**The Common Statistical Production Architecture: An Important New Tool
for Process Standardisation**

**Steven Vale**

**United Nations Economic Commission for Europe**

**Abstract**

Many statistical organizations face the challenges of rigid processes and methods, and inflexible, ageing technology environments. Process and methodology changes are time consuming and expensive resulting in an inflexible, unresponsive organization.

Statistical organizations have also often attempted to share processes, methodologies and solutions. This has usually involved an organization taking a copy of a component and trying to integrate it into its own environment. However, in most cases this has required significant work as the component was not designed in a way that facilitates sharing.

The Common Statistical Production Architecture (CSPA) addresses these issues. It builds on existing standards such as the Generic Statistical Business Process Model (GSBPM) and the Generic Statistical Information Model (GSIM) to create an agreed set of common principles and standards designed to promote greater interoperability within and between statistical organizations. It was developed under the High-Level Group for the Modernisation of Statistical Production and Services to provide the “industry architecture” for official statistics.

This paper introduces the CSPA, with examples of how it can be used in practice to promote greater standardisation and facilitate sharing between organisations.

**1. Introduction**

The High-Level Group for the Modernisation of Statistical Production and Services (HLG)[[1]](#footnote-1) was created in 2010 by the Bureau of the Conference of European Statisticians. It comprises the heads of national and international statistical organizations, and has a mandate to reflect on and guide strategic developments in the ways in which official statistics are produced.

At the end of each year HLG organises a workshop, inviting representatives of up to 30 expert groups and projects related to modernisation of official statistics. These workshops ensure coordination of activities, review progress, and determine the key priorities for the following year. In November 2012, the HLG workshop decided that developing a Common Statistical Production Architecture (CSPA) was a key priority. In November 2013, the workshop supported further work on the implementation of the CSPA.

This paper summarises the main results of these activities to date, as well as the expected outcomes in terms of improved standardisation, inter-operability and collaboration between statistical organisations.

**2. Why Develop a Common Statistical Production Architecture? (CSPA)**

Many statistical organisations have built production processes in a fairly ad-hoc way, with little standardisation. Although the processes are conceptually very similar, the individual solutions are not (as represented by the different shapes in Figure 1). Technical solutions have been developed for very specific purposes with little regard for ability to share information and tools between different statistical processes, and even less for sharing between organisations. This can be referred to as 'accidental architecture', as the processes and solutions were not designed from a holistic viewpoint. The cost of maintaining this type of approach is becoming unsustainable.

This approach makes it difficult to replace even one of the components supporting statistical production, or to apply common standards such as DDI (Data Documentation Initiative) and SDMX (Statistical Data and Metadata eXchange). Process and methodology changes are time consuming and expensive resulting in an inflexible, unresponsive statistical organisation.



Figure 1: Accidental Architectures

To address these problems, some statistical organisations are modernising their processes using an “enterprise architecture” approach. This approach enforces standards and improves collaboration across an organisation, and is illustrated in Figure 2.



Figure 2: The result of standardization within an organization

However, standardising at the level of the organisation is not necessarily the optimum solution, as each organisation may specify the components and interfaces of their statistical production processes in different ways, making it hard to share components between organisations. Figure 3 shows the example that while a component from Canada might support the same process, it will not be simple to integrate it into the Swedish environment.



Figure 3: Why sharing /reuse is hard now

Adopting a common reference architecture will make it easier for each organisation to standardise the components of statistical production, regardless of where these components originate. As shown in Figure 4, Sweden could reuse a component from Canada because they both use the same specifications. These components (or services) might be new ones, built to comply with the principles of the CSPA or they might be existing components that are “wrapped” to make them CSPA-compliant. This is shown in Figure 4 by the shapes inside the building blocks.



 Figure 4: Common architecture makes sharing and reuse easier

**3. How we developed the CSPA**

The CSPA development project (sometimes also referred to as “plug and play”), was an international collaboration activity, involving nine national and two international statistical organisations, coordinated by the UNECE. Figure 5 shows the extent of the participation in the CSPA development project. In the left-hand block, the area of the shape for each country / organisation is roughly proportionate to the volume of input.



Figure 5 – Inputs to the Common Statistical Production Architecture project

The project had two main strands. The first was to develop the architecture, and the second was to test it in practice through a proof of concept. Both strands of the project delivered their results on schedule at the end of 2013.

**3.1 The architecture strand**

This strand was responsible for the conceptual work to develop an “industry architecture” for official statistics. It started with a one-week “Architecture Sprint”, held in Ottawa in April, hosted by Statistics Canada. Fifteen participants from ten organisations reached agreement in all key areas, and produced version 0.1 of the architecture documentation, which was released for public comment.

The public consultation resulted in over 200 items of feedback, mostly favourable or seeking further clarification. This feedback was considered at a one-week “Design Sprint” in June, hosted by ISTAT in Rome. The focus at this sprint was to apply the proposed architecture to design the proof of concept.

Further drafts of the architecture were prepared, discussed, and circulated for comments, resulting in version 1.0, which was published on 24 December[[2]](#footnote-2).

**3.2 Proof of concept strand**

This strand started in summer 2013, and aimed to apply the architecture in practice by creating CSPA-compliant services, and testing them in different production environments. Given the short timeframe in which to complete the proof of concept, these services could not be built from scratch. Instead, the organisations involved in the project identified existing tools and applications that could be “wrapped” to make them CSPA-compliant.

The five tools chosen were:

* Blaise: A data collection, data editing and data processing tool developed by Statistics Netherlands. For the Proof of Concept only the collection function was involved. This was wrapped by the Australian Bureau of Statistics.
* Editrules: An error localization tool developed at Statistics Netherlands and wrapped by ISTAT.
* CANCEIS (CANadian Census Edit and Imputation System): An editing tool used for error localization and imputation developed and wrapped by Statistics Canada
* GCode: A generalized automated and assisted coding tool developed by Statistics Canada, and wrapped by Statistics Netherlands.
* Statistical Coding Service: A coding tool developed and wrapped by Statistics New Zealand.

The resulting CSPA-compliant services were implemented (in various combinations) by ISTAT, Statistics New Zealand and Statistics Sweden[[3]](#footnote-3). The proof of concept work had the following aims and results:

|  |  |
| --- | --- |
| **Aim** | **Result** |
| CSPA is practical and can be implemented by various organisations in a consistent way | The statistical organisations involved were successful in building CSPA services. |
| CSPA does not prescribe the technology platform an agency requires | Statistics New Zealand and ISTAT have different technical environments. Both organisations were successfully able to implement the same auto-coding service into their environments. |
| CSPA Statistical Services can fit into existing processes | Statistics New Zealand had an existing implementation of CANCEIS. They were able to implement the Editing and Imputation Service (i.e. a wrapped CANCEIS) into their environment. |
| CSPA compliant services can be swapped easily | Statistics New Zealand implemented both auto- coding services into their environment. It was very easy to swap the services without the need for significant IT input. |
| Reusing the same statistical service by configuration | Statistics Sweden in their implementation of the wrapped Blaise service showed that they could configure both the environment and the service for different surveys. |

The key lessons learnt during the proof of concept work included:

* CSPA is a viable approach for statistical organisations.
* Participants found that working with colleagues around the globe was stimulating and broadened their understanding, but time-zone differences caused some delays.
* Licenses for the original software were required to wrap and implement the services. Obtaining the licences took some time and caused (small) delays in starting work. In some cases, this was resolved by the direct intervention of chief statisticians. Faster and more flexible licensing arrangements would help future activities.

Overall, the CSPA development project was successful in providing the official statistical industry with the first version of an industry architecture that has been tested and shown to be viable. However, further work was needed to help organizations implement CSPA, so a new CSPA implementation project was launched in January 2014[[4]](#footnote-4).

**4. Implementing the CSPA**

The CSPA implementation project aims to take the architecture developed last year, and use it to develop some components of the statistical production process, to the point where they can be (and are being) used in regular statistical production activities. The aim is to have several components (or “services”) by the end of the year. Some will be developed by two or more statistical organisations in partnership, whilst others will be developed by individual organisations. The current list of agreed services is:

* Seasonal Adjustment – France, Australia, New Zealand
* Confidentiality on the fly – Canada, Australia
* Error correction – Italy
* SVG Generator – OECD
* SDMX transform – OECD
* Selecting sample from business register – Netherlands
* Editing components – Netherlands
* Classification Editor – Norway

Agreements have been signed with the organisations creating these components, with a general timetable that components should be designed by the end of May, built by mid-September, and implemented before the HLG workshop in November 2014.

The project will also create a “catalogue”, which will be a central store for these components, making them available to all statistical organisations. The catalogue will be global in nature, and will provide information about existing services, other resources and potential collaboration partners. The catalogue will support the lifecycle management, governance and use of components and services.

Australia, Canada, Hungary, Italy, New Zealand, Romania, Turkey and Eurostat are participating in the catalogue team.

An “Architecture Working Group” is providing over-arching support to the different activities on any issues relating to the understanding and use of the Common Statistical Production Architecture. This group will evolve to become the body responsible for the maintenance of the CSPA after the end of the current project.

The Architecture Working Group includes representatives from Australia, Austria, Canada, France, Italy, Mexico, Netherlands, New Zealand, Turkey, United Kingdom and Eurostat.

1. <http://www1.unece.org/stat/platform/display/hlgbas> [↑](#footnote-ref-1)
2. http://www1.unece.org/stat/platform/display/CSPA/CSPA+v1.0 [↑](#footnote-ref-2)
3. http://www1.unece.org/stat/platform/display/CSPA/CSPA+Proof+of+Concept+2013 [↑](#footnote-ref-3)
4. http://www1.unece.org/stat/platform/display/CSPA/Implementing+CSPA+Project+2014 [↑](#footnote-ref-4)