ESS Statistical Production Reference Architecture

31-08-2015
Table of Contents

1.0 Introduction .................................................................................................................. 2
  1.1 Objective of the document .......................................................................................... 2
  1.2 Scope and related documents ................................................................................. 2
    1.2.1 GSBPM phases covered .................................................................................... 2
    1.2.2 Relation with the ESS EARF ......................................................................... 2
    1.2.3 Level of abstraction ......................................................................................... 3
  1.3 Design drivers for the SPRA ................................................................................. 4
  1.4 Key concepts in the SPRA .................................................................................... 5
    1.4.1 Building Blocks ............................................................................................... 5
    1.4.2 Services ........................................................................................................... 6
    1.4.3 Information Objects ....................................................................................... 6
    1.4.4 Consolidation approaches and scenarios ....................................................... 6
    1.4.5 Service orchestration ..................................................................................... 7
    1.4.6 Metadata Management .................................................................................. 7

2.0 Data Collection in the SPRA .................................................................................. 9
  2.1 Context .................................................................................................................... 9
  2.2 The to-be state of Data Collection in the ESS ....................................................... 10
    2.2.1 Collect phase .................................................................................................. 12
    2.2.2 Design phase .................................................................................................. 14
    2.2.3 Build phase .................................................................................................... 15
  2.3 Design Principles .................................................................................................... 17
  2.4 ESS Member implementations - Scenarios for realizing the to-be state ............... 18
    2.4.1 Scenario 1 - Interoperable .......................................................................... 19
    2.4.2 Scenario 2 - Replicated .............................................................................. 21
    2.4.3 Scenario 3 - Shared Services scenario ....................................................... 22

3.0 Process & Analyse in the SPRA .......................................................................... 24
  3.1 Context ................................................................................................................... 24
  3.2 The to-be state of Process & Analyse in the SPRA ............................................... 25
    3.2.1 Phases, Services and ESS EARF BBs ......................................................... 25
  3.3 Design Principles .................................................................................................. 29
  3.4 ESS Member implementations - Scenarios for realizing the to-be state ............... 31
    3.4.1 Scenario 1 – Interoperable .......................................................................... 32
    3.4.2 Scenario 2- Replicated .............................................................................. 33
    3.4.3 Scenario 3- Shared Services scenario ....................................................... 33

4.0 Dissemination in the SPRA .................................................................................. 35
4.1 Context..................................................................................................................................................35
4.2 The to-be state of Dissemination in the ESS ......................................................................................36
  4.2.1 Disseminate phase..........................................................................................................................38
4.3 Design Principles ..................................................................................................................................41
4.4 ESS Member implementations - Scenarios for realizing the to-be state.............................................42
  4.4.1 Scenario 1- Interoperable...............................................................................................................43
  4.4.2 Scenario 2- Replicated..................................................................................................................44
  4.4.3 Scenario 3- Shared .........................................................................................................................45
**Version information:**

<table>
<thead>
<tr>
<th>Version 0.4</th>
<th>31/08/2015 – for public release</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1.0 Introduction

1.1 Objective of the document

This document details the ESS Statistical Production Reference Architecture (SPRA). This reference architecture represents a description of the to-be state architecture based on various key artefacts of the ESS EA Reference Framework (ESS EARF). In the current version, the SPRA outlines how the Collect, Process & Analyse, and Disseminate phases and sub-processes of the statistical production process\(^1\) should be supported by IT in the ESS. It does this by describing the Building Blocks and Services necessary to realize the architecture.

The SPRA should help statistical institutions develop IT solutions that follow the overarching models and principles outlined in the ESS EARF as well as the more specific architecture models and principles outlined here. In the end it should enable more efficient, effective and collaborative IT solutions across the ESS.

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.

1.2 Scope and related documents

1.2.1 GSBPM phases covered

This document provides the architecture models and principles for those aspects of the ESS EARF which relate to the GSBPM phases of Collect, Process & Analyse, and Disseminate. In addition, the document also addresses sub-processes of other GSBPM phases (such as Design or Build) where these inherently relate to collecting, processing, analysing or disseminating statistical information. In future versions, additional GSBPM phases may be covered, for example those associated to planning and design to obtain an even more complete, end-to-end perspective on the statistical production process.

1.2.2 Relation with the ESS EARF

This document builds on the ESS EARF currently in version 0.6. Figure 1 schematically maps the SPRA to the ESS EARF artefacts.

\(^1\) As of GSBPM.
The current SPRA provides the following viewpoints in addition to the ESS EARF:

- It models a number of Business Services needed to support the sub-processes in the GSBPM model (see Figure 4, Figure 14, Figure 8, Figure 9, Figure 10);
- It models the relation between Building Blocks and Business services i.e. in which ESS EARF Building Blocks the different Business Services are realized (see Figure 4, Figure 14, Figure 8, Figure 9, Figure 10);
- It defines a number of principles that are specific to the GSBPM phase modeled (see Table 3, Table 6, Table 5);
- It defines possible scenarios for how the architecture could be realized by the ESS members (see Sections 2.4, 3.4, 4.4).

Making reference to phases, Building Blocks and Business Services, the SPRA builds on the ESS EARF and guides the development of Information System Architectures. An Information System Architecture would typically model one or more interconnecting ESS EARF Building Blocks and prepare for the procurement and development of solutions implementing the Information System Architecture.

1.2.3 Level of abstraction

Typically, an Information System Architecture distinguishes three levels of abstraction: logical, technical, and physical.

---

2 For a full list of the ESS EARF Building Blocks, see annex.
Logical | A list of the provided Business Services and functionalities broken down in Service-Oriented Architecture (SOA) Services.

Technical | The specification of the underlying Services that are needed to cater for the Business Services defined in the logical layer. In addition, what are the functional components and products used to provide the functionalities? This could be a list of Open Source Software (OSS), Common-off-the-shelf (COTS) or new functionalities implemented and exposed as Services through a web application server. This includes components such as workflow management systems, data storage, etc.

Physical | The physical architecture specifies the physical operational implementation of the solution, including server instances, network zones, etc. These are the physical units needed to serve the non-functional requirements (scalability, speed, availability, security, integrity, etc.).

| **Table 1. Three levels of abstraction of an Information System Architecture** |

The SPRA addresses and guides the logical layer.

### 1.3 Design drivers for the SPRA

The architecture described in this document has been designed in respect of NSI specific systems, investment programmes and the principle of subsidiarity. It

- 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 in the SPRA, broken down into value-creating elements;
- Enables ESS members to utilize investments by providing different scenarios for how individual NSIs will implement the Vision;
- 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.³

---
³ 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.
<table>
<thead>
<tr>
<th>Driver</th>
<th>Impact on SPRA</th>
</tr>
</thead>
</table>
| 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 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. |

Table 2. ESS Statistical Production Reference Architecture: Driver-Impact matrix

1.4 Key concepts in the SPRA

The SPRA uses and/or introduces a set of key concepts: Building Blocks, Services, Information Objects as well as Service Orchestration and Metadata Management; the latter two ensuring that statistical production processes are orchestrated properly. These are displayed in Figure 2, in an overview graph.

1.4.1 Building Blocks

The ESS EARF defines a set of Building Blocks (BBs) which represent (potentially re-usable) components of (IT) capability. Conceptually speaking, Building Blocks comprise one or several Business Services. For further details, please see the BB chapter in the ESS EARF.
1.4.2 Services

The SPRA model provides a bridge between the Business, Application and Information Architectures and the implementation of the Vision 2020 through unique solutions (Information System Architectures), as conceptualized in Figure 1 above.

The bridge between both worlds of abstraction is facilitated by the introduction of 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 basic Services in turn can be composed to form aggregated or Business Services. Doing so, the SOA will provide the Business Services which are identified in the SPRA by analysing the processes and activities of the business, and complementing the (legacy) system view explained at the beginning of this paragraph.

1.4.3 Information Objects

The ESS EARF makes an attempt to conceptualize how information is input and output to statistical business processes, Building Blocks and Services, using GSIM Information Objects to model these in- and outputs. This perspective is not yet developed in the SPRA. It may be included in a next iteration.

1.4.4 Consolidation approaches and scenarios

The ESS EARF also defines possible consolidation approaches for the Building Blocks, Services and Information Objects in the SPRA:

- Autonomous
- Interoperable
- Replicated or
- Shared.

The political & policy ambition level within the ESS will determine the extent of consolidation and collaboration of IT across ESS partners.

Based on the proposed Consolidation approaches, the SPRA defines a number of possible implementation scenarios for the three detailed architectures: Collect, Process & Analyze and Disseminate.

The three scenarios focus on the governance of Services and the key SPRA BBs as these require clarification before investing in developing Services and the BBs.
It needs to be noted that the SPRA supports a wide range of scenarios for sharing that are not all covered in the three model scenarios, in the sense that an ESS member can combine scenarios, depending on its investment plans and priorities.

1.4.5 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 a statistical production process. From this perspective a statistical production process will consist of a number of (GSBPM) phases and steps (see Figure 2) where Services are invoked using a data set managed by the data management component (e.g. a data storage) as well as Metadata (parameters) 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 based on 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.

![Conceptual model of key SPRA elements](image)

Figure 2. Conceptual model of key SPRA elements

1.4.6 Metadata Management

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 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.

All Metadata should be available through a Metadata Management system 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.
2.0 Data Collection in the SPRA

2.1 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.

*We will invest in new IT tools and methodological development. Technology is a key ingredient in our strategy. We will invest in IT infrastructure related to Data Collection, transfer and storage.*\(^4\)

*We will continue to improve existing Data Collection methods. While there are quite a number of emerging new sources that have potential for our work, there is still room for improvement with more traditional sources. We will continue our efforts to improve these sources, in particular with respect to non-response approaches, efficient mixed-mode Data Collection strategies and exploitation of administrative sources.*\(^5\)

Reflecting the Vision 2020’s mention of new Data Collection sources, Data Collection in the SPRA puts specific emphasis on the collection from administrative and new data sources, in addition to the traditional, survey-focused view as conceptually illustrated in Figure 3.

---

\(^4\) Excerpt from Vision 2020.

\(^5\) Ibid.
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.

2.2 The to-be state of Data Collection in the ESS

Figure 4 visualizes the to-be state for the ESS Data Collection in the SPRA. It builds on the idea of making a set of ESS Data Collection Services available to ESS partners, leveraging on the ESS EARF Building Blocks.

The identification of these “Business Services” has been done by:

- Examining the activities of the GSBPM phase Collect as well as select sub-processes of the Design & Build phases relating to Data Collection;
- Analysing ESS-specific literature and documentation relevant to Data Collection.

Figure 4 follows the phases and sub-processes of the GSBPM and details for each of them, which Services and/or Building Blocks will be deployed and how they interrelate. The Collect phase is the one that is being focused on, hence it is bolded out.

For modelling purposes, Data Collection is considered as a similar phase across NSIs and DG ESTAT, even though DG ESTAT is principally collecting data from NSIs. The aspects of Data Collection related to the execution of surveys will predominantly be of relevance to NSIs.
Figure 4. ESS Statistical Production Reference Architecture - Data Collection
2.2.1 Collect phase

This phase collects all necessary data, from differing sources (including extractions from administrative databases), and loads them into the appropriate data environment. The Collect phase is broken down into four sub-processes. Data Collection in the SPRA details all four.

Select sample

This is sub-process 4.1 of the Collect phase. It builds on the Design phase where the sampling methodology has been predefined. It includes quality assurance, approval and maintenance of the sample as well as the coordination of samples between instances of the same statistical business process (for example to manage overlap or rotation), and between different processes using a common frame or register (for example to spread response burden). The sampling aspect of this sub-process is usually not relevant for processes based entirely on the use of pre-existing data sources (e.g. administrative data) as such processes generally create frames from the available data and then follow a census approach.

The following Services are introduced in this sub-process:

- “Create sample”: creating samples & frames which are coordinated across statistical domains and ESS participants;

The following ESS EARF BBs are introduced in this sub-process:

- Data supplier registry: storing all respondent-related information.

Set up collection

This is sub-process 4.2 of the Collect phase. This sub-process ensures that the people, processes and technology are ready to collect the required data. It takes place over a period of time, as it includes the strategy, planning and training activities in preparation for the specific instance of the statistical business process.

The following Services are introduced in this sub-process:

- “Train”: online training for collection staff across the ESS;
- “Generate survey”: generating the survey in the required format (e.g. ODF, WebForms, CAPI);
  - “Prefill survey”: prefilling of questionnaires where possible, thus reducing respondent burden;

The following ESS EARF BBs are introduced in this sub-process:

- Primary data storage: for pre-filling questionnaires with existing data.

---

6 For a full list of the ESS EARF Building Blocks, see annex.
Run collection

This is sub-process 4.3 of the Collect phase. This sub-process is where the Data Collection is implemented, with the different collection instruments being used to collect the data. It also includes provider management. For administrative data, this process only consists of pulling or pushing the data from the provider source until the collection meets its target.

The following Services are introduced in this sub-process:

- “Manage respondents & providers”: this Service ensures the management of initial contacts & reminders, recording contact points, response rates etc. and maintaining a positive client relationship both with respondents (for surveys) and providers (for administrative data);
- “Manage Data collection”: this Service manages survey & administrative data collection deployment and captures the metrics of Data Collection (para-data assembled in dashboards for example);
- “Plan Data collection”: this Service provides crucial input into the management of data collection by planning ahead and anticipating & mitigating Data collection peaks;

The following ESS EARF BBs are introduced in this sub-process:

- Data collection: enabling the actual Data collection;
- IT Security: ensuring the security of data to be collected;
- Metadata Management: for the capturing of para-data.

Finalize collection

This is sub-process 4.4 of the Collect phase. This sub-process includes loading the collected data and Metadata into a suitable electronic environment. It may include automatic data take-on, for example using optical character recognition tools to extract data from paper questionnaires, or converting the formats of data files received from other organizations. Relevant supporting documents are archived.

The following Services are introduced in this sub-process:

- “Extract data”: this Service extracts data from the data source;
  - “Take on”: provides for automatic data take-on to extract data from paper questionnaires;
  - “Transform”: ensures file format conversion; this could be a generic, re-usable Service across phases of the statistical production chain;
- “Load”: loads the data; this again could be a generic, re-usable Service across phases of the statistical production chain;
- “Report data and exchange”: Services used to report the collected data back to either DG ESTAT or international institutions;

For a full list of the ESS EARF Building Blocks, see annex.
“Store data”: stores the data;
“Store Metadata”: stores Metadata;
“Archive”: archives relevant documents relating to the Data Collection;

The following ESS EARF BBs are introduced in this sub-process:

ESS Data exchange: facilitating the sending and receiving sets of data among ESS members.

2.2.2 Design phase

This phase describes the development and design activities needed to define the statistical outputs, concepts, methodologies, as well as collection instruments and operational processes. The phase is highly important as it pre-defines and anticipates all parameters of the subsequent, actual Data collection. It thus inherently links to the Collect phase of the statistical production process.

Out of the Design phase, one sub-process is documented in Data Collection of the SPRA:

Design Data Collection methodology

Design Collection

This is sub-process 2.3 of the Design phase. It includes the design of questions and response templates, typically enabled by question libraries (to facilitate the re-use of questions and attributes) and questionnaire tools (to enable the quick and easy compilation of questions).

Following the concepts introduced in Figure 3, the Data Collection methodology will vary depending on the type of data to be collected: for surveys, instruments such as computer assisted interviewing or paper questionnaires will be used; for the collection of administrative data, interfaces and data integration techniques will be leveraged upon.

Currently, new approaches to survey design are being tested by statistical institutes to achieve greater process automation\(^8\). Subject matter experts are provided a survey design interface which allows them to specify the collection instruments in business terminology. “At the back-end”, this work is automatically translated into specifications in a harmonized expression language (e.g. DDI) which is transformed into code suitable to generate the questionnaire in electronic format (e.g. based on Blaise\(^9\)). Manual programming is limited to the strict minimum. Code programmers can polish and tailor the code to meet more advanced business requirements where relevant.

The following Services are introduced in this sub-process:

“Manage requirements”: helps document, analyze, prioritize and agree on data collection requirements amongst ESS partners, controls changes to requirements and communicates them to stakeholders, maps data requirements to available data and determines the best type of source to use: survey, administrative data, other;

\(^8\) Piloted by ABS and INSEE for example.
\(^9\) Based on the Statistics Netherlands Blaise Internet product.
“Design Data collection instruments”, including:
- “Manage templates”: manages (static) survey templates available to the ESS community;
- “Manage reference Metadata”: ensuring the accuracy of references for Metadata driven questionnaires, and registering and storing specifications for collection instruments;
- Manage collection variables & concept definitions: creates and/or reuses variable definitions, identifies or defines concepts and maps these to variables, matches concepts and variables across ESS partners and across data provider & data requestor;
- Define frames & samples: coordinates frames amongst surveys in an effort to reduce response burden, reuses frames, and prepares samples.

The following **ESS EARF BBs** are introduced in this sub-process:
- Catalogue of reusable solutions and standards: providing access to concepts and methods for Data Collection;
- Collaboration: enabling interaction (on the design of collection instruments) amongst ESS partners;
- Design of collection instruments: to access and use the survey templates and/or design functionalities;
- Metadata Management: the direct connection of collection instruments to the statistical Metadata system, ensuring that Metadata can more easily be captured in the subsequent sub-processes and phases, from Collect to Disseminate;
- Process Orchestrator: BPM-standards based system executing workflows according to pre-defined rules; allowing users, such as reviewers, approvers, and certifiers of collection instruments to interact and execute business processes- in the case of this sub-process the validation and approval of data collection instruments.

### 2.2.3 Build phase

This phase builds and tests the production systems, including systems for data collection, to the point where they are ready for use in the “live” environment. It thus also inherently ties to the Collect phase of the statistical production process.

Out of the Build phase, the following GSBPM sub-processes are included in the architecture view for Data Collection:
- Build data collection instrument
- Test production systems
**Build data collection instrument**

This is sub-process 3.1 of the Build phase. This sub-process describes the activities to build the collection instruments to be used during the Collect phase. The collection instrument is generated based on the design specifications created during the Design phase.

The following Services are introduced in this sub-process:

- “Author survey”: Service which allows users to specify survey instrument Metadata in business terms;
  - “Transform Metadata”: Service transforming the business-authored survey instrument into form-suitable code to generate the survey;
- “Configure Data collection instruments”: configuring the collection instrument to the very specific needs of a data collection round. The Service includes the setting up and maintaining of a data interface where administrative data needs to be retrieved. This Service is deployed either by each ESS partner individually or by one/a team of ESS partners charged with setting up an instance of a Data Collection (e.g. in a specific domain);
  - “Implement checks”: build appropriate controls into the survey;
  - Configure workflow: ensure the Data collection is executed in a way that activities are properly orchestrated and executed in the right order;
- “Manage data collection instruments”: managing the lifecycle of data collection instruments from strategy, design and transition to operation and continual improvement\(^\text{10}\);
  - “Share data collection instruments”: managing the effective sharing of data collection instruments amongst ESS partners, including Service Level Agreements (SLAs), governance and licensing agreements.

**Test production systems**

This is sub-process 3.4 of the Build phase. The testing of production systems is concerned with the testing of computer systems and tools. In the case of Data Collection, this would include the testing of components & their interaction as well as the actual roll out of a small scale Data Collection to test the instruments.

The following Services are introduced in this sub-process:

- “Test Data collection”: allowing for small scale Data Collections simulating “live” conditions, including the exploration of Data collections specific to administrative data;
- “Certify”: certifying Data Collection Services for sharing within the ESS. Certification may examine areas like security, governance, technological and semantic interoperability and availability.

\(^{10}\) Example following ITIL V.3 lifecycle
2.3 Design Principles

The next table summarizes the Design Principles for the to-be state of Data Collection in the SPRA.\textsuperscript{11}

<table>
<thead>
<tr>
<th>Name</th>
<th>Standard definition and description language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statement</td>
<td>Data collection instruments are designed and defined using a standard definition and description language.</td>
</tr>
<tr>
<td>Explanation</td>
<td>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.</td>
</tr>
<tr>
<td>Rationale</td>
<td>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.</td>
</tr>
<tr>
<td>Implications</td>
<td>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.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>Unified Metadata system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statement</td>
<td>Metadata from Data Collection is stored in the unified Metadata system.</td>
</tr>
<tr>
<td>Explanation</td>
<td>The Metadata produced during the design and execution of data collection instruments is captured in the unified Metadata system.</td>
</tr>
<tr>
<td>Rationale</td>
<td>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.</td>
</tr>
<tr>
<td>Implications</td>
<td>Metadata is captured in the Data Collection systems in a format enabling the storage and re-use of Metadata in the unified Metadata system.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>Re-use of existing instruments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statement</td>
<td>Data collection instruments are built reusing already existing instruments where possible.</td>
</tr>
<tr>
<td>Explanation</td>
<td>Data collection instruments should - when possible - reuse or build on existing and proven Data Collection instruments, including surveys, administrative data sources and other.</td>
</tr>
<tr>
<td>Rationale</td>
<td>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.</td>
</tr>
<tr>
<td>Implications</td>
<td>Design of Data collection instruments is never done from scratch. It starts with analysing existing collection instruments, including those from other statistical domains.</td>
</tr>
</tbody>
</table>

Table 3. Design Principles for Data Collection in the ESS Statistical Production Reference Architecture

\textsuperscript{11} Each architecture Design Principle is presented through Name, Statement, Rationale and Implications as stipulated in the ESS EARF.
2.4 ESS Member implementations - Scenarios for realizing the to-be state

Data Collection in the SPRA acts like a framework and provides guidance for developing Information System Architectures impacting the whole of the statistical production chain.

There are three scenarios for NSIs to realize Data Collection as per the SPRA:

- **Scenario 1**: Data is collected via NSI specific data systems. The NSI can make use of some shared ESS Data Collection Services;
- **Scenario 2**: The NSI has replicated the ESS Data Collection Services;
- **Scenario 3**: The NSI integrates the ESS Data Collection Services.
2.4.1 Scenario 1 - Interoperable

This scenario can be visualized as follows:

Figure 5. Interoperable scenario for Data Collection in the ESS Statistical Production Reference Architecture

In this scenario, data is collected via the ESS member’s own data systems which are either autonomous or have a varying degree of interoperability.

In addition, the ESS member can make use of (some) replica of ESS shared services for designing data collection and collecting data in its local environment.
For harmonisation reasons, there is already some value to standardise data collection metadata at ESS level with the obvious benefit of using a local replication of an ESS metadata management system.

Pre-condition for this is that the data collection architecture of the NSI is implemented in an interoperable manner, and/or is facilitated by suitable interoperability middleware. In all cases, the different modularization of statistical Services at the NSI (i.e. different approaches to scoping and generalization) is likely to limit reuse of Services amongst ESS partners.

The NSI will also be able to provide access to select ESS Data Collection Services for replication by other ESS members and thus contribute to the wider ESS community.
2.4.2 Scenario 2 - Replicated

This scenario can be visualized as follows:

![Diagram of Scenario 2 - Replicated]

Figure 6. Replicated scenario for Data Collection in the ESS Statistical Production Reference Architecture

In this scenario, an ESS member replicates the ESS Data Collection Services in a local environment and mixes them with locally developed Data collection Services that are not shared with the ESS community. In addition, the ESS member can make use of shared Services for designing Data collection instruments. This levelling between sharing and replicating reflects the greater likelihood to collaborate in the Design phase than in the Collect phase which involves
the collection and handling of sensitive (micro)data. Microdata are stored in local interoperable instance.

Overall, this scenario implies that the ESS member implements a local Service catalogue for both the replicated and the local Data collection Services.

2.4.3 Scenario 3 - Shared Services scenario

This scenario can be visualized as follows:

![Diagram of Shared Services scenario for Data Collection in the ESS Statistical Production Reference Architecture](image-url)

**Figure 7.** Shared Services scenario for Data Collection in the ESS Statistical Production Reference Architecture
The scenario implies that the ESS member integrates and consumes ESS Data Collection Services on a local replica of a dissemination platform and leverages on a shared ESS Metadata Management system to create, maintain and use its Data Collection instruments. The Data Collection Services are thus shared & central amongst the ESS community whilst still allowing for customization and tailor making to local Data Collection requirements. Microdata are stored locally in a standardised form on a local replica of an ESS platform.
3.0 Process & Analyse in the SPRA

3.1 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.
3.2 The to-be state of Process & Analyse in the SPRA

3.2.1 Phases, Services and ESS EARF BBs

The next series of figures visualize the to-be state for Process & Analyse in the SPRA. More specifically, Figure 14 to Figure 16 show which Business Services are needed to support a particular phase closely related to the statistical production process. All the Services shown are used during the execution of these various phases. Some Services, such as data collection and data integration, are aggregated Services, which can be further decomposed into various other, more basic Services.

In order for the Services to realize their functionalities, they have to “call” or make use of the underlying ESS EARF BBs, as represented by the lower layer of the SPRA model. The ESS EARF BBs, on their turn, might need to make use or “call” underlying systems, databases or data warehouses.
Figure 8. ESS Statistical Production Reference Architecture – Process phase (5.1–5.3)
Figure 9.  ESS Statistical Production Reference Architecture – Process phase (5.4 – 5.7)
Table 4 contains a brief description of the Services identified in the SPRA for Process and Analyze. “Higher level Services” i.e. Services which aggregate other Services are displayed in italic font.

<table>
<thead>
<tr>
<th>Service name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data integration Service</td>
<td>Services which ensures the integration of data by aggregating the Match/record and Anonymize data Services</td>
</tr>
<tr>
<td>Match/record</td>
<td>Matches, records and links to structural Metadata and Metadata models and prioritizes data sources for variables</td>
</tr>
<tr>
<td>Anonymize data</td>
<td>Anonymizes sensitive data</td>
</tr>
<tr>
<td>Data loading</td>
<td>Loads statistical data into the production system</td>
</tr>
<tr>
<td>Classify &amp; code</td>
<td>Provides different classifications and coding of data including geographical coding</td>
</tr>
<tr>
<td>Data review and visualization</td>
<td>Reviews and visualizes data for analysis</td>
</tr>
<tr>
<td>Remove duplicates</td>
<td>Checks for and removes duplicate records</td>
</tr>
<tr>
<td>Data inspection and editing</td>
<td>Provides visualizations of data for inspection and supports different means of editing of data</td>
</tr>
<tr>
<td>Validation</td>
<td>Provides validations of data at different levels</td>
</tr>
<tr>
<td>Structural validation</td>
<td>Check that the data is in the right format</td>
</tr>
<tr>
<td>Content validation</td>
<td>Provides various content validation mechanisms (checksums, identification of outliers, etc.)</td>
</tr>
<tr>
<td>Imputation</td>
<td>Identifies missing data values and performs imputation using different imputation methods</td>
</tr>
<tr>
<td>Adjustments and weightings</td>
<td>Service which derives new units and adjusts &amp; provides weightings on the data, aggregating other Services</td>
</tr>
<tr>
<td>Seasonal adjustments</td>
<td>Performs seasonal adjustments to data</td>
</tr>
<tr>
<td>Other adjustments</td>
<td>Adjusts for e.g. coverage, inflation, coherence, geography</td>
</tr>
<tr>
<td>Time series analysis</td>
<td>Creates, visualizes and analyzes time series</td>
</tr>
<tr>
<td>Aggregation</td>
<td>Creates aggregate statistical data from micro-data</td>
</tr>
<tr>
<td>Finalize output</td>
<td>Finalizes outputs for loading in the dissemination systems</td>
</tr>
<tr>
<td>Check disclosure</td>
<td>Checks and removes values that could identify individuals in the output data set</td>
</tr>
</tbody>
</table>

Table 4. Services in the SPRA: Process & Analyze

### 3.3 Design Principles

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

<table>
<thead>
<tr>
<th>Name</th>
<th>There is a shared statistical process language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statement</td>
<td>The SPRA will be based on a shared process description language such as Business Process Model and Notation (BPMN) 2.0, Web Services Business Process Execution Language (WS-BPEL) 2.0 or XML Process Definition Language (XPDL) 2.1.</td>
</tr>
<tr>
<td>Rationale</td>
<td>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.</td>
</tr>
<tr>
<td>Name</td>
<td>Statement</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Statistical Services are shared</td>
<td>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).</td>
</tr>
<tr>
<td>Services are usable whilst generic</td>
<td>Statistical Services should be generic where feasible.</td>
</tr>
<tr>
<td>New statistical processes are made available for replication</td>
<td>All new statistical production processes should be made available as a replicable process using the shared process language and a set of shared Services.</td>
</tr>
<tr>
<td>Process Metadata is managed by the Process Orchestrator</td>
<td>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.</td>
</tr>
<tr>
<td>Statistical design takes place in a production system sandbox</td>
<td>The development/design of new statistical production will take place in a sandbox implementation of the process designer and Orchestrator component.</td>
</tr>
</tbody>
</table>

Table 5. Design Principles for Process and Analyse in the ESS Statistical Production Reference Architecture
3.4 ESS Member implementations - Scenarios for realizing the to-be state

There are three scenarios for NSIs to implement the SPRA. These scenarios can be supported simultaneously with the architecture and are thus not mutually exclusive:

- **Scenario 1** – In this scenario the ESS member implements a non-ESS Statistical Production Architecture that is interoperable with the ESS Production System enabling it to utilize a selection of the shared statistical Services.

- **Scenario 2** – In this scenario the ESS member implements the SPRA and replicates locally the provided Services and combines them with locally developed Services.

- **Scenario 3** – In this scenario the ESS member implements the full SPRA locally and utilizes the shared statistical Services.

No matter whether an ESS member chooses to implement one or the other scenario, a gradual implementation remains possible. The effort of making the statistical Services available by wrapping\(^{12}\) existing applications or building new ones is significant. Also the design and implementation of the actual statistical processes is a significant change effort. For these reasons the suggested architecture also takes into account the NSI’s specific production systems of those ESS members who cannot find positive business cases for moving the production of specific stable statistics to the new production system.

---

\(^{12}\) Wrapping is an approach whereby existing functionality in a computer program is provided with an interface that makes it accessible via e.g. a web Service interface.
3.4.1 Scenario 1 – Interoperable

This scenario can be visualized as follows:

In this scenario the ESS member chooses to implement its Statistical Production Architecture either based on a commercially available product such as SAS or a system based on e.g. an R implementation that also supports the design and management of processes.

Provided that the Statistical Production Architecture of the ESS member is implemented in an interoperable manner, the NSI will be able to utilize select shared statistical Services. It may also reuse aspects of the production processes designed in the SPRA, provided that a statistical version of a language such as BPEL can be agreed upon. However, the different modularization of statistical Services (i.e. different approaches to scoping and generalization\textsuperscript{13}) in the ESS members’ Statistical Production Architecture is likely to limit this reuse.

The ESS member will also be able to provide standard versions of statistical Services for other ESS members to use and thus contribute to the wider ESS community. All the available resources will be registered in the ESS catalogue.

\textsuperscript{13} The scope of a Service can e.g. differ in terms of offering a specific imputation method vs. multiple methods and differences in generalization will affect the differing context in which it can be used (e.g. the input data structure).
Primary data are stored in a local system.

3.4.2 **Scenario 2 - Replicated**

This scenario can be visualized as follows:

![Diagram](image)

**Figure 12. Replicated scenario for ESS Statistical Production Reference Architecture**

In this scenario, the ESS member uses a statistical process orchestrator replica. It selects from the ESS catalogue replica of the common statistical Services in a local environment and mixes them with locally developed statistical Services. It is more likely that selected shared services focus on statistical process design and process metadata report. This also implies that the ESS member implements a local Service catalogue for both the replicated and local statistical Services. The primary data storage remains local.

This scenario would suit ESS members (for example in the larger European countries) with a higher percentage of locally developed statistical Services and security or performance concerns preventing them from using the shared statistical Services directly.

3.4.3 **Scenario 3 - Shared Services scenario**

This scenario can be visualized as follows:
Figure 13. Shared scenario for ESS Statistical Production Reference Architecture

The scenario implies that the ESS member implements a full replica of the ESS process Orchestrator utilizing the shared statistical Services. The primary data storage is also a replication of the ESS primary data storage implementing a shared data model.

In this scenario the ESS member realizes the maximum benefit of the shared investments.
4.0 Dissemination in the SPRA

4.1 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.
4.2 The to-be state of Dissemination in the ESS

Figure 14 visualizes the to-be state for Dissemination in the SPRA. It builds on the idea of making available a set of ESS Dissemination Services to ESS partners, leveraging on the ESS Statistical Production System and Dissemination Platform.

The identification of these “Business Services” has been done by:

- examining the activities of the GSBPM phase Dissemination as well as select sub-process of the Build phase relating to Dissemination;
- analysing ESS-specific literature and documentation relevant to Dissemination;
- Workshops held with DG ESTAT.

Figure 14 follows the phases and sub-processes of the GSBPM and details for each of them, which Services and/or ESS EARF BBs will be deployed and how they interrelate.

The to-be state represents both Dissemination to end users and reporting of data from NSIs to DG ESTAT.
Figure 14. ESS Statistical Production Reference Architecture – Disseminate phase
4.2.1 Disseminate phase

This phase manages the release of the statistical products to DG ESTAT and other, end users. For statistical output produced regularly, this phase occurs for each iteration. It is made up of five sub-processes, which are generally sequential, from left to right, but can also occur in parallel, and can be iterative. These sub-processes are:

- Update output systems
- Produce dissemination or reporting products
- Manage release of dissemination or reporting products
- Promote the products

Update output systems

This sub-process\(^{14}\) (7.1 per GSBPM) manages the update of systems where data and Metadata are stored for dissemination or reporting purposes, including:

- formatting data and Metadata ready to be put into output databases;
- loading data and Metadata into output databases.

The following Services are introduced in this sub-process:

- “Update systems”: Updating the Building Blocks Primary Data Storage and Metadata Management with the relevant data and Metadata;
- “Manage dissemination & reporting data”: allowing to format and load the data and Metadata, as well as establish the link between data and the relevant Metadata;
- “Check dissemination data”: performs a verification of the existence of the necessary links between the concerned data and Metadata;
- “Data conversion”: allows the conversion of a given input data format to the required format for loading data into the Dissemination system.

As mentioned above, the following ESS EARF BBs are either updated by the Services or needed for the realization of the Services in this sub-process:

- Primary data storage
- Metadata Management
- Dissemination data storage

\(^{14}\) Note: formatting, loading and linking of Metadata should preferably mostly take place in earlier phases, but this sub-process includes a check that all of the necessary Metadata are in place ready for dissemination.
**Produce dissemination products**

This sub-process (7.2) produces the dissemination and reporting products to meet user needs. The products can take many forms including printed publications, press releases and web sites. For each Data collection, one or many dissemination products can be created.

The following **Services** are introduced in this sub-process:

- “Prepare product components” allows the provision of explanatory texts, tables, charts etc.;
- “Manage branding/look & feel” allows to give the products a DG ESTAT or NSI specific look & feel;
- “Edit dissemination products” allows the editing of the products and checking whether they meet publication standards;
- “Mapping service”: service which maps the available data versus what actually needs to be disseminated to correspond to the view the user requires. As one example, a web application may need to show the full set of dimensions of data, whilst a mobile app may only wish to display a reduced set;
- “Content review and amendments”: allows doing a final check on the dissemination and reporting products by visualizing the end product in a “real life” environment.

The following **ESS EARF BBs**¹⁵ are needed to realize the functionalities provided by the Services in this sub-process:

- Collaboration
- Dissemination data storage
- Process Orchestrator

**Manage release of dissemination products**

This sub-process (7.3) ensures that all elements for the release are in place including managing the timing of the release. It includes briefings for specific groups such as the press or high level political executives, as well as the arrangements for any pre-release embargoes. It also includes the provision of products to subscribers.

The following **Services** are introduced in this sub-process:

- “Manage dissemination & reporting channels” allows to create and maintain the various dissemination channels;
- “Manage embargo” allows the management of embargoes applicable to the dissemination products;
- “Provide products to subscribers” ensures the provision of the dissemination products to the subscribers;

¹⁵ For a full list of the ESS EARF Building Blocks, see annex.
"Publish on website" allows the publishing of statistical information on DG ESTAT, NSI or EC websites;

"Data browsing and advanced visualization" allows users to browse, download and visualize the data available on the website;

"Machine-to-machine data access": provision of data feeds for further integration by subscribers.

The following ESS EARF BBs are needed to realize the functionalities provided by the Services in this sub-process:

- Collaboration
- Dissemination data storage
- Process Orchestrator

**Promote dissemination products**

Whilst marketing in general can be considered to be an over-arching process, this sub-process (7.4) concerns the active promotion of the statistical products produced in a specific statistical business process, to help them reach the widest possible audience. It includes the use of customer relationship management (CRM) tools, to better target potential users of the products, as well as the use of media including web sites, wikis and blogs to facilitate the process of communicating statistical information to users.

The following Service is introduced in this sub-process:

- "Manage communication of statistical information" to allow the management of communication of statistical information, by means of accessing information through CRM tools, websites, wikis and blogs.

The following ESS EARF BBs are needed to realize the functionalities provided by the Services in this sub-process:

- Process Orchestrator
- Dissemination data storage

**Manage user support**

This sub-process (7.5) ensures that customer queries are recorded, and that responses are provided within agreed deadlines. These queries should be regularly reviewed to provide an input to the over-arching quality management process, as they can indicate new or changing user needs.

The following Service is introduced in this sub-process:

- "Manage customer relationship data” to allow the management of customer data, queries and follow-up of tasks.
The following ESS EARF BBs are needed to realize the functionalities provided by the Services in this sub-process:

- Process Orchesturator

### 4.3 Design Principles

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

<table>
<thead>
<tr>
<th>Name</th>
<th>Adherence to Standards for data and Metadata Exchange</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statement</td>
<td>Data and Metadata standards are implemented among ESS participants</td>
</tr>
<tr>
<td>Rationale</td>
<td>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</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>Use of data warehousing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statement</td>
<td>Data which needs to be disseminated is registered, stored and updated in standardized form in a (unified) data warehouse</td>
</tr>
<tr>
<td>Rationale</td>
<td>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.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>Control only once</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statement</td>
<td>If possible, Dissemination data should only be checked and made ready for Dissemination once in the European production chain.</td>
</tr>
<tr>
<td>Rationale</td>
<td>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.</td>
</tr>
</tbody>
</table>
### Name | 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.

### 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.

### 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.

**Table 6. Design Principles for Dissemination in the ESS Statistical Production Reference Architecture**

### 4.4 ESS Member implementations - Scenarios for realizing the to-be state

There are three scenarios for NSIs to interconnect with the European Dissemination platform:

- **Scenario 1** – where an NSI can use dissemination functions from the European Dissemination platform but is in essence basing itself on (specific, possibly legacy) dissemination solutions developed either locally or collaborating with others;

- **Scenario 2** – where an NSI is using a local replicated version of the European Dissemination platform and possibly combines it with NSI-specific Dissemination Services;

- **Scenario 3** – where an NSI is only using the European Dissemination platform. This may be a long-term to-be state for some NSIs.
4.4.1 Scenario 1- Interoperable

This scenario can be visualized as follows:

In this scenario, the ESS member essentially uses specific data storage and dissemination solutions, which could be both legacy solutions as well as solutions developed by the ESS members. It continues disseminating data and metadata to its end users through the end user interface(s) of its choice. There is value to standardise dissemination metadata at ESS level and therefore to use a replica of an ESS metadata management system. The scenario allows the ESS member to utilize selected ESS functional components from a European Dissemination Platform requiring that data is delivered in SDMX format.

The NSI continues to report data to Eurostat and other international organisations, using the data flow already known today. The main improvement to status quo is that this data (increasingly) conforms to SDMX-based standards for data and Metadata exchange. The data enters the Eurostat production systems, in combination with data from other ESS members.
4.4.2 Scenario 2- Replicated

In this scenario, the ESS member replicates the ESS Dissemination platform. Data are stored locally in an interoperable system. This results in aligned business processes, solutions across the NSI and DG ESTAT. Some examples of concrete products would be data modelling conventions, standard transformations, database schemas, ETL templates.

Locally, the ESS member continues disseminating its data through the end user interface(s) of its choice, either making use of a replicated dissemination component(s) or using (a) dissemination component(s) which is specific to the NSI.

Interoperability of ESS member’s Dissemination Data Storage means that data can be transferred to other members in a seamless way. There is an obvious gain to also implement an ESS replicated solution that would provide built in standardised push or pull mode.

Figure 16. Replication scenario for Dissemination
This scenario could support, at the time when methodology and quality control are sufficiently well aligned, that data available at the NSI Dissemination data warehouse flow directly to the European Dissemination Platform.

### 4.4.3 Scenario 3 - Shared

This scenario can be visualized as follows:

**Figure 17. Replication scenario for Dissemination**

In this scenario, an ESS member solely uses the ESS Dissemination Platform to disseminate its data. It does so by replicating locally the ESS Dissemination Data Storage and exchanging data as from there. A precondition for realizing this scenario is that both Metadata and data fully adhere to pre-defined standards such as SDMX.

The ESS member disseminates its data through a unique ESS endpoint that can be customised to local needs (language, look and feel, etc.).

Again, all scenarios outlined are example scenarios, ESS members may choose to implement a single scenario or combinations of them.
Any questions regarding this Report should be addressed to:

Jean-Marc Museux  
EA team leader  
DG ESTAT  
Joseph Bech building  
5 Rue Alphonse Weicker  
L-2721 Luxembourg  
Telephone: +(352) 430 135 021  
Email: jean-marc.museux@ec.europa.eu