UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE

ECE/CES/BUR/2014/JAN/15
20 January 2014

CONFERENCE OF EUROPEAN STATISTICIANS

Meeting of the 2013/2014 Bureau
Geneva (Switzerland), 28-29 January 2014

For discussion and recommendations

Item 3 (g) of the Provisional Agenda

Outcomes of the 2013 HLG Projects

Note by the Secretariat

The note provides a progress report of the activities overseen by the High-level Group for the Modernisation of Statistical Production and Services in 2013. ***The Bureau reviewed the outcomes of work and provided advice on the way forward.***

I. INTRODUCTION

1. The High-level Group for the Modernisation of Statistical Production and Services (HLG) 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.

2. Each year HLG organises a Workshop, inviting representatives of up to 30 expert groups and projects related to modernisation of official statistics. These workshops help to ensure coordination of activities. They also review progress and determine the key priorities for the following year. In 2012, the HLG workshop decided that creating a Common Statistical Production Architecture, and further developing the Frameworks and Standards for Statistical Modernisation were the two highest priorities. HLG launched international collaboration projects to address these priorities. Both projects ran during the calendar year 2013. This paper summarises the main results of these projects.

II. COMMON STATISTICAL PRODUCTION ARCHITECTURE (CSPA)

3. The background to this project is that many statistical organisations have built production processes in a fairly ad-hoc way, with little standardisation. The cost of maintaining this type of approach is becoming unsustainable.

4. 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 process and solutions were not designed from a holistic viewpoint.

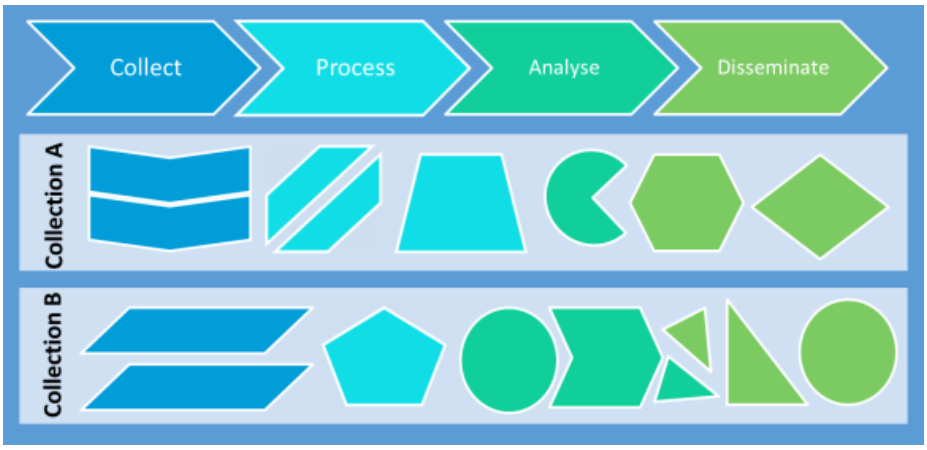

Figure 1: Accidental Architectures

5. 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.

6. 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.

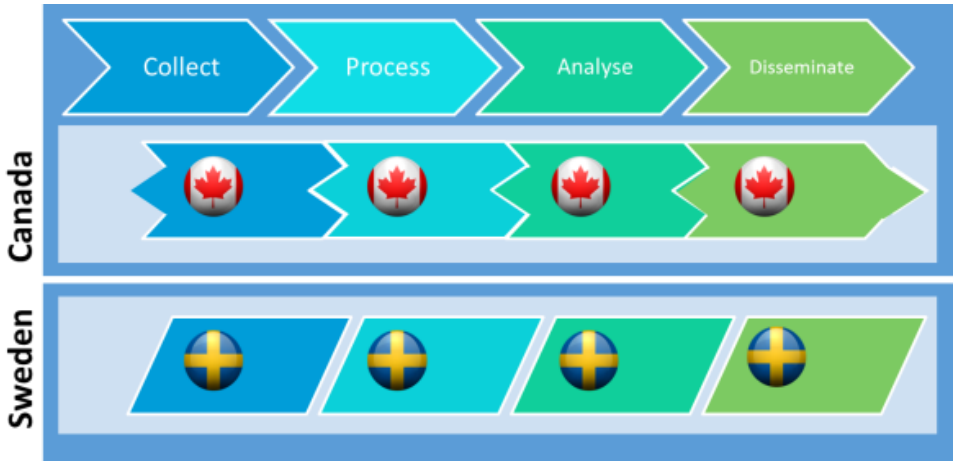

Figure 2: The result of standardization within an organization

7. 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.

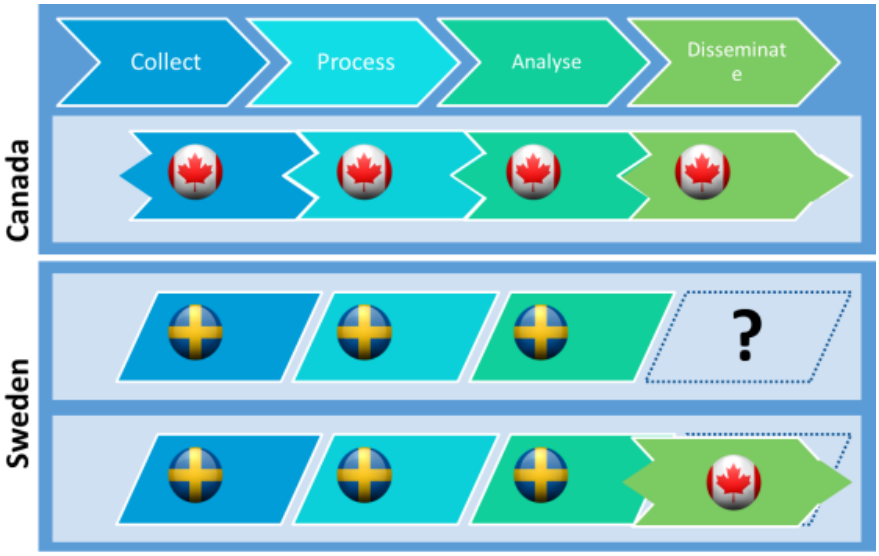

Figure 3: Why sharing /reuse is hard now

8. 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.

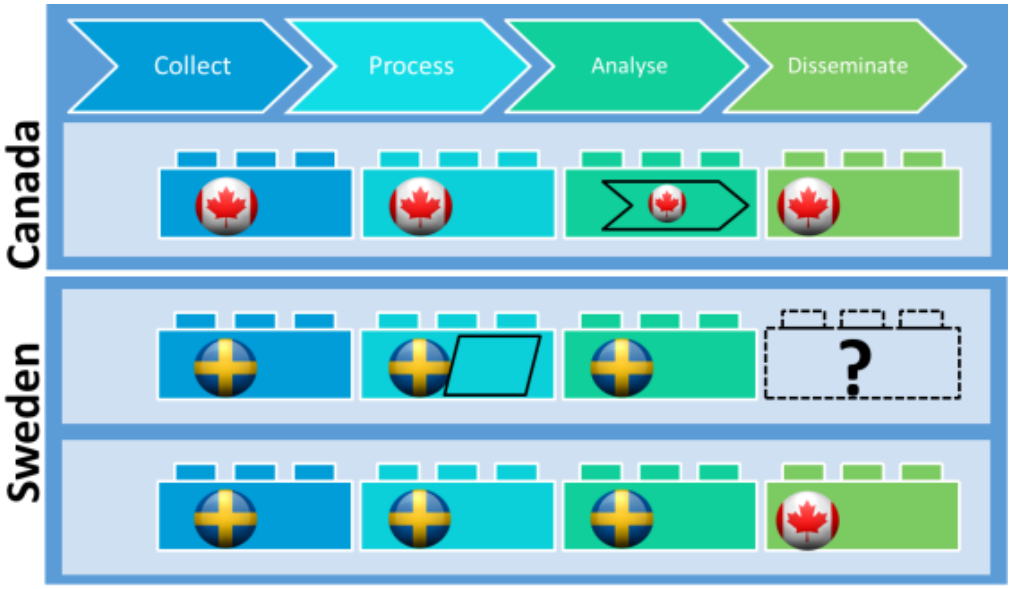

Figure 4: Common architecture makes sharing and reuse easier

9. The CSPA project (sometimes also referred to as "plug and play"), 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.

A. Architecture Strand

10. This strand was responsible for the conceptual work to develop an "industry architecture" for official statistics. It started with a one-week "Architecture Sprint[1]", 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.

11. 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.

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

B. Proof of Concept Strand

13. 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.

14. 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.

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

[1] A "sprint" is a very intensive workshop, bringing together key experts from different organisations and disciplines to deliver a specific output in a short period of time.
[2] See http://www1.unece.org/stat/platform/display/CSPA

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.

16. 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.

17. Overall, the CSPA 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. Further work is needed to help organizations implement CSPA, so one of the key priorities for 2014 is a follow-up project, under HLG, focussing on CSPA implementation.

III. FRAMEWORKS AND STANDARDS FOR STATISTICAL PRODUCTION

18. This project was a priority for the official statistics community because a lot of effort had gone into the development of standards such as the Generic Statistical Business Process Model (GSBPM) and the Generic Statistical Information Model (GSIM). To realise the full benefits of this investment, continued coordination was necessary during the implementation phase, to ensure that the lessons learned during implementation were shared.

19. This project had four main objectives:

- To ensure that the international statistical community has access to the standards needed to support the modernisation of statistical production and services

- To increase coherence between these standards
- To provide support mechanisms for the practical implementation of these standards within national and international statistical organisations
- To ensure effective promotion and maintenance of the GSBPM and the GSIM, including the release of new versions as appropriate.

20. The work packages within the project, and their main outcomes were:

Work Package 1: Support for the implementation of GSIM

21. This work package included providing support for a community of GSIM "early adopters", by creating a "GSIM Implementation Group" which met regularly by web conference, and shared ideas and questions through a wiki-based discussion forum. The main output was a set of proposals for enhancing GSIM and its associated documentation.

Work Package 2: Enhancement of GSIM

22. This work package gathered feedback from the users of GSIM, the model describing the information objects used in statistical production, created by a project overseen by HLG in 2012. Feedback was sought from those with experiences of practical implementations, as well as potential user groups within and outside official statistics. The main output was GSIM version 1.1, released on 24 December[3].

23. This new version simplified some areas of the model, where complexity was a potential barrier to implementation, and increased the alignment of GSIM with other models and good practices. For example, a review of the Neuchâtel Model of classification terminology was taking place independently as a joint activity between the UN Expert Group on Classifications, and the Conference of European Statisticians Steering Group on Statistical Metadata. The reviewers and the project team realised that there was considerable overlap in terms of what the two models were trying to cover. The result is that the revised Neuchâtel model has been incorporated as the classification part of GSIM[4].

Work Package 3: Update of GSBPM

24. This work package gathered feedback from users on potential changes to GSBPM, the model that defines and describes the different steps in the statistical production process. As GSBPM version 4.0 was adopted in 2009, and already implemented by over 50 statistical organisations around the world, changes were deliberately kept to a minimum. They had to be supported by a good business case, and a majority of the participating organisations. This was to ensure that the impact of the new version on the user community was kept to a minimum. The main output was version 5.0 of GSBPM, released on 24 December[5].

Work Package 4: GSIM / DDI / SDMX mapping

25. This work package included detailed mappings between the information objects in GSIM and those in the information models of DDI and SDMX. The aim was to identify any

[3] http://www1.unece.org/stat/platform/display/gsim
[4] http://www1.unece.org/stat/platform/display/gsim/Statistical+Classification+Model
[5] http://www1.unece.org/stat/platform/display/GSBPM

issues affecting the coherence of these standards, and propose solutions where possible. The main outputs were mappings between these standards[6].

26. Building on the success of GSIM to DDI mapping group, several statistical organisations expressed an interest in further developing this work to create a set of DDI profiles. These profiles identify the specific parts of the DDI standard to use in particular scenarios related to the production of official statistics. Five such profiles were created and published. This work will continue under the new Modernisation Committee on Standards, which reports to HLG.

Work Package 5: Training, advocacy and communication

27. This work package included all activities to communicate and promote the idea of standards-based modernisation. It included presentations to expert groups and conferences, and preparing brochures, posters and other promotional materials. The main outputs were the various presentations and brochures published on the UNECE wiki platform.

Work Package 6: Links to geo-spatial standards

28. This work package provided an initial assessment of the role of geo-spatial standards in the modernisation of official statistics. The main output was a paper on this topic[7].

Work Package 7: Set up a virtual standards helpdesk

29. This work package established a central point for information relating to the standards necessary to support statistical modernisation. This virtual helpdesk facility links to the Global Inventory of Statistical Standards, but focuses more on supporting cross-cutting statistical modernisation. The main output was the helpdesk wiki site[8], which will be further enhanced by the Modernisation Committee on Standards.

Work Package 8: Mapping GSBPM to the Fundamental Principles of Official Statistics

30. This requirement was identified during a high-level seminar for leaders of statistical organisations in SPECA (Special Programme for the Economies of Central Asia) countries. The main output was a paper, published on the HLG wiki[9].

---

[6] http://www1.unece.org/stat/platform/display/gsim/GSIM+and+standards

[7] http://www1.unece.org/stat/platform/download/attachments/63931489/Geospatial%20Paper%20%281%29.docx?version=1&modificationDate=1382920480519&api=v2

[8] http://www1.unece.org/stat/platform/display/VSH

[9] http://www1.unece.org/stat/platform/download/attachments/63931489/Mapping%20GSBPM%20to%20the%20FPs.pdf?version=1&modificationDate=1376318548313&api=v2

IV. PROJECT RESOURCES

31. Apart from a full-time project manager, sponsored by the Australian Bureau of Statistics, and working with the UNECE Secretariat, all resources in both projects were voluntary. This is typical of HLG projects, and means that the results reflect as far as possible the needs of the participating organisations. It is also the reason why the results differ to some extent to those proposed in the original project plans. Both projects delivered more in certain areas where there was a strong interest from participants, and slightly less in a few other areas. This flexibility to respond to evolving requirements and changing priorities is a key factor in the success of the projects.

32. The following diagrams show the extent of the participation in the two projects. In the left-hand block, the area of the shape for each country / organisation is roughly proportionate to the volume of input.

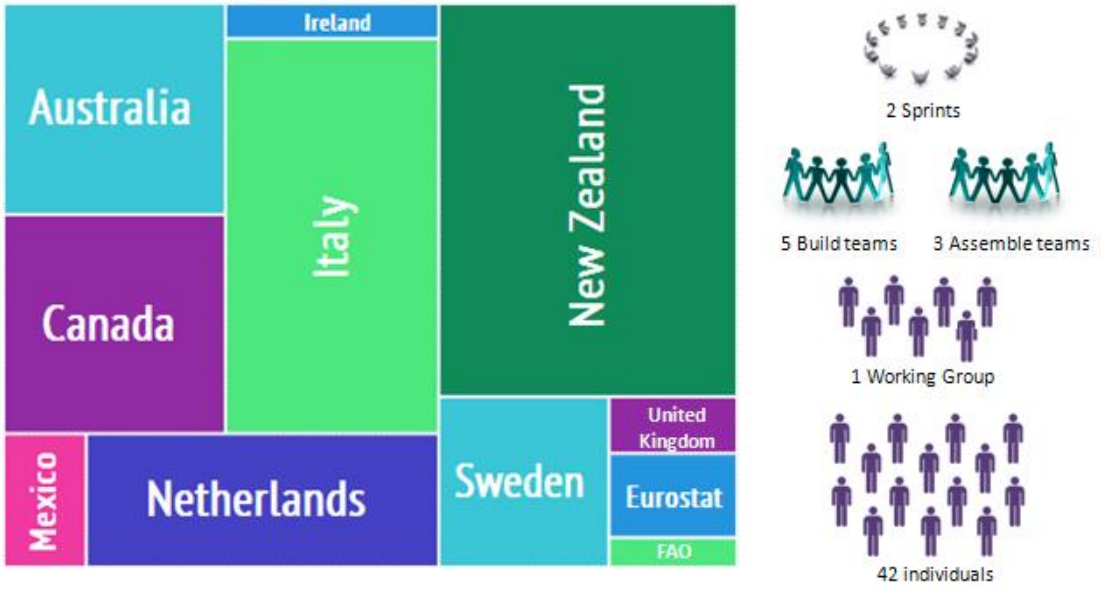

Figure 5 – Inputs to the Common Statistical Production Architecture Project

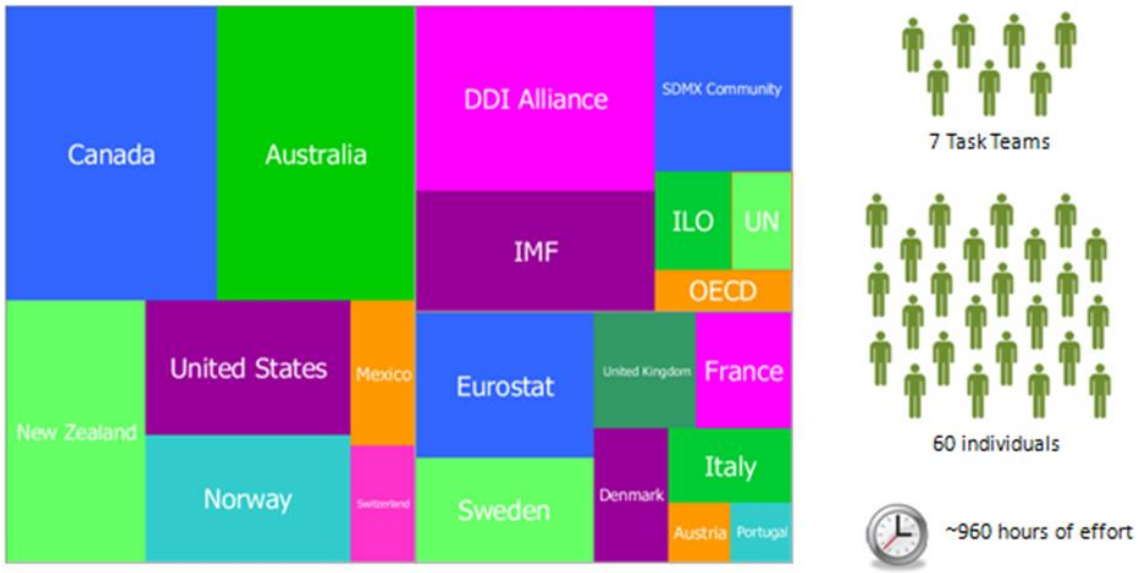

Figure 6 – Inputs to the Frameworks and Standards for Statistical Production Project

V. NEXT STEPS

33. The work on the Common Statistical Production Architecture has moved from development to implementation. This is the subject of a major international collaboration project under HLG in 2014.

34. The work on managing and developing the Frameworks and Standards for Statistical Production has passed to the new Modernisation Committee on Standards, which has the responsibility for maintaining and supporting GSBPM and GSIM.

35. In 2014, HLG is also overseeing an international collaboration project on the use of "Big Data" for official statistics purposes. GSBPM, GSIM and CSPA will all influence the development of methods and procedures for the use of Big Data sources, so the relevance of these standards for statistical modernisation activities will be further tested.

VI. CONCLUSION

36. Whilst it remains rather early to give a final judgement on the value of the two HLG projects undertaken in 2013, the initial signs are very positive. This is largely due to the strong focus for HLG projects to be demand-driven, and to respond to the needs of participating organisations. The success of these projects can, therefore, be seen as a triumph for the processes of international collaboration. The outputs can truly claim to be the intellectual property of the international statistical community as a whole.

* * * * *