above is visualised in the figure below.



Figure 4 Visualisation of Regional Index production using Building Blocks

With one of his colleagues, Marco is testing out a new algorithm using **Statistical Processing** and **Statistical Analysis** for calculating the index based on two additional data sets located in **Non-production data store** from a national regulator as well as an additional private data provider. He uses **Data loading** to load the data sets in the **Non-production data storage** and register the data sets in the **Metadata management** as data that are not used for production.

He is using functionality from **Statistical analysis** and **Statistical processing** combining his existing data sets with the additional new ones using **Data virtualisation** to easily combine the data. After approving the revised approach, it is modelled and tested in the **Process Orchestrator** before the revised production process is put into productions using the data sets.

* 1. Actual use and templates

The ESS EARF Building Blocks are may be used as starting points for initiatives to realise the future state architecture for the ESS. Some of the building blocks form candidates for sharing (such as metadata management or design of collection instruments), while ensuring interoperability at the level of building blocks (such as e.g. statistical processing) will support reuse and sharing. **Examples of implementations** of BBs can be found in the **Annex**, Implementations of Building Blocks.

1. Services in the SPRA: Data Collection
   1. Context

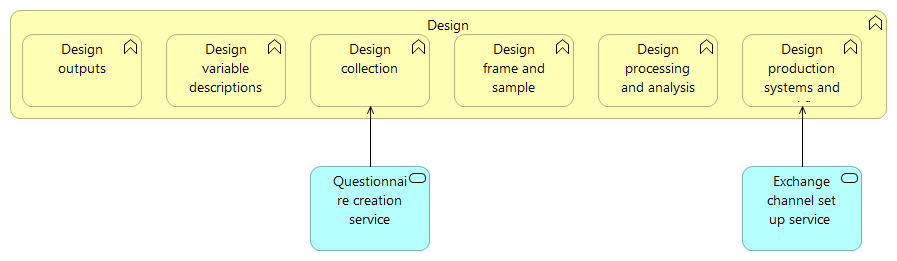
The Vision 2020 gives explicit mention to the modernization of data collection enabled by information and communications technologies and enhancements of the data collection process.

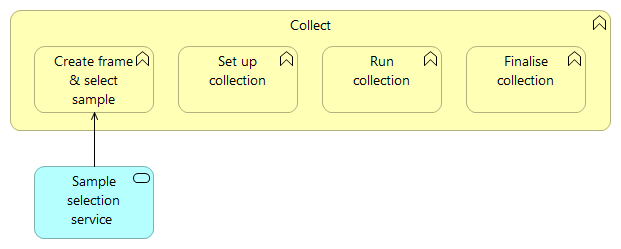
The architecture shall enable new modes of data collection and exchange such as “pulling” data from providers and leverage on the rapidly accruing amounts of data that could be used in the statistical production chain (key word: “Big data”) by allowing to collect administrative and other non-survey data.

* 1. ESS relevant services

Figure 4 identifies 4 relevant services which could be made available to ESS partners.

For modelling purposes, Data Collection is notmade specific to an NSI and Eurostat despite Eurostat is principally collecting data from NSIs and from some commercial providers (Group register) and NSIs are predominatly focusing on the execution of surveys. With the advent of new sources, this paradigm may change with some NSIs collecting and preprocessing for the ESS some big data sources





1. ESS Statistical Production Reference Architecture – Design and Data Collection phases

Table 2 contains a brief description of the Services identified in the SPRA for the Collect phase.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **GSBPM Phase** | **GSBPM sub-phase** | **Name** | **Description** | **Category** |
| 2. Design | Design collection | Questionnaire creation service | Service which allows the creation of questionnaires. | Application service |
| 2. Design | Design production systems and workflow | Exchange channel set up service | Service which sets up exchange channels for the exchange of data between organizations | Application service |
| 4. Collect | Create frame & select sample | Sample selection service | Service which selects units and records from populations | Application service |

1. SPRA services for Collect phase
   1. Design Principles

The next table summarizes the Design Principles for a to-be state architecture of Data Collection.[[4]](#footnote-5)

|  |  |
| --- | --- |
| **Name** | **Standard definition and description language** |
| **Statement** | Data collection instruments are designed and defined using a standard definition and description language. A Data collection instrument such as a survey is described using a standard description and definition language defined for the statistical domain as guided by the list of agreed standards. |
| **Rationale** | To ensure that Data collection instruments as well as data can easily be exchanged and re-used between statistical domains and between NSIs within the same statistical domain. |
| **Implications** | Data collection systems are designed to use the agreed standards for description and definition. Legacy Data collection systems are provided with interfaces supporting the communication of Data collection instruments. |
|  | |
| **Name** | **Unified Metadata system** |
| **Statement** | Metadata from Data Collection is stored in the unified Metadata system. The Metadata produced during the design and execution of data collection instruments is captured in the unified Metadata system. |
| **Rationale** | The capture of Metadata should support the vision of reusing Metadata across the production cycle as well as the transition from passive to active Metadata. |
| **Implications** | Metadata is captured in the Data Collection systems in a format enabling the storage and re-use of Metadata in the unified Metadata system. |
|  | |
| **Name** | **Re-use of existing instruments** |
| **Statement** | Data collection instruments are built reusing already existing instruments where possible. Data collection instruments should - when possible - reuse or build on existing and proven Data Collection instruments, including surveys, administrative data sources and other. |
| **Rationale** | The cost of building new collection instruments and therefore designing new statistics is reduced. The quality of data collection instruments is improved through the systematic re-use. |
| **Implications** | Design of Data collection instruments is never done from scratch. It starts with analysing existing collection instruments, including those from other statistical domains. |

1. Design Principles for Data Collection in the ESS Statistical Production Reference Architecture
2. Services in the SPRA: Process & Analyse
   1. Context

The Vision 2020 of the ESS has provided some guidelines and context for establishing Process & Analyse in the SPRA. A number of specific elements in the Vision 2020 shape the design of this part of the architecture

*“We want to intensify our collaboration by moving towards sharing data, Services and resources. The collaboration will be based on standards and common elements of IT and statistical infrastructure, all of them using the modern design of statistical production.”*

*“Further identification and implementation of standards for statistical production is crucial for improving the comparability of statistical outputs. Standards are required to ensure a smooth communication in the system and to make process components interoperable.”*

*“For specific steps of the statistical production process we can use the same methods and tools across NSIs. Well-established examples include seasonal adjustment, disclosure control and administrative data validation methods. We will explore other areas. In particular in dealing with new technology-driven areas like big data, open data and visualization techniques we expect concrete opportunities.”*

*“Several factors should be considered to harness new sources in statistical production. The use of these data in a multi-source statistical production environment requires changes in the production architecture and investment on different dimensions both at national and ESS level.”*

*“…we must improve our efficiency through systematic collaboration, while fully respecting the subsidiarity principle.”*

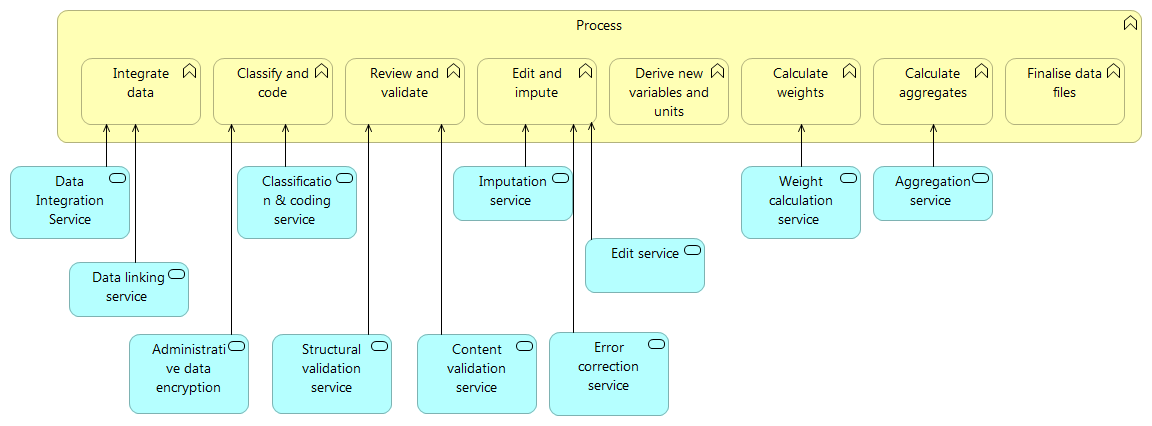
The statistical production of a specific statistical domain in the context of the presently described architecture comprises the phases starting with Data Collection and ending with Dissemination, including the processing and analysis of data. However, the GSBPM Design and Build phases are also crucial.

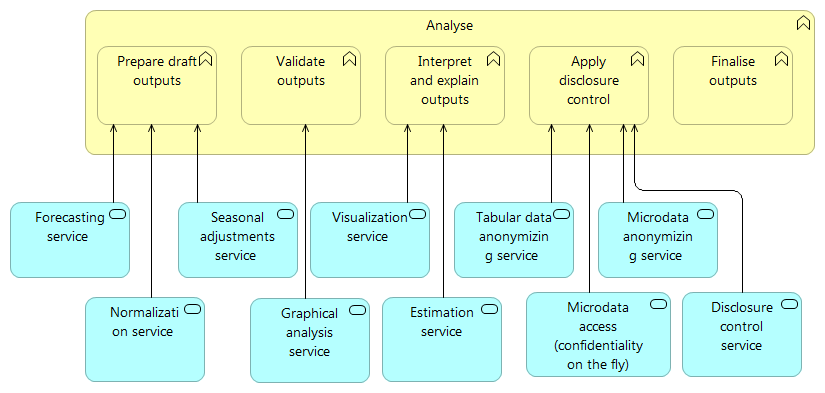
Concepts for the Statistical Production System have been developed in a number of efforts, most notably and thoroughly in the Common Statistical Production Architecture (CSPA) project under the High-Level Group for the Modernisation of Statistical Production and Services (HLG). CSPA defines a Service-oriented approach for the implementation of statistical production architecture; the ideas of CSPA underlie the architecture proposed in this document.

Statistical Process & Analyse in the SPRA implemented according to the guidelines suggested here would make statistical production significantly more efficient. The development and implementation of new statistics would in particular become much more efficient as the Statistical Production System supports the sharing of investments in efficient processes and statistical Services. Reaping these benefits would require significant changes and assumes that a number of standards are agreed in the ESS (including a business process description language) and that there is a uniform approach to Metadata. The Statistical Production System would be the key element in realizing the movement from passive to active Metadata described in the Information view of the ESS EARF.

* 1. ESS relevant service

Figure 5 below visualizes the to-be state for Process & Analyse in the SPRA.





1. ESS Statistical Production Reference Architecture – Process and Analyse phases

Table 4 contains a brief description of the Services identified in the SPRA for Process and Analyze.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **GSBPM Phase** | **GSBPM sub-phase** | **Name** | **Description** | **Category** |
| 5. Process | Edit and impute | Imputation service | Service which inserts new values for data considered missing, incorrect and unreliable based on different algorithms | Application service |
| 5. Process | Calculate weights | Weight calculation service | Service which calculates weights to statistical data | Application service |
| 5. Process | Calculate aggregates | Aggregation service | Service which creates aggregate statistical data from micro data | Application service |
| 5. Process | Classify and code | Classification & coding service | Service which assigns classification categories to unstructured data fields | Application service |
| 5. Process | Review and validate | Content validation service | Service which provides various content validation mechanisms (checksums, identification of outliers, etc.) | Application service |
| 5. Process | Review and validate | Structural validation | Service which checks whether the data has the right structure | Application service |
| 5. Process | Integrate data | Data integration | Allows combining data from different sources by providing a unified view | Application function |
| 5. Process | Integrate data | Data linking service | Service which allows matching of data records | Application service |
| 5. Process | Edit and impute | Edit | Software prrovided by ESTAT to process data | Application function |
| 5. Process | Edit and impute | Error correction | If data are considered incorrect, missing or unreliable, new values are inserted according to deterministic rules. The rules include the determination of whether to add or change data, adding or changing data values, writing of data values back to the data set and flagging them as changed. | Application service |
| 5. Process | Classify and code | Administrative data encryption | Service that provides the mechanism that converts data for the purpose of disabling the recognition of the unit | Application service |
| 6. Analyze | Interpret and explain outputs | Estimation | Service that provides estimation service for projection | Application service |
| 6. Analyze | Apply disclosure control | Tabular data anonymizing service | Service which protects tabular data against disclosure | Application service |
| 6. Analyze | Apply disclosure control | Microdata anonymizing service | Service which anonymizes data to prevent the identification of specific units in the data set | Application service |
| 6. Analyze | Interpret and explain outputs | Visualization service | Service which visualizes data for analysis purposes | Application service |
| 6. Analyze | Prepare draft outputs | Seasonal adjustments service | Service which decompose series into different components ( trends/cycle, seasonal component, calendar effects, irregular) and produce specifc output like seasonally adjusted series | Application service |
| 6. Analyze | Prepare draft outputs | Forecasting service | Service which predicts time series future behavior based on historical data | Application service |
| 6. Analyze | Prepare draft outputs | Normalization service | Service which normalizes unstructured data | Application service |
| 6. Analyze | Apply disclosure control | Disclosure control | Service that ensures that the data to be disseminated do not breach the confidentiality on collectd information. This may include checks for primary and secondary disclosure, as well as the application of data suppression or perturbation techniques. | Application service |
| 6. Analyze | Apply disclosure control | Microdata access (confidentiality on the fly) | Service where the disclosure limitation routine is  applied dynamically when the data items  are retrieved, after any selection.. | Application service |

1. Services in the SPRA: Process & Analyze
   1. Design Principles

The next table summarizes the Design Principles for the to-be state of the SPRA.

|  |  |
| --- | --- |
| Name | **There is a shared statistical process language** |
| Statement | The statistical production process use a shared process description language such as Business Process Model and Notation like : BPMN,VTL, Other technical standards are used when ever relevant (e.g. Web Services Business Process Execution Language (WS-BPEL) 2.0 or XML Process Definition Language (XPDL) 2.1. |
| Rationale | The agreement of a shared language for describing statistical processes will enable a fast and low-cost replication or adaption of a new statistical production processes. |
| Implication | When, implementing a process orchestrator, the shared process description language should be supported. |
|  | |
| Name | **Statistical Services are shared** |
| Statement | Statistical Services are preferably made available to the ESS community under a shared distribution model (if feasible), otherwise replicated or interoperable (in this order of preference). |
| Rationale | In order to reap the investments made in shared Services, ESS members should adopt the principle of first using an available shared Service, then of replicating it, and only in the case this would not be feasible, opt for only an interoperable Service. |
| Implication | ESS members should consider the practicality and value of reusing a service in the design phase to ensure that the design and implementation is ready for sharing. |
|  | |
| Name | **Services are usable whilst generic** |
| Statement | Statistical Services should be generic where feasible. |
| Rationale | Statistical Services should strike a balance between being easily useable by various statistical domains and general availability. Usability should always supersede generality. |
| Implication | The design of statistical services should be designed with the view to bes used in other context than for which it is developed but should not be designed to be shared if there is not an obvious case for sharing. |
|  | |
| Name | **Process Metadata is managed by the Process Orchestrator** |
| Statement | All process Metadata, which documents the data sources, Services and specific parameters used for invoking a Service, is (logically) created and managed by the Process Designer and Orchestrator. |
| Rationale | The design and execution of statistical production processes will produce and be primary user of process Metadata. The process Metadata will then be made available to other processes, such as quality management or producing reference Metadata via the unified Metadata Management system. |
| Implication | Process Metadata is managed separately from structural metadata and reference metadata. |
|  | |
| Name | **Statistical design takes place in a production system sandbox** |
| Statement | The development/design of new statistical production will take place in a sandbox implementation of the process designer and Orchestrator component. |
| Rationale | The design of statistical production will imply that a very limited development effort is required to support a new production process. The development effort involved will be limited to the possible development/wrapping of a new statistical Service necessary in the production process. |
| Implication | The design and implementation of statistical production processes will to a much larger extent take place without the involvement of IT developers. |

1. Design Principles for Process and Analyse in the ESS Statistical Production Reference Architecture
2. Services in the SPRA: Dissemination
   1. Context

Vision 2020 outlines the dissemination strategy of the ESS to be based on two pillars: a data pool on the one hand; and a flexible suite of products and Services on top of the data pool on the other hand.

* *Data pool* –This is a pool of European statistics in a machine-readable open data format, supporting the ESS’ move towards an open data philosophy and the promotion of re-use of ESS data. This data pool should be publicly available at any time, any place and for all user groups.
* *Products and Services* –these are tailored Products and Services the ESS intends using based on its own data pool. Products and Services include visualizations, animations, multilingualism, interactive tools, apps, etc.

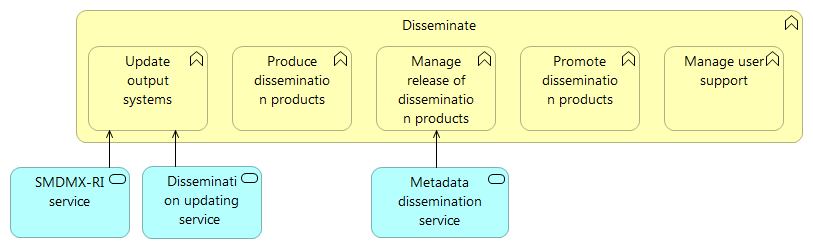
Vision 2020 describes Dissemination as user-friendly, clearly communicating European statistics as a trusted brand for official statistics.

The below to-be state for Dissemination in the SPRA has been produced with the aim to facilitate the realization of the dissemination aspects of Vision 2020 as described above, supporting a long term outlook to the dissemination of European statistics.

The suggested architecture describes a to-be state architecture which:

* Supports static content and dynamic content access methods and supports information tailored to specific user group needs (in particular “power users”):
  + - Static: disseminated information labelled as static is an information set that will be delivered equally to all users or systems;
    - Dynamic: disseminated information labelled as dynamic is an information set that will be a custom subset of data responding to a specific query of a user or a system.
    - Ad hoc requests: requests that cannot be satisfied with dynamic products.
* Supports more voluminous data sets:
  + - Access to new data sources (e.g. admin & big data): New sources of data are accessible, thereby extending the possibilities of creating new statistical products on demand;
* Set up of an ESS Data pool:
  + - The ESS Data pool will be defining and implementing a reference platform and standard Services that allow interoperability among different organizations involving the exchange/sharing of data and Metadata between their respective information systems addressing specific issues of data warehousing including data storage, data integration and a common IT network for data exchange;
    - Data will be easier to reuse (“collect or publish once, use many times”) and more efficient governance and maintenance will be enabled. Over time a reduction of dissemination costs for European statistics is expected;
    - An ESS Data pool should also allow integration of data from individual ESS members with non-ESS ("external") sources for example including less structured information available on the Web.
* Support open data standards
  1. ESS relevant services

Figure 6 visualizes the to-be state for Dissemination in the SPRA.



1. ESS Statistical Production Reference Architecture – Disseminate phase

Table 6 contains a brief description of the Services identified in the SPRA for Disseminate.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **GSBPM Phase** | **GSBPM sub-phase** | **Name** | **Description** | **Category** |
| 7. Disseminate | Manage release of dissemination products | Metadata dissemination | Service that provides dissemination of information on the source, concept, definition, methodology and details on collection, processing, interpretation and dissemination as well as availability of data. | Application service |
| 7. Disseminate | Update output systems | Dissemination updating service | Service which loads data to the dissemination environment | Application service |
| 7. Disseminate | Update output systems | SDMX-RI | Service translating relational data to SDMX for dissemination | Application service |

1. Services in the SPRA: Disseminate
   1. Design Principles

The next table summarizes the Design Principles for Dissemination in the SPRA.

|  |  |
| --- | --- |
| **Name** | **Adherence to Standards for data and Metadata Exchange** |
| **Statement** | Data and Metadata standards are implemented among ESS participants |
| **Rationale** | Standardized file formats for data and Metadata and standardized contents of these files are the pre-condition for the automated production, processing and exchange of data and Metadata files between national and international statistical organizations and registering products for example in (pan)European open data portals. Standardization thus leads to more efficient processes for exchange and sharing of data and Metadata |
| **Implication** | ESS should use SDMX supplemented by other standards for exchanging data and metadata. |
|  | |
| **Name** | **Use of data warehousing** |
| **Statement** | Data which needs to be disseminated is registered, stored and updated in standardized form in a (unified) data warehouse |
| **Rationale** | The data warehouse is the most appropriate solution for combining data extracted from various source systems (including external sources) in a single architecture.  It is an enabler for data analysis, allowing recording and distinguishing time variant versions of the same data point or selecting the appropriate level of data grain depending on the business need. |
| **Implication** | ESS members should gradually replace stovepiped data stores with a data warehouse approach |
|  | |
| **Name** | **Control only once** |
| **Statement** | If possible, Dissemination data should only be checked and made ready for Dissemination once in the European production chain. |
| **Rationale** | Shared methodologies and quality controls should enable data to be made ready for Dissemination only once to save possible duplicated efforts in NSIs and DG ESTAT. |
| **Implication** | The validations and checks being performed on data must be transparent across the production chain. |
|  | |
| **Name** | **Machine to machine Web Service based access to Dissemination components** |
| **Statement** | Dissemination Services can be accessed by another ESS partner’s application (such as a generally available, ubiquitous protocols and transports). |
| **Rationale** | Web Services provide the following benefits:   * Interoperability – due to standards-based communications methods; * Reusability – through the closest to, possibly zero-coding deployment of Services; * Deployability – deployment over standard internet technologies including over fire wall to servers running on the Internet. |
| **Implication** | Access to statistical output should be made available to the public supplementing acces through a data browser. |
|  | |
| **Name** | **Aligned branding of European official statistical products** |
| **Statement** | The online presence of ESS partners is aligned through harmonized look & feel |
| **Rationale** | A unified visual identity and consolidation of access points to official European statistics (potential “Single entry point to European Statistics”) helps to better communicate the joint “brand” of ESS partners. Benefits are among others a more distinguishing positioning as a provider of statistics, and increased user trust. |
| **Implication** | ESS should develop joint guidelines for branding of European official statistics. |
|  | |
| **Name** | **User-friendly Dissemination** |
| **Statement** | The European Dissemination platform makes available best-in class Dissemination Services to end users |
| **Rationale** | The collaboration among ESS partners allows deploying the most advanced Dissemination solutions available in the community (e.g. through re-using functionalities/components from leading NSIs). Users have access to larger quantities of data, and can possible combine them directly with (international coverage) third party data. |
| **Implication** | ESS member should exchange experience and knowledge on user experience and get inspired from other settings publishing data sets. |

1. Design Principles for Dissemination in the ESS Statistical Production Reference Architecture

# Attachments

## Implementations of Building Blocks

The table below provides an initial set of implementations of Building Blocks available to ESS partners for reuse. In future versions, a more detailed list should be investigated and a repository for these should be set up.

|  |  |
| --- | --- |
| ESS EARF BBs | Available implementations |
| Statistical processing | SDMX-RI (mapping)  Validation |
| Data collection |  |
| Collaboration | CIRCABC |
| Data exploration and analysis |  |
| Consumer management and tracking |  |
| Design of collection instruments |  |
| ESS Data exchange | ESDEN, SIMSTAT, Census Hub, SDMX-RI |
| Dissemination | DISSEMINATION |
| Metadata Management | Census Hub/SIMSTAT based pull mechanism |
| Quality assessment |  |
| Dissemination Data Storage |  |
| Primary Data Storage | Census Hub/SIMSTAT based pull mechanism |
| IT Security | ESDEN, SIMSTAT, Census Hub, SDMX-RI |
| Identity and access management (IAM) | ECAS |
| Catalogue of reusable solutions and standards | Interactive Profiling Tool (IPT)  RAMON Standards Catalogue |
| Data supplier registry |  |
| Process Orchestrator | SERV  Census Hub/SIMSTAT |

Table 3 Initial proposal of standard IT solution Building Blocks

## Illustration of consolidation approaches

What follow is an Illustration of what the four consolidation approaches: shared, replicated, interoperable, and autonomous could mean for implementing a Validation service, an SDMX converter or a Metadata management solution.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Validation** | *Shared* | *Replicated* | *Interoperable* | *Autonomous* |
| The business function validation performs structural validation of data sets for statistics.  The business function involves multiple specific validations but it takes three inputs:  A data set  Metadata for the data set (e.g. DSD)  Additional parameters  and produces one output:  - Validation report. | Here validation is implemented as a shared service.  The validation is performed using a central operational instance and is accessed in a uniform manner for different users/systems.  Protocols are defined and shared for accessing the validation service. | Here validation is a centrally developed and maintained software implementations implementing the same logic as the shared service.  The software implementation is then hosted in the IT environments of different ESS members.  The software implementation may be customized to fit the needs of the specific ESS members but must maintain a common core to ensure that the system can be maintained and further developed centrally. | Here validation is multiple decentrally developed software implementations of the validation business function.  The software implementations are developed according to shared definitions of data set and metadata, which ensures interoperability and possible easy transfer of validation algorithms from one implementation to the other. | Here validation is implemented independently within each ESS member.  ESS members will share validation algorithms and exchange knowledge about validation |
| **SDMX converter** | *Shared* | *Replicated* | *Interoperable* | *Autonomous* |
| The SDMX converter is a business function ensuring that the following can be performed on a data set:  A data set in another format (e.g. CSV) is converted into valid SDMX.  An SDMX data set is converted to another variant of SDMX.  An SDMX data set is converted into another format (e.g. CSV) | The SDMX converter is developed, maintained and operated as a central services, which is available to all ESS members.  The ESS member calls the single available instance of the ESS SDMX converter through a user/system interface using a common protocol. | The SDMX converter is developed and maintained centrally in the ESS.  The SDMX converter is installed and operated within each ESS member using the same software and the common protocol for access. | The ESS partner utilizes his own converter that however is conform with the SDMX community’s technical standards (including the Information Model) and statistical guidelines. | Not relevant |
| **Metadata management** | *Shared* | *Replicated* | *Interoperable* | *Autonomous* |
| The business function Metadata Management is managing metadata both horizontally and vertically. Som across different statistical domains and across different phases of production, metadata is managed here. The function will support the management (i.e. creating, maintaining, retriring) of metadata. It will also provide access to metadata for production systems (e.g. serving DSD’s for use in validation) | The Metadata Management function is performed in a centrally developed, maintained and operated solution that is used by everyone in the ESS.  The ESS member access the single available instance metadata management solution through a user/system interface. | The Metadata Management function is implemented as a centrally developed and maintained software implementation.  This software implementation is customized and hosted in the individual ESS member.  Shared metadata such as code lists, DSD’s, etc. will be replicated between the operational instances in the ESS members. | The ESS member implements a metadata management solution, which fits their local requirements.  All solutions agree on a common set of metadata that are exchanged using a common protocol. | The ESS member manages metadata in locally developed solutions.  All solutions agree on a common set of metadata to enable the exchange of statistical data.  The specific exchange of metadata is agreed bi-lateraly between user and the one responsible for managing the metadata. |

## Consolidation approaches and their impact on interoperability requirements

The EIRA identifies four layers of interoperability, which require consideration when designing and assessing the interoperability of a unit of functionality.

|  |  |
| --- | --- |
| **Interoperability layers** | **Shared scenario**  **“The functionality realized in the software solution executes identical business processes through a single solution instance, possibly but not necessarily utilizing information that is shared across all ESS participants.”** |
| Legal interoperability | * Legislation can be a supporting instrument, in the sense that it enforces closer collaboration between ESS partners and provides legal certainty on how this collaboration needs to be implemented |
| Organizational Interoperability | * Jointly defined business processes * Service contracts between solution provider and users * Single sourcing |
| Semantic Interoperability | * Solution manages semantic and syntactic IOP |
| Technical Interoperability | * Shared solution manages uniform technical IOP with other building blocks * Single instance * common requirements for end-user technology |

|  |  |
| --- | --- |
| **Interoperability layers** | **Replicated**  **“The functionality realized in the software solution supports the execution of similar business processes through duplication (possibly with configuration) by the ESS partner”** |
| Legal interoperability | * Legislation can be a supporting instrument, in the sense that it enforces closer collaboration between ESS partners and provides legal certainty on how this collaboration needs to be implemented |
| Organizational Interoperability | * Similar businesses processes * Coordinated definition of the solution * Software maintenance contracts between providers and users |
| Semantic Interoperability | * Prescriptive syntax: common structure or data format * Prescriptive semantic standards: common data model |
| Technical Interoperability | * Identical (or closely identical) solution * Conformance to a common technological infrastructure * Common requirements for end-user technology |

|  |  |
| --- | --- |
| **Interoperability layers** | **Interoperable**  **“The functionality is realized through an individual software solution of the ESS partner which nevertheless effectively exchanges information and operates together with others in a coordinated way”** |
| Legal interoperability | * A more light-weight approach (such as Harmonization through Recommendations) can be used instead of Legislation |
| Organizational Interoperability | * Responsibility for the required capability remains with the ESS partner * Coordinated processes * Interoperability agreements such as existing data exchange agreements |
| Semantic Interoperability | * Can leverage (statistical domain specific) industry standards * Messaging middleware manages syntactic IOP * Application manages semantic IOP |
| Technical Interoperability | * Conformance to a common transport protocol |

Any questions regarding this Report  
should be addressed to:

Jean-Marc Museux

Enterprise Architect

Eurostat

Joseph Bech building

5 Rue Alphonse Weicker

L-2721 Luxembourg

Email: [Jean-Marc.Museux@ec.europa.eu](mailto:Jean-Marc.Museux@ec.europa.eu)

1. The table is an illustration of applying the technique of “Implementation Factor Assessment and Deduction matrix” as defined in TOGAF. The Implementation Factor Assessment and Deduction matrix is used to document factors impacting an architecture implementation, their descriptions and the deductions that motivate the actions or constraints to be taken into consideration in the Information Systems architecture. [↑](#footnote-ref-2)
2. See http://pubs.opengroup.org/architecture/archimate2-doc/chap04.html#\_Toc371945184 [↑](#footnote-ref-3)
3. Adopted from “Framework of metadata requirements and roles in the S-DWH” a report from ESS – NET ON MICRO DATA LINKING AND DATA WAREHOUSING IN PRODUCTION OF BUSINESS STATISTICS [↑](#footnote-ref-4)
4. Each architecture Design Principle is presented through Name, Statement, Rationale and Implications as stipulated in the ESS EARF. [↑](#footnote-ref-5)