



Economic and Social Council

Distr.: General
3 April 2014

English only

Economic Commission for Europe

Conference of European Statisticians

Sixty-second plenary session

Paris, 9-11 April 2014

Item 5 of the provisional agenda

Work of the High-level Group for the Modernisation of Statistical Production and Services

Annual report of the High-Level Group for the Modernisation of Statistical Production and Services

Note by the Secretariat

Summary

This paper informs the Conference of European Statisticians about the activities and outputs of the High-Level Group for the Modernisation of Statistical Production and Services since June 2013. The purpose of the Group is to guide strategic developments in the ways official statistics are produced.

The Conference is invited to express views on the progress of work during this period, and the priorities for the coming year.

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 selected 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. 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 help to ensure coordination of activities. They also review progress and identify the key priorities for the following year. In November 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. In November 2013, the workshop selected Big Data in Official Statistics, and Implementation of the Common Statistical Production Architecture.

3. The HLG launched international collaboration projects to address these priorities. HLG projects typically run for one calendar year, therefore the projects selected in November 2012 were completed by the end of 2013, and the projects selected in November 2013 will be completed by the end of 2014. This paper summarises the main results of the 2013 projects, the progress so far on the 2014 projects, and the working methods and governance structures put in place to deliver the necessary results.

II. Projects completed in 2013

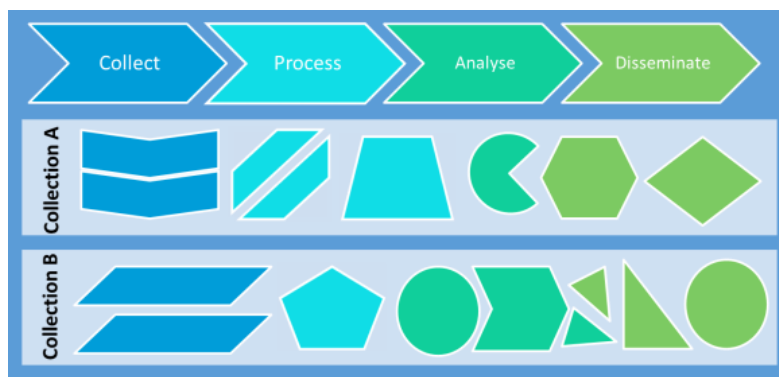
A. Common Statistical Production Architecture

4. 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.

5. 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

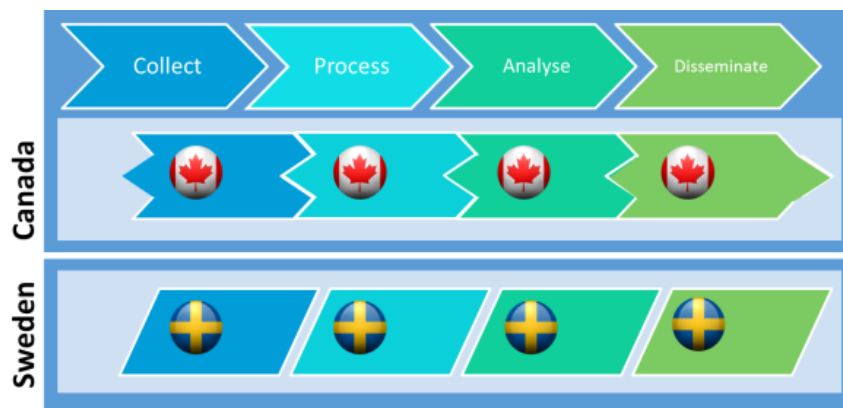


6. This approach makes it difficult to replace even one of the components supporting statistical production, or to apply common standards such as the Data Documentation Initiative (DDI) and the Statistical Data and Metadata eXchange (SDMX). Process and methodology changes are time consuming and expensive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

7. 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

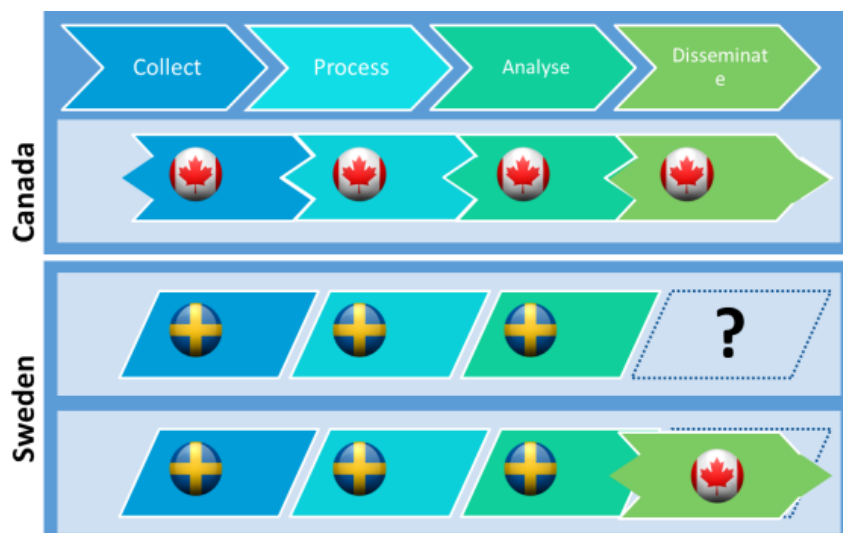
The result of standardization within an organization



8. 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

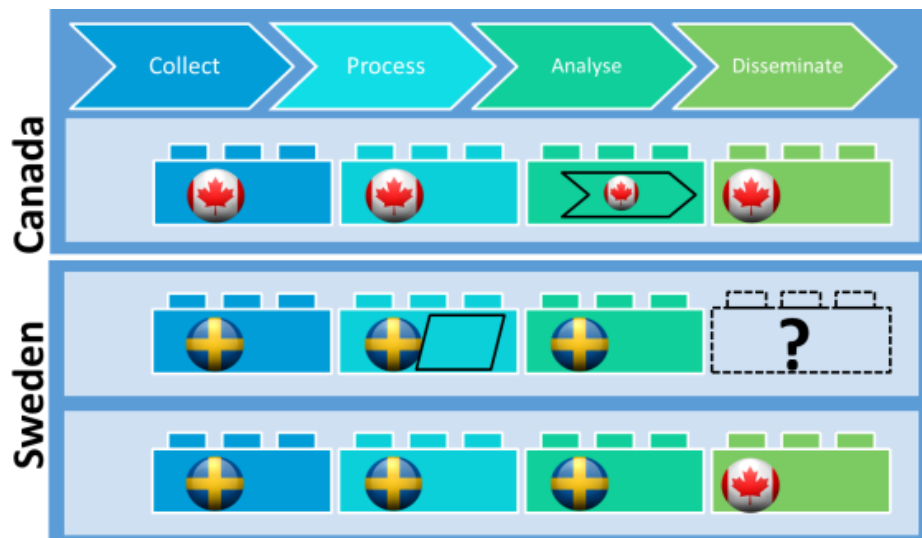
Why sharing /reuse is hard now



9. 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 Common Statistical Production Architecture (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



10. 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.

1. Architecture strand

11. This strand was responsible for the conceptual work to develop an “industry architecture” for official statistics. It started with a one-week “Architecture Sprint¹”, 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.

12. 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 the National Institute for Statistics (ISTAT) in Rome. The focus at this sprint was to apply the proposed architecture to design the proof of concept.

13. Further drafts of the architecture were prepared, discussed, and circulated for comments, resulting in version 1.0, which was published on 24 December².

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>

2. Proof of concept strand

14. 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.

15. The five tools chosen were:

(a) A data collection, data editing and data processing tool (Blaise) developed by Statistics Netherlands. For the Proof of Concept only the collection function was involved. This was wrapped by the Australian Bureau of Statistics;

(b) An error localization tool (Editrules) developed at Statistics Netherlands and wrapped by ISTAT;

(c) The Canadian Census Edit and Imputation System (CANCEIS): An editing tool used for error localization and imputation developed and wrapped by Statistics Canada;

(d) A generalized automated and assisted coding tool (GCode) developed by Statistics Canada, and wrapped by Statistics Netherlands;

(e) Statistical Coding Service: A coding tool developed and wrapped by Statistics New Zealand.

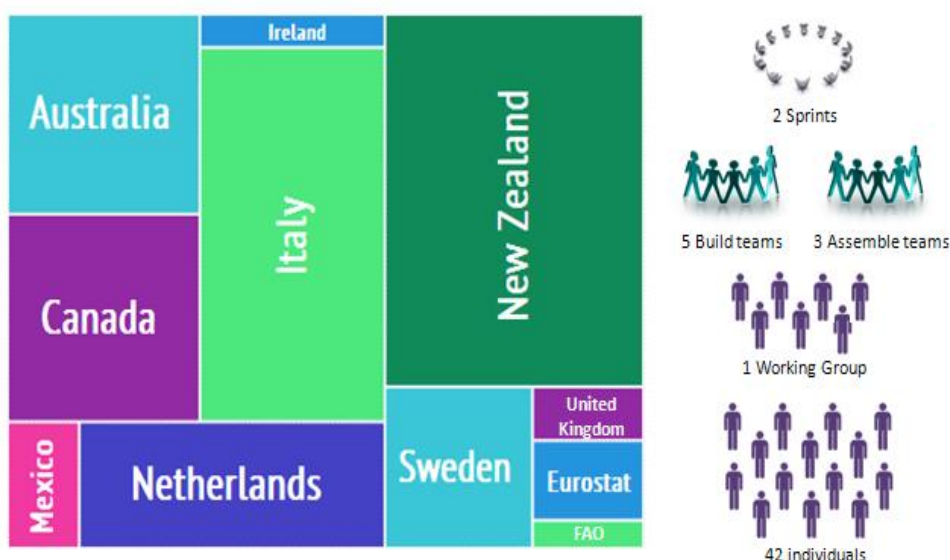
16. The resulting CSPA-compliant services were implemented (in various combinations) by ISTAT, Statistics New Zealand and Statistics Sweden. The proof of concept work had the following aims and results:

| 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. |

17. 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.
18. 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.
19. The following diagram 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



B. Frameworks and standards for statistical modernisation

20. 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.
21. 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;

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

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

1. Work package 1: Support for the implementation of the Generic Statistical Information Model

23. 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.

2. Work package 2: Enhancement of the Generic Statistical Information Model

24. 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 December³.

25. 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 United Nations 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 GSIM⁴.

3. Work package 3: Update of the Generic Statistical Business Process Model

26. 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 December⁵.

4. Work package 4: Mapping of the Generic Statistical Information Model, Data Documentation Initiative and the Statistical Data and Metadata eXchange

27. 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 standards⁶.

28. 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.

5. Work package 5: Training, advocacy and communication

29. 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 United Nations Economic Commission for Europe (UNECE) wiki platform.

6. Work package 6: Links to geo-spatial standards

30. 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 topic⁷.

7. Work package 7: Set up a virtual standards helpdesk

31. 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 site⁸, which will be further enhanced by the Modernisation Committee on Standards.

8. Work package 8: Mapping the Generic Statistical Business Process Model to the Fundamental Principles of Official Statistics

32. This requirement was identified during a high-level seminar for leaders of statistical organisations in the Special Programme for the Economies of Central Asia (SPECA) countries. The main output was a paper, published on the HLG wiki⁹.

33. The following diagram 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.

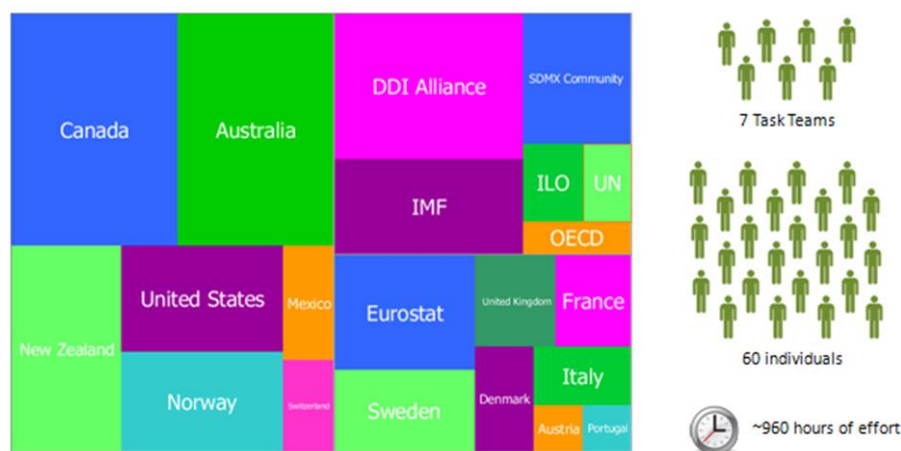
⁶ <http://www1.unece.org/stat/platform/display/gsim/GSIM+and+standards>

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>

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



III. Projects ongoing in 2014

34. The two projects described below have recently started, and will be completed within the calendar year 2014. They are open to participation from any interested country or organisation, including from outside the official statistics community. Participation is on a voluntary, unfunded basis, therefore the specific outcomes will automatically focus on areas considered to be priorities for the participating organisations. Participation can be at many levels, for example throughout a project, just in the specific aspects of a project that are of most interest to an organisation, or just as a reviewer of outputs.

A. Implementation of the Common Statistical Production Architecture

35. This 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. We expect 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:

- (a) Seasonal Adjustment – France, Australia, New Zealand;
- (b) Confidentiality on the fly – Canada, Australia;
- (c) Error correction – Italy;
- (d) Scalable vector graphics (SVG) generator – Organisation for Economic Cooperation and Development (OECD) for programming animated graphics;
- (e) SDMX transform – OECD;
- (f) Selecting sample from business register – Netherlands;
- (g) Editing components – Netherlands;
- (h) Classification Editor – Norway.

36. 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 2014 HLG workshop in November.

37. The project will also create a “catalogue”, which will be a central store for these components, making them available to all statistical organisations. Australia, Canada, Italy, New Zealand, Romania, Turkey and Eurostat are participating in the catalogue team.

38. 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 comprises of representatives from Australia, Austria, Canada, France, Italy, Mexico, Netherlands, New Zealand, Turkey and Eurostat.

B. Big Data in official statistics

39. There is a lot of talk about Big Data. Is it a sort of short-term fashion trend that will soon pass, or is it something that will stay and become part of our work? This project is investigating what Big Data means for official statistics, on the strategic, methodological and technical levels.

40. The first stage of the project was to define the key questions and priorities for the international official statistics community, through a “virtual sprint” session in the week of 10-14 March. Participants from Australia, Canada, Ireland, Italy, Mexico, Netherlands, New Zealand, United Kingdom, Eurostat and OECD joined a web conference for two hours each day during this week, and also worked off-line individually or in small groups. The output of the sprint was a paper on the priority issues to be tackled during the rest of the project. This was circulated for a short public consultation, and then finalised at a workshop hosted by ISTAT in Rome on 2-3 April (back-to-back with a Eurostat Big Data event).

41. In parallel, a “Sandbox” environment is being created, with support from the Central Statistics office (CSO) of Ireland and the Irish Centre for High-End Computing. This will provide a technical platform to load Big Data sets and tools. The sandbox will be launched at a workshop in Dublin on 16 April, on the margins of the meeting on the Management of Statistical Information Systems (14-16 April). It will give participating organisations the opportunity to:

(a) Test the feasibility of remote access and processing – Statistical organisations around the world will be able to access and analyse Big Data sets held on a central server. Could this approach be used in practice? What are the issues to be resolved?;

(b) Test whether existing statistical standards / models / methods etc. can be applied to Big Data;

(c) Determine which Big Data software tools are most useful for statistical organisations;

(d) Learn more about the potential uses, advantages and disadvantages of Big Data sets – “learning by doing”;

(e) Build an international collaboration community to share ideas and experiences on the technical aspects of using Big Data.

IV. Working methods and governance

42. The philosophy behind HLG activities is that all statistical organisations face similar pressures to modernise statistical production. At the same time, resources for this work are limited. It is therefore preferable to pool available resources across organisations, so that each organisation can benefit from the resources and intellectual capital of others. In this way, HLG activities should not impose a burden on participating organisations, instead they should provide much better value for money than is possible if each organisation works on

its own. Greater international harmonisation of official statistics is a useful by-product of this approach.

43. HLG activities and projects are open to all interested organisations, including those outside official statistics. There are already examples of academics and external standards bodies getting involved in specific activities. All outputs are published, and are freely available for anyone to use or adapt.

44. Apart from a full-time project manager working with the UNECE Secretariat, all resources for HLG projects are in the form of volunteers from interested organisations. This means that project results closely reflect the interests and requirements of the participating organisations. It is also the reason why project results can differ to some extent to the original project plans. Projects tend to deliver most in areas where there is a strong interest from participants. This flexibility to respond to evolving requirements and changing priorities is a key factor in the success of the projects.

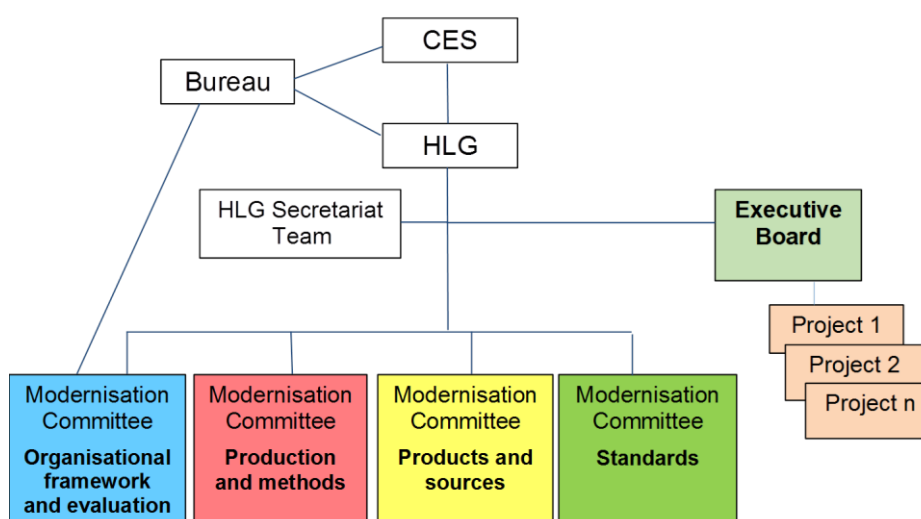
45. HLG projects have pioneered a number of new working methods for international statistical collaboration activities. The concept of the “sprint” session has been borrowed from the “Agile” approach, increasingly used in the IT industry. Sprints are very intensive workshops with around 12-15 participants, which have to deliver a specific output in a given time frame (usually one week). Within the Big Data project, the concept of a “virtual sprint” is being tested. This involves participants meeting for two hours each day for a week by web conference, as well as several hours per day working in smaller groups or individually.

46. Momentum is maintained between sprint sessions by the use of short-term virtual task teams, meeting regularly by web conference, and using the UNECE wiki platform for collaborative drafting, recording issues and solutions, and sharing information.

47. To oversee the HLG projects, and ensure necessary inputs from relevant groups of specialists, the HLG put in place a new governance structure during 2013. This structure is illustrated in Figure 7 below. It has absorbed all existing initiatives and expert groups related to statistical modernisation, operating under the Conference of European Statisticians, reallocating tasks and responsibilities in a way that encourages a more multi-disciplinary approach than was previously the case.

Figure 7

Governance structure for HLG activities



48. The **Executive Board** is responsible for the strategic management of on-going HLG projects. It is comprised of senior managers from national and international statistical organisations, with the authority to commit the resources of their organisations to supporting HLG activities. It also includes the chairs of the four modernisation committees, to ensure coordination, and give access to broader perspective on modernisation activities.

49. The four **modernisation committees** generate ideas for projects, identify good practices, and highlight issues and developments in their respective fields of competence:

(a) **Organizational Framework and Evaluation** – This committee considers human resources, training, legal, licensing and similar issues. It is also responsible for evaluating the success of HLG activities;

(b) **Production and Methods** – This committee considers the technical and methodological and architectural aspects of modernising statistical production. It will oversee the maintenance of the Common Statistical Production Architecture;

(c) **Products and Sources** – This committee considers how to use and integrate a growing range of data sources, and provide new types of outputs and services for users of official statistics. It looks beyond the official statistics community for new data sources and new opportunities to add value by enhancing statistical outputs;

(d) **Standards** – This committee considers how to develop, enhance, integrate and support the range of standards needed for statistical modernisation. It is responsible for the on-going support and maintenance of the Generic Statistical Business Process Model (GSBPM) and the Generic Statistical Information Model (GSIM).

50. As the work of the HLG has expanded, and results such as those outlined in Section II of this paper, the interest in HLG activities has grown. This is clearly indicated in Figure 8, which shows the level of engagement of different national and international statistical organisations in HLG groups and projects as at March 2014.

Figure 8

Engagement map for HLG groups and activities, March 2014



V. Challenges

51. Whilst participation in HLG activities is on a voluntary, self-funded basis, there is a need for a central resource to coordinate these activities. UNECE has taken on the secretariat role for HLG and the modernisation committees, but additional resource is needed to manage the HLG projects. So far, this role has been filled by staff on loan from

national statistical organisations, and financed by generous donations from Australia, Canada, the Netherlands and the Republic of Korea. Other countries have provided contributions in-kind, such as hosting project events or providing infrastructures for project activities.

52. This approach to resourcing HLG activities can, however, be described as “ad-hoc”. A more systematic approach is needed. The UNECE Secretariat has established a simplified process for receiving contributions of up to 30,000 US Dollars, and continues to investigate how to better facilitate contributions, including in the form of subscriptions. The total cost of supporting and managing HLG projects is around 200,000 US Dollars per year.

53. Looking to the future, HLG has started a review of its vision and strategy. These were endorsed by the Conference in 2011 and 2012 respectively. However, they recognised that the target of modernisation is not fixed. New developments in various areas mean that it would be dangerous to set long-term goals without reviewing them from time to time to ensure they are still relevant. This is also in line with the “agile” approach to project management, which has been successfully employed in the HLG projects to date. Revised versions of the vision and strategy are expected to be available by the end of 2014.

VI. Conclusion

54. Whilst it remains rather early to give a final judgement on the value of the outputs of HLG activities, the feedback so far from national and international statistical organisations is very positive. This is largely due to HLG projects and activities being collaborative and strongly demand-driven, responding to the current needs of participating organisations. The main difference between the activities of HLG, and other previous international collaboration projects, is that the HLG work is strongly supported, and actively overseen by chief statisticians, so that it includes a strategic view and does not just focus on technical issues. The success of HLG activities can therefore be seen as a success for the international statistical community as a whole.

55. The Conference is invited to:

- (a) Comment on the usefulness of the outputs of the various HLG activities described above;
 - (b) Express opinions on priorities for future HLG activities;
 - (c) Give guidance and suggestions on dealing with the issues raised in Section V of this paper.
-