Modernisation Committee on Production and Methods

Members

- Gillian Nicoll (Australia)
- Claude Poirier (Canada)
- Rob McLellan (Canada)
- Gergely Horvath (Hungary)
- Joe Treacy (Ireland)
- Alessio Cardacino (Italy)
- Marton Vucsan (Netherlands) (Co-chair)
- Rosemary McGrath (New Zealand)
- Rune Gløersen (Norway) (Co-chair)
- Janusz Dygaszewicz (Poland)
- Pedro Revilla Novella (Spain)
- Christine Wirtz (Eurostat)
- David Barraclough (OECD)
- Valentin Todorov (UNIDO)

Terms of Reference

Suggestions for Key issues and priorities

- Regarding future work, we think that it is about time to start working on methodology. Make inventories, identify best practices, look at implementation issues within the CSPA etc. We also think that we should identify roles and competencies that we need, and try to harmonize descriptions etc. One example is roles like business or enterprise architects. Equally we should perhaps do something like this with best practice working processes? The reason is basically to provide for better understanding between partners and within projects, to help cooperation succeed.

- Looking at key issues in our back yard we might consider not only thinking about a more efficient or better production process, but also in terms of production value chain looking at what we can make others to produce for us. Until now we have been looking at the improvement of the currently defined production. I think we should also address the question of whether we see a paradigm change and if we need to think about defining a changing position in the information chain. My point is this: There is ever more pressure to reduce costs. We will certainly not have the money to continue with our traditional production. I do not think we will have the money to process all the big data we need for an adequate statistical picture of our society. We need to look into this and think about ways to achieve results in a maybe fundamentally different way. I can see a future in which we no longer will process micro data but we could “get” or buy aggregates of various levels and extract quality by combining them or managing the quality in the manufacturing contracts. It is one of the lessons of big data; it is the emerging patterns that is the data, the individual records do not have much meaning. In my view there is no fundamental reason that a statistical agency should use micro data; our job is to provide aggregates and to explain. Research in this area might not be a candidate for top priority but I think we should do some work on it.

- Methodology has been identified as a 'gap' in all of the foundation projects, what does that gap mean, is it perceived differently between methodologists and non-methodologists? Before we dig too deep and solve the wrong thing it would be great to understand the mechanisms by which Statistical Methods are currently agreed (or not as the case may be) and implemented - what is needed to support international agreement on best practice for statistical methodologies and where would we focus first - this ties into some of the suggestions below. Also, do we need to address the question of statistical methodologist vs data scientist (I know it's a buzz word, but does it really mean a step change or is it an additional capability?)

- We all need to “do more with less”, and many of our organisations, public services, are reorganised. How to change our production methods to face the challenges by competitive pressures such as the growing volume of statistics published by private companies and the advent of new data sources, including big data?
- What changing skillsets are needed by statistical offices today (e.g. data engineers, blending of IT and other disciplines)?
- Collaboration between disciplines (e.g. statistics, IT and methodology);
- External or international collaboration to source skills and knowledge.
- Streamlining statistical business processes
- How IT services are configured and delivered (e.g. potential of shared services / cloud, AGILE delivery).
- Change management

Ai Determine what further technical elements are still required for CSPA eg:

- · develop consistent process design approaches and tools
- · develop standards based editing rules
- · agree how to develop CSPA aligned services - governance/guidelines/requirements
- · develop a global catalogue of assets, incl process design catalogue of assets
- · defining what a statistician/assembler (the CSPA user role) workbench may look like when we are assembling/running services

Why – meets the need of the CSPA project

Aii How would we integrate our highly technical (CSPA) solutions into our BAU world. Which are the early adopters to test elements for BAU adoption:

- · Metadata; Data; Methods; Assembling and configuring systems

Why – meets the need of the CSPA project

Aiii What are the next practical steps for CSPA to enable sharing of IT statistical solutions. What else needs to be in place e.g. licensing regimes and/or risk approaches agreed, technical mechanisms for sharing such as a catalogue of sharable solutions? What are the first cabs of the rank for sharing?

Why – meets the need of the CSPA project

Bi Progress the Big Data initiatives

- Develop a taxonomy of the different types of big data
- Identify the statistical potential of each type of data
- Start understanding quality issues, and determine fitness-for-purpose
    - Agree the methodological challenges and opportunities arising from big data, and determine how these should be resolved:
    - · How valid will the results be (eg the acquisition of mobile phone data from private industry are significantly different to those associated with the acquisition of satellite imagery from government sources)
    - · Assess and develop methods of analysis of Big Data (detection and description of statistical structure)

Why – meets the need of the big data project

Bii Determine the opportunities for sharing Big Data IT designs and solutions (including IT strategies, investigation and implementation)

Why – meets the need of the big data project

C What other methodology issues are priorities:

- · accounts compilation - methods for detecting anomalous values and reconciling 'random' deviations from account constraints
- · Machine learning -application of supervised machine learning to manual processes eg coding and resolving uncertain linkages
- · Responsive Survey Design - for the data collection process, develop and assess methods for assessing the quality, cost and resource allocation in real time and adjusting the collection process to optimise the balance between these.
- · How can we better support rapid updating of standards and classifications such as SNA, ISIC etc

Why – greater efficiency and effectiveness as we undertake statistical processing

D Other IT issues

Cloud Computing and ICT Infrastructure Futures - What are the opportunities/threats? What is the future for technical infrastructure for statistics given government drivers around the world for things such as shared services and ICT efficiencies via the cloud. What are the experiences of others and what are their plans?

- For IT - I like the suggestions made so far - A question - why the split?

Presentation given at November 2013 HLG

workshop