Common Statistical Production Architecture (CSPA)

Monica Scannapieco

ESTP Training Course “Enterprise Architecture and the different EA layers, application to the ESS context – Advanced course”
Rome, 11 – 14 October 2016
What is CSPA?

• **CSPA** – **Common Statistical Production Architecture** is an industry architecture for official statistics.
• CSPA provides a **template architecture** for official statistics, describing:
  
  • **What** the official statistical industry wants to achieve
  • **How** the industry can achieve this, i.e. principles that guide how statistics are produced
  • **What** the industry will have **to do**, compliance with the CSPA
CSPA Scope

• Provide guidance for building reliable and high quality services to be shared and reused in a distributed environment (within and across statistical organizations)

• Enable international collaboration initiatives for building common infrastructures and services

• Foster alignment with existing industry standards such as the Generic Statistical Business Process Model (GSBPM) and the Generic Statistical Information Model (GSIM)
CSPA project - 1

- CSPA 2013: first CSPA specification
- CSPA 2014: CSPA Services Build/Deploy
  - Architecture Working Group: refinement of specification
  - CSPA catalogue
CSPA project - 2

- CSPA 2015: CSPA Implementation project
- To extend the support offered for the implementation of CSPA compliant statistical services
  - Classification retrieval service
  - Probabilistic record linkage service
  - Web dissemination service
  - Confidentialised analysis of microdata service
CSPA project - 3

• To have a plan in place for continued investment in the development of CSPA services
• To facilitate the transitioning of CSPA governance from HLG project governance arrangements
• 2016: CSPA Implementation group
CSPA Assets built 2013 - 2015

• 2013 – Defined the concept and proved it

• 2014 – Proved that we can develop production services and make them visible and available
  • Istat developed CSPA Error Correction Service

• 2015 – Address remaining implementation challenges and made the assets sustainable

• Assets:
  • CSPA – the architecture v1.5
  • Logical Information Model
  • Catalogues
CSPA and the EA Layers

- **Business Architecture** which defines what the industry does and how it is done (statistics in our case),

- **Information Architecture** which describes the information, its flows and uses across the industry, and how that information is managed,

- **Application Architecture** which describes the set of practices used to select, define or design software components and their relationships, and

- **Technology Architecture** which describes the infrastructure technology underlying (supporting) the other architecture perspectives.
CSPA Business Architecture

• "Business Architecture covers all the activities undertaken by a statistical organization, including those undertaken to conceptualize, design, build and maintain information and application assets used in the production of statistical outputs. Business Architecture drives the Information, Application and Technology architectures for a statistical organization."

CSPA Specification 1.5
CSPA Business Architecture

• **CSPA focuses on statistical production as bounded by GSBPM**
• **Business concerns NOT specific of statistical production are out of CSPA scope**
  • *E.g.: Recruiting, retaining and developing staff with relevant skills*
CSPA Business Architecture

- Main Concepts (defined as of GSIM)
  - **Business Function**: something an enterprise does, or needs to do, in order to achieve its objectives
  - **Business Process**: set of process steps to perform on or more Business Functions to deliver a Statistical Program
  - **Business Service**: is the means of accessing a Business Function. It will perform one or more Business Processes. A Statistical Service is a kind of Business Service
CSPA Information Architecture

• “Information Architecture (IA) classifies the information and knowledge assets gathered, produced and used within the Business Architecture. It also describes the information standards and frameworks that underpin the statistical information. IA facilitates discoverability and accessibility, leading to greater reuse and sharing.”

CSPA Specification 1.5
CSPA Information Architecture

- **Reference Frameworks:**
  - GSBPM for business processes
  - GSIM for information input to, and output from, business processes
  - Logical information Model (LIM) for bridging the conceptual and implementation layer in the specification of a CSPA service
# CSPA Logical information Model - 1

<table>
<thead>
<tr>
<th>Conceptual</th>
<th>GSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logical</td>
<td>CSPA Logical Information Model (LIM)</td>
</tr>
<tr>
<td>Physical</td>
<td></td>
</tr>
<tr>
<td>Relevant modelling from SDMX</td>
<td>Relevant modelling from DDI</td>
</tr>
<tr>
<td>SDMX instance (SDMX-ML/SDMX-JSON/etc...)</td>
<td>DDI-XML DDI-RDF</td>
</tr>
</tbody>
</table>
CSPA Application Architecture

• “Application architecture is a description of the major logical grouping of capabilities that manage the data objects necessary to process the data and support the business - it details the structure of components, their inter-relationships, and the principles and guidelines governing their design and evolution over time.”

CSPA Specification 1.5
CSPA & SOA - 1

- CSPA is based on Service Oriented Architecture (SOA)
  - Services are self-contained and can be reused by a number of business processes (either within or across statistical organizations)
CSPA & SOA - 2

- CSPA Services can be of two types
  - Statistical Function Service
  - Statistical Entity Service
CSPA & SOA - 3

- A **Statistical Function Service** will perform a task in the statistical process, at different levels of granularity:
  - An atomic or fine grained statistical function service may, for example, support the application of a methodological step within a GSBPM sub process
  - Coarse grained or aggregate Statistical Services will encapsulate a larger piece of functionality, for example, a whole GSBPM sub process.
CSPA & SOA - 4

- A **Statistical entity service** provide access to statistical information entities (objects) in order to support statistical production processes.

- Examples:
  - Classification services - for statistical classifications
  - Register services - for business, address, and household register information
  - Geography services - for geographic information
  - Statistical metadata services - for statistical metadata throughout GSBPM statistical production
CSPA & SOA - 5

• Statistical Organizations do also have **Utility Services**, i.e. reusable service components of non-statistical nature to create solutions

• Out of CSPA scope
CSPA Service Layers

- CSPA Statistical Services
  - Statistical Function Services
  - Entity Services

- IT Community Sharing
  - Utility Services
  - Libs & Tools

- IT Experience Sharing
  - Environment Platform (e.g. Messaging)
  - Environment Platform (e.g. Pub/Sub)
  - Environment Platform (e.g. R)
  - Environment Platform (e.g. SAS)
  - Technical Infrastructure

Eurostat
Service Definition, Specification, Implementation
CSPA Service Design & Implementation

Service Definition
- GSIM Objects

Service Specification
- Inputs
- GSIM Impl. Objects

Service Implementation Description
- GSIM Impl. Objects

Capabilities
- Statistical Service
- Inputs
- Outputs
- GSIM Objects

Functionality
- Statistical Service
- Inputs
- GSIM Impl. Objects

Operations
- Statistical Service
- Inputs
- GSIM Impl. Objects

Existing App
- Dependency

Infrastructure (SAS, SPSS, C compiler, JVM, DBMS, etc)

Metrics
Methodology
Non-functional requirements

Protocol
Tech dependencies
CSPA Technology Architecture

- "Technology Architecture (TA) describes the IT infrastructure required to support the deployment of business services, data services and applications services, including hardware, middleware, networks, platforms, etc."
Statistical services and communication platform
Statistical services and communication platform

• Communication platform:
  
  • Enterprise service bus
  • CORE
  • Whatever platform (even no platform) chosen at National level

• Transitional state for the Coding 2 Statistical Service that does not (yet) implement CSPA specification
Example: Istat’s CSPA Error Correction Service

Istat in particular developed the Error correction service by:

- Defining the CSPA service definition
- Defining the CSPA service specification
- Implementing the service
CSPA Error
Correction service
definition

It is a conceptual-level
definition of the service
where are described:

• the principal business
  functionalities

• the inputs and outputs
  according to GSIM

<table>
<thead>
<tr>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
</tr>
<tr>
<td>GSBPM</td>
</tr>
<tr>
<td>Business Function</td>
</tr>
<tr>
<td>Outcomes</td>
</tr>
<tr>
<td>Restrictions</td>
</tr>
<tr>
<td>GSIM Inputs</td>
</tr>
<tr>
<td>GSIM Outputs</td>
</tr>
<tr>
<td>Service dependencies</td>
</tr>
<tr>
<td>Process Method</td>
</tr>
</tbody>
</table>
CSPA Error Correction service specification

It is a logical-level specification of the service in which:

- some design issues are addressed with respect to the methods of the service (e.g. how to invoke such methods)

- GSIM inputs and outputs implementation is specified according to a defined “logical” model (e.g. SDMX or DDI)
CSPA Error Correction issues & solutions

The development of this service was more than a technological exercise.

Issues emerged on:

- the format/model of I/O data
  - JSON Table schema
- the protocol and on the interface to implement
  - REST Web Service

The final output is compatible with the solutions provided by the AWG to such issues.
CSPA architecture and technologies

- Each NSI or organization that implements a CSPA compliant architecture can be both a service consumer or provider.

Step 1: POST job details
Step 2: poll job status
Step 3: GET job results
### Global CSPA Catalogue

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge Base (Virtual Help Desk)</td>
<td>Provides access to standards such as GSIM, GSBPM and GAMSO, as well as information about DDI and SDMX</td>
</tr>
<tr>
<td>Investment Catalogue</td>
<td>Provides information about the developments that have been planned or are in progress by statistical organizations</td>
</tr>
<tr>
<td>Capability Catalogue</td>
<td>Provides information about the developments that have been completed</td>
</tr>
<tr>
<td>CSPA Service Catalogue</td>
<td>Provides access to a list of Statistical Services that have been developed. This catalogue is hosted by Eurostat</td>
</tr>
<tr>
<td>Technical Repository</td>
<td>Currently accessed via the Service Implementation Specification for the CSPA Service, the current repository is Github.com</td>
</tr>
</tbody>
</table>

Currently accessed via the Service Implementation Specification for the CSPA Service, the current repository is Github.com.