

Workshop on the Modernisation of Statistical Production
(Geneva, Switzerland, 15-17 April 2015)

REPORT OF THE WORKSHOP

1. The Workshop on the Modernisation of Statistical Production was held in Geneva, Switzerland, from 15-17 April 2015. It was attended by representatives from the statistical offices of Australia, Azerbaijan, Bulgaria, Canada, Croatia, Estonia, Finland, Germany, Hungary, Ireland, Israel, Italy, Mongolia, Netherlands, New Zealand, Norway, Poland, Republic of Korea, Russian Federation, Slovenia, Spain, Sweden, Switzerland, The former Yugoslav Republic of Macedonia, Turkey, Ukraine, United Kingdom and United States as well as by representatives from the Eurasian Economic Commission, Eurostat, International Labour Organization (ILO), International Monetary Fund (IMF), Organization for Economic Cooperation and Development (OECD), Partnership in Statistics for Development in the 21st Century (PARIS21), United Nations Educational, Scientific and Cultural Organization (UNESCO), United Nations Industrial Development Organization (UNIDO), and the World Trade Organization (WTO).

2. Mr Steven Vale, Head of the UNECE Statistical Management and Modernisation Unit, opened the workshop and welcomed participants. He introduced Mr Rune Gløersen of Norway, who chaired the workshop.

3. The keynote presentation was given by Mr Pádraig Dalton (Ireland). He outlined the different external pressures that drive the need to modernise statistics, stressed the importance of collaboration, and summarised the work of the HLG so far.

4. The agenda included the following substantive topics, the outcomes of which are documented in the annex:

- (i) Business and IT changes that will impact statistical production
- (ii) Enterprise architecture and its role in the modernisation of statistical production
- (iii) Cost/benefits of international collaboration in our field
- (iv) Innovation in technology and methods driving opportunities for modernisation

5. All background documents and presentations for the workshop are available at <http://www.unece.org/stats/documents/2015.04.msp.html#/>

6. Key items identified for future work included:

- Developing mechanisms for sharing experiences, information, case studies and lessons learned in all areas
- Enhancing communication about the implementation of the Common Statistical Production Architecture (CSPA) at all levels, including about work in progress
- Creating guidelines on how to modernise and become CSPA compliant
- Exploring potential platforms for integrating CSPA services
- Providing more opportunities for interaction between methodologists and IT specialists
- Further exploring the potential of machine learning for statistical production

Annex: Summary of discussions on substantive topics

A. Topic (i): Business and IT changes that will impact statistical production

7. This topic was organized by Rune Gløersen (Norway), and Lorna Drennen (United States). It included the following presentations:

- Slovenia, From data store to data services - Developing scalable data architecture at SURS
- Sweden, Manage change
- USDA/NASS, Implementation of a case management service at a federal statistical agency
- Netherlands et.al, ESSnet ValiDat Foundation
- Eurostat, Modernisation of statistical production

8. The following points were raised in the discussions:

- The extent to which the ideas and solutions presented are shareable
- The merits and experiences regarding different versions of the Agile approach
- How to spread change through an organization, particularly changes in culture
- The risks associated with temporary solutions becoming permanent if they work satisfactorily

9. Small groups discussed the following topics and identified areas for future work:

Data warehouse

- Sharing experiences, identifying centres of excellence
- As a community we could formulate requirements for off-the-shelf products. It may be easier to approach vendors as a community rather than as individual organizations
- Raise awareness of standards / models – GSBPM / GSIM / CSPA
- Cooperation in methodology may be easier than in technology

Business / technology agility

- Sharing of experiences / information / lessons learned – case studies and workshops. The ability to access honest information about lessons learned is very valuable
- Make more of the LinkedIn community, e.g. the group “Modernising Official Statistics”
- Creating a recipe for modernisation
- Share examples of agile cooperation at the international level

Mobile and social media

- Have a particular project to use mobile data in production
- Build open source CSPA-compliant services for using mobile applications (collection and dissemination)
- Develop standards for using mobile data
- Global survey into use of mobile data / devices for official statistics, including types of interactions, what works and what doesn't, which channels are most successful, etc.. This should be a multi-NSI survey.

Cloud computing

- Standardized technology and semantics
- Analysis framework to identify opportunities
- Look to other industries for examples
- Update paper from ABS on current status of cloud computing in statistics

B. Keynote speech

10. Mr Pádraig Dalton, head of the Irish Central Statistics Office and Chair of the High-Level Group for the Modernisation of Statistical Production and Services (HLG) gave the keynote speech. He described the main challenges of modernisation, and encouraged participants to dream of new things and ask “why not?”

11. The following points were raised in the discussion:

- The relationship between modernisation and collaboration within organizations and between organizations
- How to quantify the benefits of collaboration
- The need to embed modernisation and collaboration activities at the organizational rather than personal level
- Branding HLG outputs
- How to convince methodologists of the need to modernise and standardize
- The importance of developing new skills and capabilities, for example data scientists
- How to define, and hence ensure, relevance in a changing world
- How to embrace and learn from failure, rather than fearing it
- How to manage standards so that they enable modernisation and support agility
- How to balance conflicting demands for change and stability
- The importance of having a dedicated budget for innovation, and collaborating to get the best results from limited resources

C. Topic (ii): Enterprise Architecture and its role in the Modernisation of Statistical Production

12. This topic was organized by Rosemary McGrath (New Zealand), and Joe Treacy (Ireland). It included the following presentations:

- Slovenia: Modernisation of statistical processing at SURS
- Turkey: Harzemli, The DDI based statistical production platform
- Eurostat: Implementing CSPA compliant guidelines to describe identified generic business needs
- Netherlands: Becoming CSPA compliant
- New Zealand: How to make EA-work impact everyday-work?
- United Kingdom: Enterprise Architecture Project of the Office for National Statistics
- New Zealand: A common methodology architecture for statistical organizations?

13. The following points were raised in the discussions:

- Whether to use existing capabilities to develop new services, or to purchase from commercial vendors
- The governance of the CSPA, and how to ensure services are truly compliant
- The development of the CSPA services catalogue, hosted by Eurostat
- How to motivate people to adopt CSPA, and make the necessary compromises at the local level to deliver benefits at the organization level

14. Small groups discussed the following topics and identified areas for future work:

Interaction with CSPA

- Joint lobbying of organizations at HLG level
- Joint lobbying of vendors to adopt CSPA
- Much better communication and visibility of the work in progress around CSPA (not just already developed modules)

- CSPA evangelists in statistical organizations - CSPA needs to be sold at the Director level, and then they can delegate CSPA responsibility to someone in organization
- Maintain international funding of CSPA initiative

International collaboration

- Set up a global enterprise architecture (using work already done by ESS)
- Define a project for doing an overall survey of all NSIs to see how they are composed / structured and their maturity
- Looking for group-ware alternatives

Need for an enterprise architecture

- Create a cookbook on how to become CSPA compliant, including stories from NSOs who have done it
- Need champions to build support in organizations for CSPA - HLG could produce business cases, information to help people in their organizations

Organizational and cultural challenges

- A group or network of architects (3-4 years ago Eurostat established a working group for IT architects similar to this that was good, but maybe it was too early)
- Knowledge base on these issues
- Advice on portfolio management when you have competing demands in production environment and EA

Managing methodological knowledge

- More opportunities for interaction between methodologists and IT specialists, nationally and internationally.

Modernisation

- Define research projects so that they involve methodology and IT
- Encourage a harmonious vision from lowest level to highest level and have leadership
- Develop something like the GSBPM for methodology - to help identify overlaps
- Leveraging Lean 6 Sigma as a tool to find efficiencies.

15. The topic also included a special session on recent developments with the CSPA, focusing on the outcomes of recent “sprint” sessions in Ottawa and Canberra. The main issues raised in the discussion were:

- How best to communicate CSPA within statistical organizations
- How to manage versioning of CSPA, though it was noted that new versions tend to add more guidance rather than significantly changing the underlying architecture
- Whether organizations would sign-up to the proposed manifest. Chief statisticians will be invited to sign-up at the Conference of European Statisticians in June, but it would help if they could be briefed beforehand
- Whether CSPA is as appropriate for research activities as it is for regular production
- Statistical organizations were invited to express interest in leading or participating in the development of new services, focusing on, but not strictly limited to, those on the candidate list. Organizations interested as implementers of the proposed services were also encouraged to express an interest
- The varying levels of granularity in the candidate list
- The need for consistent and user-friendly interfaces to services, including multi-lingual support
- The need to develop further layers of the capability model (based on the work of the European Statistical System Enterprise Architecture Task Force)
- The best way to identify priority services to be developed

- Limited availability of support for shared services may be perceived as a barrier
- The potential application of the new Validation and Transformation Language (VTL) in validation services
- Whether candidate services should focus on current needs or new types of needs expected in the future

D. Panel Discussion: Cost/benefits of international collaboration in our field

16. A panel consisting of Ms Gillian Nicoll (Australia), Mr Robert McLellan (Canada), Mr Carlo Vaccari (Italy) and Mr Trevor Fletcher (PARIS21), and moderated by Mr Steven Vale (UNECE) gave their views on this topic. Points raised in the discussion included:

- We will reduce costs through international collaboration but it is a long term investment
- We must remember where we started. We have made considerable progress by combining our resources. When valuing international collaboration we really need to look at what we have done. If we want to get further we need a bigger commitment from our organizations
- Developing countries are a perfect environment to spread standards. They do not have so many existing systems, so it is easier for them to accept change
- The Big Data sandbox is an excellent example of collaboration, providing a resource that few statistical organizations could afford on their own
- The importance of training, rather than just providing tools
- The risks and benefits of pooling capabilities and creating centres of excellence. This approach may be most appropriate for new capabilities, e.g. processing Big Data
- It should be “illegal” to work on new things unless they are based on the principles that we have created e.g. CSPA
- Outcomes of international collaboration are more likely to be adopted by national cross-government initiatives than proposals that just come from the national statistical organization
- We need to ask “why not?” international collaboration, rather than “why?”
- Collaboration is about creating opportunities as well as reducing costs
- The role of the UNECE in facilitating and motivating international collaboration was acknowledged. This should extend to managing an international investment fund. However the UNECE needs more resources for this work, to ensure greater continuity and engage in more communication activities. With more resources, UNECE could seek larger amounts of funding from external sources.

E. Topic (iii): Innovation in technology and methods driving opportunities for modernisation

17. This topic was organized by Marton Vuksan (Netherlands), and Carlo Vaccari (Italy). It included the following presentations:

- Spain: Data collection and selective data editing in a systematized and integrated way. An experience in progress in INE
- Republic of Korea: Progress and future plans for big data use at Statistics Korea
- Irish Centre for High-End Computing: The UNECE Big Data Sandbox: What Means to What Ends?
- Canada: Machine learning documentation initiative
- UNIDO: Using the Software Environment R in Official Statistics and Survey Methodology
- Italy: CORE: an implementation of CSPA

18. The following points were raised in the discussions:

- The increasing pressure to validate and impute at the time of collection
- We are at the early stages of exploring the use of Big Data, further work is needed in areas such as quality measurement and data protection
- Should national statistical organizations have a role of data curators for all data holdings across government?
- The extent to which the systems and ideas presented are shareable across organizations
- Whether it is a good idea to buy data from third parties, or whether this sets a dangerous precedent
- The impact of new approaches on organizational structures
- The costs of setting up a sandbox-like environment are mostly associated with support and having the right expertise in the context of rapidly changing technology

19. Small groups discussed the following topics and identified areas for future work:

General opportunities

- Developing the methods and skills to work with new data sources
- Regular meetings organized according to the GSBPM structure
- Standards for machine to machine transfers of open data

Data input processing

- A unified approach for outlier detection and management
- Development of record linkage approaches including mapping and coding
- Document best practices on data input processing according to GSBPM sub-process and statistical domain
- Assessing the impact of imputation processes on the data

Big Data

- Develop new methodologies
- Further development of the Sandbox initiative
- Cooperation with academics to develop new methodology and define potential areas for use in official statistics
- Cost/effectiveness analysis of investments into Big Data collection, processing, storing and analysis technologies.

Machine Learning

- Explore intelligent data cleaning
- Assess potential for forecasting and analysis in areas such as national accounts and the Euro-Groups Register
- Publish the detailed results of the survey used for the Canadian paper as case studies
- A competition on an imputation problem or something similar
- Greater use of synthetic data to test methods with external partners
- Develop a synthetic data generator
- Consider using data mining to support key statistics such as labour and National Accounts
- Create intelligent systems to hunt for metadata in “data lakes”
- Include machine learning in the Big Data project
- Partner more with the scientific world, e.g. in the use of satellite data

CSPA: The next level

- Further develop the repository of CSPA components, including governance issues All kinds of licensing arrangements should be supported
- More work on orchestration for CSPA services, drawing on the work done by Italy and Eurostat