

# **Workshop on Implementing Standards for Statistical Modernisation 2016**

**Geneva, 21-23 September 2016**

## **Complementing the GSBPM with Quality Indicators for admin data and mixed sources**

Joint work of Modernisation Committee on Standards and  
Modernisation Committee on Production and Methods

# Members

## From the Modernisation Committee on Standards

**Alice Born and Kaveri**

**Mechanda** (Statistics Canada)

**Jan Planovsky** and **Jean-Pierre**

**Poncelet** (Eurostat)

**Nilgün Dorsan** and **Deniz Ozkan** (Turkstat)

**Marina Signore** and **Giovanna Brancato**  
(ISTAT)

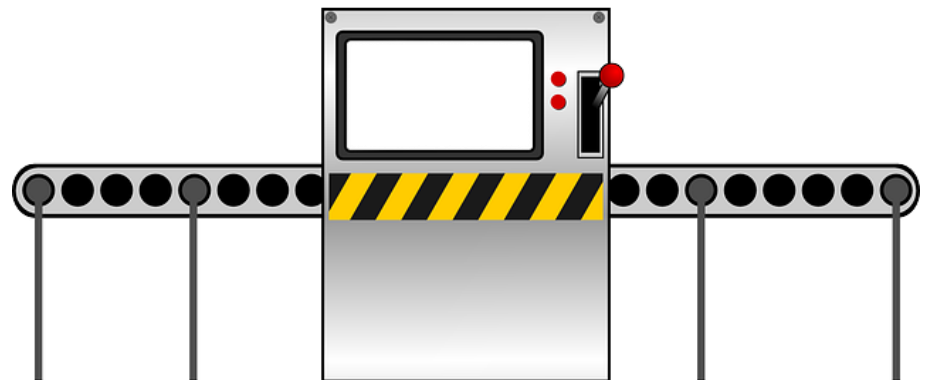
## From the Modernisation Committee on Production and Methods

**Claude Poirier** and **Laurie**

**Reedman** (Statistics Canada)

**Felipa Zabala** and **Anapapa Mulitalo**

(Statistics New Zealand)



# Presentation

- ✓ Developing QIs for GSBPM phases and sub-processes
- ✓ 2014 – 2015: Mapping QIs to the GSBPM for survey-derived statistics
- ✓ 2016-2017: Extending the mapping of QIs to admin data and mixed sources

# Survey-derived statistics

Quality indicators were developed for the Generic Statistical Business Process Model (GSBPM) with the aim of expanding the quality management layer for the GSBPM

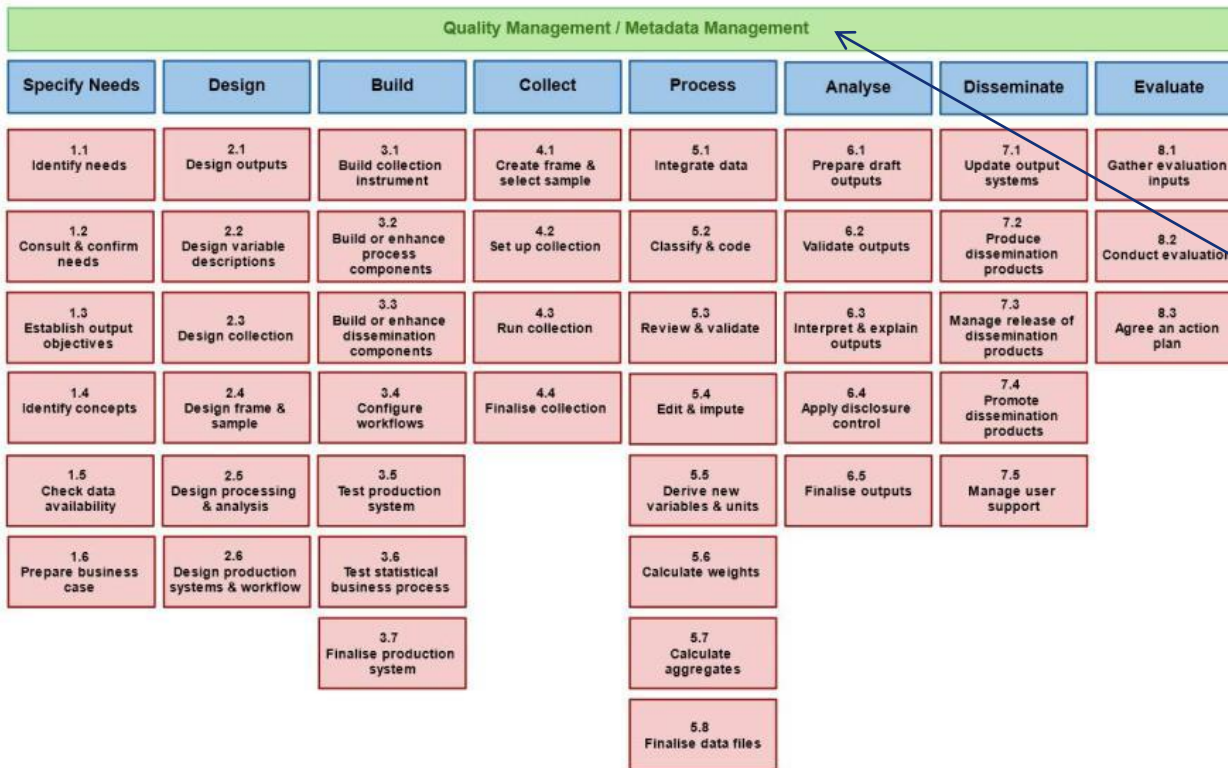


**Quality Indicators for the Generic Statistical Business Process Model (GSBPM) - For Statistics derived from Surveys**

(Version 1.0, May 2016)

- Version 1.0 of Quality Indicators for the GSBPM was released May 2016
- <http://www1.unece.org/stat/platform/pages/viewPage.action?pageId=123142969>
- Task team: Statistics Canada, Istat, Turkstat and Eurostat
- Potential users and stakeholders were consulted on the GSBPM QIs in 2015 via
  - i) a group work session during the Standards-Based Modernisation Workshop,
  - ii) an open consultation on the UNECE website.

# Survey-derived statistics



Quality management overarching process

Quality indicators were developed for each phase (1 to 8) and sub-processes of GSBPM

# Developing QIs for Survey-derived statistics

- ✓ Generic indicators
- ✓ Coherence with existing frameworks:
  - UN National Quality Assurance Framework (NQAF), EU CoP, ESS Q&P Indicators,  
National quality assurance frameworks (e.g. Statistics Canada Quality Guidelines ...)
- ✓ Quantitative indicators whenever possible
- ✓ Qualitative indicators
  - yes/no
  - low/medium/large



Personalisation of the indicators left to NSIs, e.g. setting targets or levels

Quality Management / Metadata Management							
Specify Needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Build collection hardware	4.1 Create frame & define sample	5.1 Ingest raw data	6.1 Prepare arch outputs	7.1 Update output coverage	8.1 Gather & evaluate feedback
1.2 Contact & confirm needs	2.2 Design variable descriptions	3.2 Build or enhance process capabilities	4.2 Set up collection	5.2 Classify & code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.3 Establish input relationships	2.3 Design collection	3.3 Build or enhance dissemination capabilities	4.3 Run collection	5.3 Review & v/criteria	6.3 Interpret & explain outputs	7.3 Manage release or dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design theme & sample	3.4 Configure hardware	4.4 Prepare extraction	5.4 Edit & merge	6.4 Apply metadata control	7.4 Produce dissemination products	
1.5 Check data availability	2.5 Design processing & storage	3.5 Test production system		5.5 Derive new variables & links	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare hardware data	2.6 Design production systems & extraction	3.6 Test operational business process		5.6 Calculate weights			
	2.7 Produce dissemination system			5.7 Calculate aggregates			
				5.8 Finalise data files			

# Admin data and mixed sources

Extending QIs for GSBPM to the use of admin and mixed sources was one of the priorities for the WG in 2016-2017 as emerged from the consultation process in 2015

## Key features

- ✓ Build on existing work and frameworks
- ✓ Coordination with ongoing activities
- ✓ Follow similar guiding principles as for developing QIs for survey-derived statistics



Work is still in progress

# Literature review

- **Many experiences from NSIs**
  - ❖ e.g. Stats Netherlands Checklist, Stats Canada Guidelines, Istat Guidelines and assessment questionnaire, Statistics New Zealand's Guide to reporting on administrative data quality, ...
- **International/ European experiences**
  - ❖ MIAD: Methodologies for an Integrated Use of Administrative Data in the Statistical Process (2013 – 2014)
  - ❖ EU FP7 Blue-Ets: BLUE-Enterprise and Trade Statistics (2010-13)
  - ❖ Essnet Admin Data: Use of Administrative and Accounts Data for Business Statistics (2009-2013)
- **On-going activity at International/ European level**
  - ❖ 2016 HLG-MOS project on Data Integration
  - ❖ Essnet Quality of multisource statistics (2016-2020)



# Use of admin data

## Administrative Body



## NSI



## Statistics from admin sources



### **Administrative process not under the NSI control**

- strategy to increase **USABILITY** for statistical purposes

### **Data Acquisition**

- relationships with data owners, protocols, ....

### **Input Quality**

- analysis of data source; metadata availability; ...

### **Different usages**

- admin data based statistics, mixed sources based statistics, indirect usage

### **Input, throughput, output quality**

- more focused input quality indicators
- errors generated during the statistical process
- quality of the estimates

# Usage Mapping

The MIAD project identified the following usages of admin data sources for statistical purposes

## DIRECT

1. Direct tabulation (for full coverage admin sources)
2. Substitution and Supplementation for Direct Collection

## INDIRECT

3. Creation and maintenance of survey frames
4. Construction of sampling designs
5. Editing and imputation
6. Indirect estimation and weighting
7. Data validation/confrontation

Quality Management / Metadata Management							
Specify Needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Build collection hardware	4.1 Create forms & write scripts	5.1 Import raw data	6.1 Prepare arch outputs	7.1 Update output content	8.1 Gather evaluation plans
1.2 Consult & confirm needs	2.2 Design variable descriptions	3.2 Build or enhance physical infrastructure	4.2 Set up collection	5.2 Classify & code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design collection	3.3 Build or enhance management infrastructure	4.3 Run collection	5.3 Review & validate	6.3 Interpret & explain outputs	7.3 Manage release of dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design theme & sample	3.4 Configure hardware	4.4 Prepare collection	5.4 Edit & update	6.4 Apply metadata control	7.4 Produce dissemination products	
1.5 Check data availability	2.5 Design processing & storage	3.5 Test production system	4.5 Derive new variables & links	5.5 Finalise outputs	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare business case	2.6 Design production systems & workflow	3.6 Test operational business process	4.6 Calculate weights	5.6 Calculate weights	6.6 Calculate aggregates	7.6 Prepare data files	
	2.7 Prepare production system	3.7 Prepare production system	4.7 Prepare data files	5.7 Prepare data files	6.7 Prepare data files	7.7 Prepare data files	

# Developing QIs for admin data and mixed sources

- ✓ Review of the GSBPM descriptions for phases and sub-processes to verify that the use of admin data and mixed sources is duly accounted for
- ✓ Screening of QIs already mapped to the GSBPM for survey-derived statistics that are also meaningful for admin data and mixed sources
- ✓ Developing additional QIs for admin data and mixed sources **taking into account the different usages**



Some examples are presented in next slides

# QIs for Admin data and mixed sources

1

Quality Management / Metadata Management									
Specify Needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate		
1.1 Identify needs	2.1 Design outputs	3.1 Build collection instrument	4.1 Create frame & select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs		
1.2 Consult & confirm needs	2.2 Design variable descriptions	3.2 Build or enhance process components	4.2 Set up collection	5.2 Classify & code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation		
1.3 Establish output objectives	2.3 Design collection	<b>Quality Dimension</b>					<b>Indicator</b>		<b>Notes</b>
1.4 Identify concepts	2.4 Design frame & sample	Statistical confidentiality and security					To what extent have legal constraints regarding statistical outputs been considered, for example but not limited to ensuring confidentiality of data and preventing the disclosure of sensitive information?		
1.5 Check data availability	2.5 Design processing & analysis	Relevance					To what extent have all statistical needs been addressed by the proposed outputs?		
1.6 Prepare business case	2.6 Design production systems & workflow	Accuracy and reliability					To what extent are the proposed outputs and their quality measures suitable to user needs?		

# QIs for Admin data and mixed sources

2

## Quality Management / Metadata Management

Specify Needs		Design	Quality Dimension	Indicator	Notes
1.1 Identify needs	2.1 Design outputs		Cost effectiveness	To what extent is the process planning to re-use systems for coding, E&I, data integration, weighting, estimation	
1.2 Consult & confirm needs	2.2 Design variable descriptions		Soundness of implementation	To what extent is the business process using standard or well-known methods for subsequent phases (e.g. coding, E&I, data integration, weighting, estimation, revision,...), in a transparent way?	See also phase 5 and 6 yes/partly/no indicator Corresponds to the appropriate statistical procedures principle in the ES Code of Practice
1.3 Establish output objectives	2.3 Design collection				
1.4 Identify concepts	2.4 Design frame & sample		Soundness of implementation	When have the methodologies for subsequent phases (e.g. coding, E&I, data integration, weighting, estimation,...) last been assessed?	See also phase 5 and 6 for outputs produced on a regular basis
1.5 Check data availability	2.5 Design processing & analysis				
1.6 Prepare business case	2.6 Design production systems & workflow		Soundness of implementation	Specifications for coding, editing, imputing, estimation, integrating, validating and finalizing datasets take into consideration the type of data being processes, i.e. respondent data or ADS or a combination.	Take into consideration of ADS in the process, including specification of routines for coding, editing, imputing, estimating, integrating, validating and finalising data sets of ADS data.

# QIs for Admin data and mixed sources

4

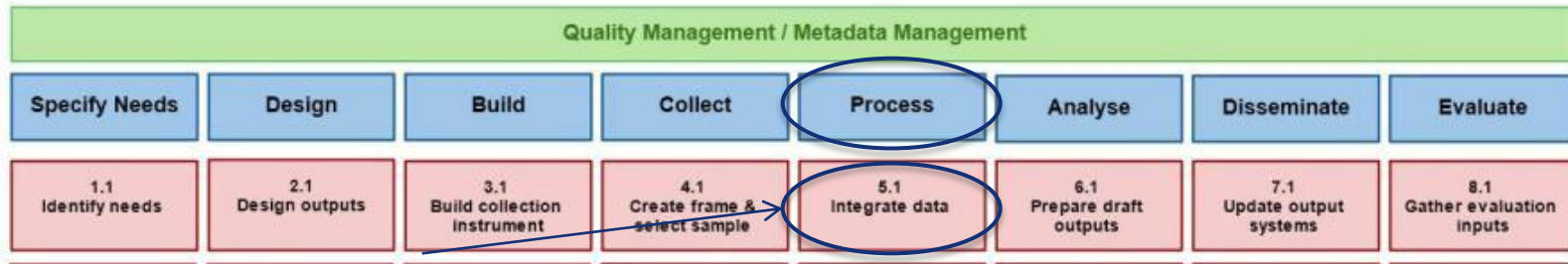
## Quality Management / Metadata Management

Specify Needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Build collection instrument	4.1 Create frame & select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult & confirm needs	2.2 Design variable descriptions	3.2 Build or enhance process components	4.2 Set up collection	5.2 Classify & code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.3 Establish output	2.3	3.3 Build or enhance	4.3 Run collection	5.3 Review & validate	6.3 Interpret & explain	7.3 Manage release of	8.3 Agree an action

Quality Dimension	Indicator	Notes
Accuracy and reliability	Extent to which admin data supplement direct collection (e.g. % or records from admin sources; % of variables from admin sources)	
Accuracy and reliability	Extent to which administrative data sources are used to create/ maintain or assess the quality of the frame	
Accuracy and reliability	Extent to which administrative data sources are used as auxiliary variables in the construction of the sampling designs	Also in the design phase

# QIs for Admin data and mixed sources

5



Quality Dimension	Indicator	Notes
Accuracy and reliability	<p>If record linkage is required, percentage of records that were successfully matched</p> <p>Linkage rate - the proportion of objects in each dataset which can be connected with units on the other datasets</p>	
Accuracy and reliability	<p>Reliability of the linkage results</p> <p>False positive and false negative rates</p> <p>Precision and recall</p> <p>An indicator of the effectiveness of the cut off weight for determining the threshold of passes in probabilistic matching</p> <p>Proportion of duplicated records in linked data</p>	

## Concluding remarks

- The QI study requires a good process knowledge of GSBPM
- Each of the sub-processes needs to be reviewed carefully as some of the sub-processes have natural links with each other
- In the forthcoming reviews of the GSBPM, we need to think about the link between GSBPM, GAMS0 and also the quality indicators
- As this study is important for measuring the quality of the processes and products and will be used for a resource document, the feedbacks from international community is welcome
- Next steps : QI-GSBPM version 2.0 to be approved by MC Standards and presented at next Workshop (2017)