



Statistics Canada

www.statcan.gc.ca

Confidentialized Analysis of Microdata CSPA project

Michelle Simard, Thérèse Lalor

Statistics Canada CSPA Project Manager

UNECE Work Session on Statistical Data Confidentiality

Helsinki, October 2015



Statistics
Canada

Statistique
Canada

Canada

Outline

- Background
- High-Level Group (HLG) Structure
- Confidentiality groups
- From ABS to StatCan
- Next Steps

Background

- Generalized systems traditionally developed by each National Statistics Organisation (NSO)
 - Ad hoc and minimal sharing between countries
 - With recent financial constraint
 - No longer sustainable to work all by ourselves
- ➔ NSO need to help each other more comprehensively

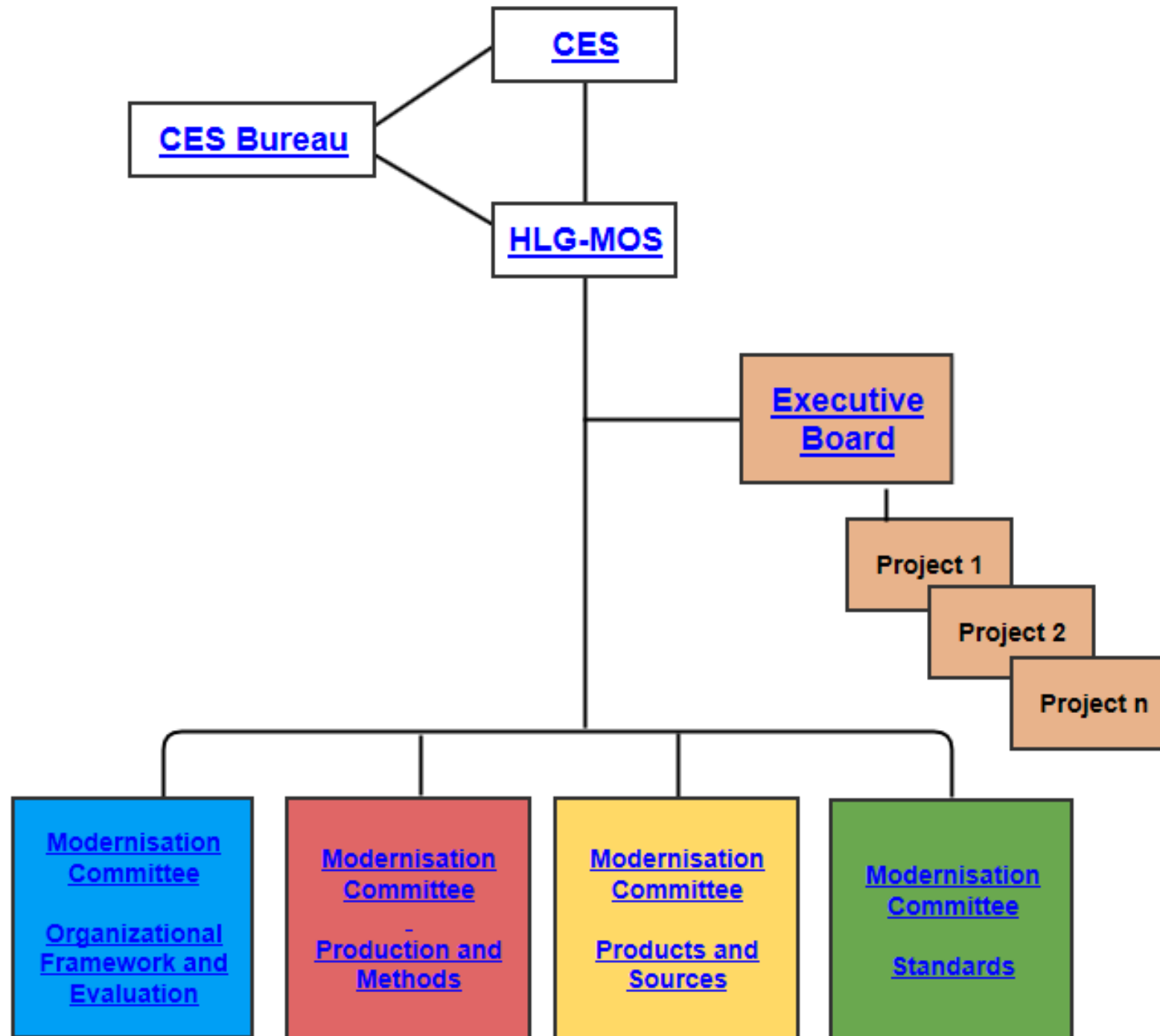
The UNECE High-Level Group for the Modernisation of Official Statistics (HLG-MOS)

- Set up by the Bureau of the Conference of European Statisticians (CES) in 2010
- Coordinates international work
- Promotes standards-based modernisation of statistical production and services
- The missions
 - Overseeing development of frameworks, and sharing of information, tools and methods
 - Improving the efficiency of the statistical production process

Background

- Key: using the same concepts/ languages
- Developed common frameworks
- Some results:
 - GSBPM (Generic Statistical Business Process Model)
 - Decomposition of all processes in a statistical organisation
 - CSPA (Common Statistical Production Architecture)
 - Decomposition of common IT components
 - GSIM (Generic Statistical Information Model)
 - Decomposition of common information / metadata
 - GAMSO (Generic Activity Model for Statistical Organisations)

HLG Structure



HLG Structure

- Participating members:
 - UNECE, Eurostat, OECD, Ireland, Australia, Canada, Italy, Netherlands, New Zealand, Republic of Korea, Slovenia, UK, Sweden, Finland, Norway, Mexico...
- Annual meeting, teleconferences and wikis, sprints and sandbox

Note: As of 2015, the Statistical Network activities are incorporated into the HLG

HLG Projects – CSPA Implementation

- Standardisation of the infrastructure - common architecture (CSPA work)
 - Implemented on NSO common generalised tools such as editing, sampling, coding, linkage, confidentiality, etc...
- Creation of a common “language” (wrapping)
- The idea
 - Make it CSPA-compliant
 - Then share with other countries
- One country may have developed a tool, another one will be the “wrapper”, others will use the wrapped tool

Confidentiality

- Started in 2014
 - Statistical Network Innovation in Dissemination (SNID)
- Australia, Norway, Italy, UK, Canada
- Exchanges on systems and methods for confidentialized output tool
- Lead: Australian Bureau of Statistics (ABS)

Confidentialized Analysis of Microdata

- Through CSPA
 - Partnership ABS (builder) – StatCan(wrapper)
- ABS DataAnalyzer was imported and made functional in StatCan environment
- Architect built a ‘Confidentialized Analysis of Microdata’ CSPA service
 - Removed the ABS “outside layers” and connections
 - Kept the engine – statistics and confidentiality
 - Wrap a CSPA compliant architecture around the “engine” for easy recycling to other countries
 - Then for StatCan use, they developed a internal GUI prototype (not web-based)

ABS DataAnalyzer

- Online product with an Interface (ABS GUI)
- Web-based
- Explore (tabulate), manipulate, and analyse microdata
 - Linear Regression Model
 - Generalized Linear Model
 - Multivariate Model
- Confidentiality of outputs (diagnostics and model parameters)
- Privacy of individuals' data kept

ABS DataAnalyzer

- All outputs (tabular or graphics) are confidentialized
- Perturbation is the main method of protection; adding random noise to any estimates
- Perturbation is used in Tables (counts and means) and Regressions (coefficient estimation)
- Other methods used:
 - Sparsity, Field Exclusion Rules, Range Restrictions, Dropping Units, Suppression of Small Counts, X-only Variables, Leverage Protection
- 2011 UNECE worksession, Ottawa

Implementation

- Canada, New Zealand, Australian and Finland are implementing the Confidentialized Analysis of Microdata CSPA service
- Canada – How does it work:
 - User submits code through GUI/ functions/options (prototype) - StatCan codes
 - Code of the “engine” parsed, validated executed - ABS codes
 - Results are parsed, validated and returned back to the user – StatCan codes
- Early evaluation and assessment of the tool
 - Many years of development and resources saved by the organisation
 - Remaining issues
 - IT architecture and methodological

Canada

- Discussion and coordination about the potential integration of the tool within StatCan
 - Generalised system, remote access, common tool, internal tool for economists, analysts, others...
 - StatCan RTRA is a “remote submission” using a FTP
 - Ideally build a web-based application
 - Evaluate the informatics infrastructure and how it fits with StatCan architecture

Canada- Next steps

- Evaluate methods for calculating statistics and model diagnostics
 - If applicable, propose modifications, testing and implementation
- Evaluate coherence between ABS SDC methods and StatCan SDC methods (including graphics)
- Validation and Approval processes for methods (statistical procedures and disclosure controls)

Concluding remarks

- Very exciting and promising
- StatCan definitely will find some use

- More information: Confidentialized Analysis of Microdata
CSPA Service on UNECE wiki