United Nations Economic Commission for Europe
Statistical Division

Using the GSBPM in Practice

Steven Vale
UNECE
steven.vale@un.org

Introduction

- The original aim of the GSBPM was to standardise terminology for discussions on statistical metadata systems and processes
- But now it has many other uses:

  (Quotes from GSBPM 5.1)

Documentation

- “The GSBPM can provide a structure for organising and storing documentation within an organisation, promoting standardisation and the identification of good practices”

Process quality management

- “If a benchmarking approach to process quality assessment is to be successful, it is necessary to standardise processes as much as possible. The GSBPM provides a mechanism to facilitate this”

Integrating metadata and quality

- “The common framework provided by the GSBPM can help to integrate international work on statistical metadata with that on data quality by providing a common framework and common to describe the statistical business process”

Mapping Statistical Production Processes to GSBPM

- All processes that result in data outputs can be mapped to the GSBPM

Example 1

- Australian project to improve production of prices data
- Identified activities within GSBPM sub-processes
- Some phases “out of scope”

Australian Bureau of Statistics

Prices System Improvement Project

Prices concordance of high level 'to-be' business processes to the GSBPM v 4.0

version 0.9 - 20 January 2011

Quality Management / Metadata Management

1 Specify Needs

1.1 Determine needs for information

1.2 Consult and confirm needs

1.3 Establish output objectives

1.4 Identify concepts

1.5 Check data availability

1.6 Prepare business case

2 Design

2.1 Design outputs
- Design Price Index characteristics
- Design Price Index classification/s (structures)
- Design weighting patterns
- Design Price Index measures
- Design Price Index products

2.2 Design variable descriptions
- Design variables collected via data collection instrument
- Design derived variables and transformation formula

2.3 Design data collection methodology
- Determine suitable methods for data collection
- Design data collection instrument
- Design formal agreements to collect data
- Design provider management method

2.4 Design frame and sample methodology
- Desing survey frame methodology
- Design survey sample methodology

2.5 Design statistical processing methodology
- Design statistical method for
  - integrating data
  - validating data
  - imputing data
  - calculating aggregates
  - micro editing
  - macro editing
  - classifying and coding data
  - calculating weights
  - finalising data

2.6 Design production systems and workflow
- Determine the work flows from data collection to dissemination of outputs
- Define criteria to assess the quality of the production systems and work flows
- Design systems /work flows integration, migration and roll forward processes
- Determine fitness for purpose of existing production systems and work flows
- Undertake gap analysis to determine re-use of existing systems and work flows

3 Build

3.1 Build data collection instrument

3.2 Build or enhance process components

3.3 Configure work flows

3.4 Test production system

3.5 Test statistical business process

3.6 Finalise production system

4 Collect

4.1 Select sample
- Establish frame according to frame design
- Select sample according to sample design
- Establish maintenance procedures for frame and sample
- Maintain frame and sample

4.2 Set up collection
- Investigate and establish collection strategy for sample and administrative data
- Establish training regime for staff to enable data collection
- Check data collection tools, technology and processes are available and ready for use
- Establish and review security procedures for data collection
- Prepare data collection systems for collection and receipt of data

4.3 Run collection
- Establish provider contact procedures and SLA's
- Collect data according to schedules
- Follow up with providers for data
- Record provider contact information and response rates
- Record and resolve queries by providers and data collection staff

4.4 Finalise collection
- Convert data for data loading
- Upload data into processing systems
- Load metadata into metadata storage systems
- Archive paper forms, and administrative data sets

5 Process

5.1 Integrate data
- Match data records
- Prioritise data records
- Eliminate duplicate data records
- Anonymise data records

5.2 Classify and code
- Classify data according to pre-defined collection requirements and definitions
- Assign codes to data items according to classification

5.3 Review, validate and edit
- Detect and treat all Quality Adjustments
- Detect and treat all significant anomalous data

5.4 Impute
- Logical imputation
- Historical imputation
- Seasonal imputation
- Subjective imputation
- Mean imputation
- Regression
- Donor Imputation

5.5 Derive new variables & statistical units
- Derive a final price for reference data
- Derive a final price for pre-processed data
- Derive a final price for all Price Observations collected
- Derive a current Price Relative for all Price Observations which have a Base period Price recorded

5.6 Calculate weights
- Calculate Reporting Unit sample weight
- Calculate Price Observation sample weight

5.7 Calculate aggregates
- Calculate the Raw Index, C Index, Average Price and Median for Price Samples using assigned Compilation method
- Price update previous period value aggregates for all elementary aggregates
- Sum all child value aggregates for each upper level Price Index component

5.8 Finalise data files
- Calculate all additional aggregate data for Price Index (P Indexes, points contribution & change, percentage change)

6 Analyse

6.1 Prepare draft outputs
- Record quality characteristics for each Price Index number produced
- Produce data visualisation outputs

6.2 Variable outputs
- Validate Price Index outputs against body of knowledge
- Validate Price Index quality characteristics against a quality framework
- Macro analyse Price Index outputs
- Investigate inconsistencies between macro data and body of knowledge

6.3 Scrutinise and explain
- Inspect macro statistical outputs
- Explain macro statistical outputs against body of knowledge

6.4 Apply disclosure control
- Assess likelihood of identification of data
- Recommend protection techniques for data
- Apply protection techniques to data

6.5 Finalise outputs
- Apply consistency checks
- Produce clearance documentation
- Set level of data release
- Conduct clearance meetings with senior management
- Clear data for release

7 Disseminate

7.1 Update output systems
- Load data and metadata to output data stores
- Resolve issues
- Validate and sign off

7.2 Produce dissemination reports
- Prepare draft of dissemination products
- Finalise dissemination products
- Validate and Sign off release of dissemination products

7.3 Manage release of dissemination products
- Brief authorised stakeholders

7.4 Promote dissemination products
- Release dissemination products
- Release other products

7.5 Manage user support
- Record Query in Single repository
- Categorise/Link Query by Release Period/Index/Component
- Resolve Query
- Refer to Delegate
- Advise response to client
- Record response

8 Archive

8.1 Define archive rules

8.2 Manage archive repository

8.3 Preserve data and associated metadata

8.4 Dispose of data and associated metadata

9 Evaluate

9.1 Gather evaluation inputs
- Determine persons / team to conduct evaluation
- Gather inputs required for evaluation

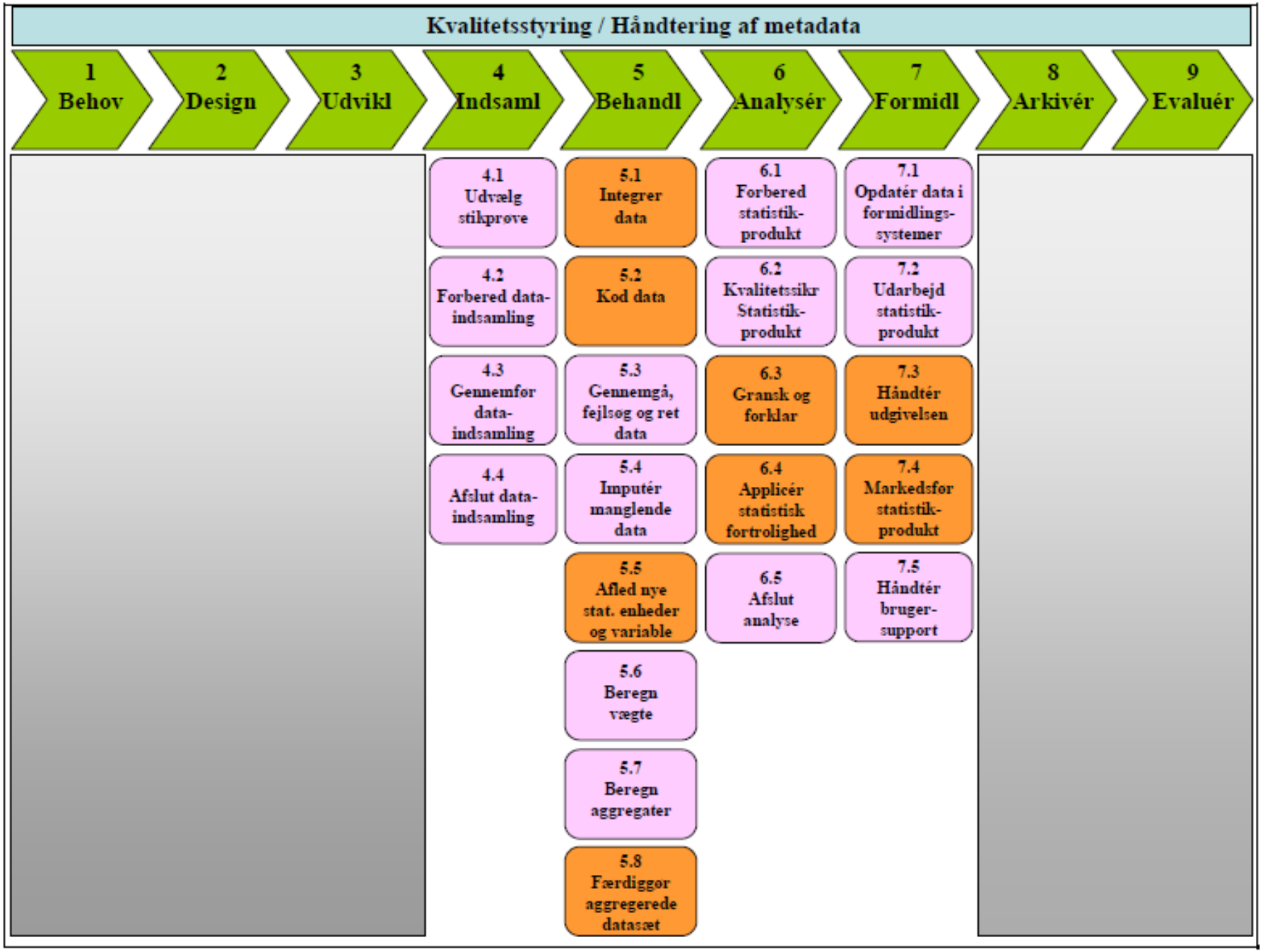
9.2 Conduct evaluation
- Conduct detailed analysis and evaluation of all gathered inputs
- Produce report detailing finding, and recommendations for improvement

9.3 Agree on action plan
- Present evaluation report to appropriate corporate consultative boards for discussion
- Agree on action plan for either implementing or amending the proposed recommendations
- Set up metrics to monitor the success and benefits derived from implementing recommendations

Example 2

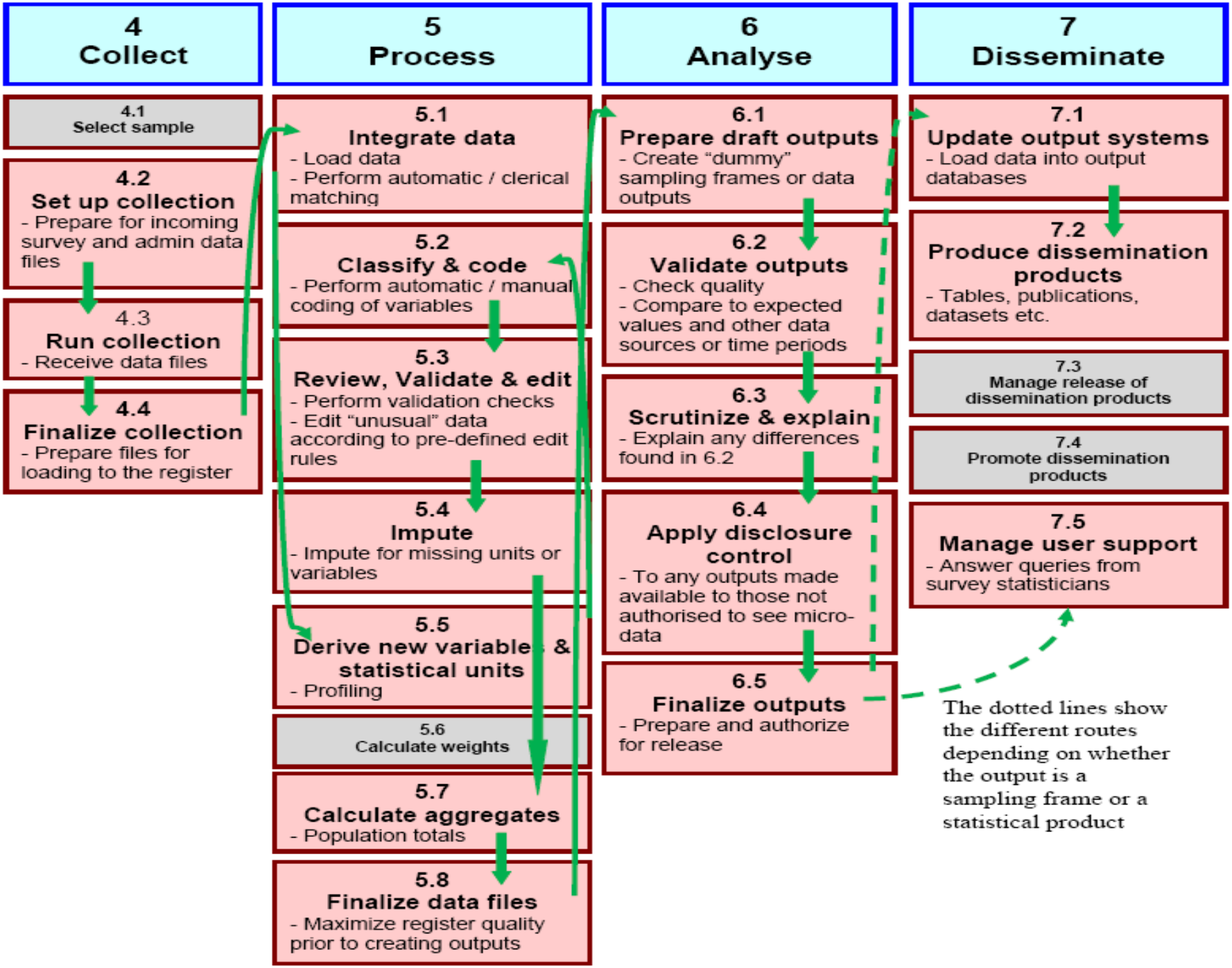
- Denmark – quarterly survey on employment in construction
- Established, regular survey
  - Just phases 4 to 7

METIS phases for: Employment in construction industries (marked with pink)



Example 3

- Mapping register processes to GSBPM
- Register maintenance is a continuous activity, not a single process
- But statistical registers have:
  - Inputs “collected” from different sources
  - A sequence of processing and analysis
  - Outputs – statistics and sampling frames
- Therefore, register maintenance is similar to other statistical production


4
Collect
5
Process
6
Analyse
7
Disseminate
4.1
Select sample
4.2
Set up collection
- Prepare for incoming survey and admin data files
4.3
Run collection
- Receive data files
4.4
Finalize collection
- Prepare files for loading to the register
5.1
Integrate data
- Load data
- Perform automatic / clerical matching
5.2
Classify & code
- Perform automatic / manual coding of variables
5.3
Review, Validate & edit
- Perform validation checks
- Edit "unusual" data according to pre-defined edit rules
5.4
Impute
- Impute for missing units or variables
5.5
Derive new variables & statistical units
- Profiling
5.6
Calculate weights
5.7
Calculate aggregates
- Population totals
5.8
Finalize data files
- Maximize register quality prior to creating outputs
6.1
Prepare draft outputs
- Create "dummy" sampling frames or data outputs
6.2
Validate outputs
- Check quality
- Compare to expected values and other data sources or time periods
6.3
Scrutinize & explain
- Explain any differences found in 6.2
6.4
Apply disclosure control
- To any outputs made available to those not authorised to see micro-data
6.5
Finalize outputs
- Prepare and authorize for release
7.1
Update output systems
- Load data into output databases
7.2
Produce dissemination products
- Tables, publications, datasets etc.
7.3
Manage release of dissemination products
7.4
Promote dissemination products
7.5
Manage user support
- Answer queries from survey statisticians
The dotted lines show the different routes depending on whether the output is a sampling frame or a statistical product

Why do this?

There are practical benefits:

- Standardisation of terminology
- Standard framework for benchmarking
- Facilitates use of common tools / methods
- Efficiency savings
- Tool for managing process quality

Documentation

Example: Armenia - 2011 Population Census

Phase 5 – Process

5.1. Integrate data.

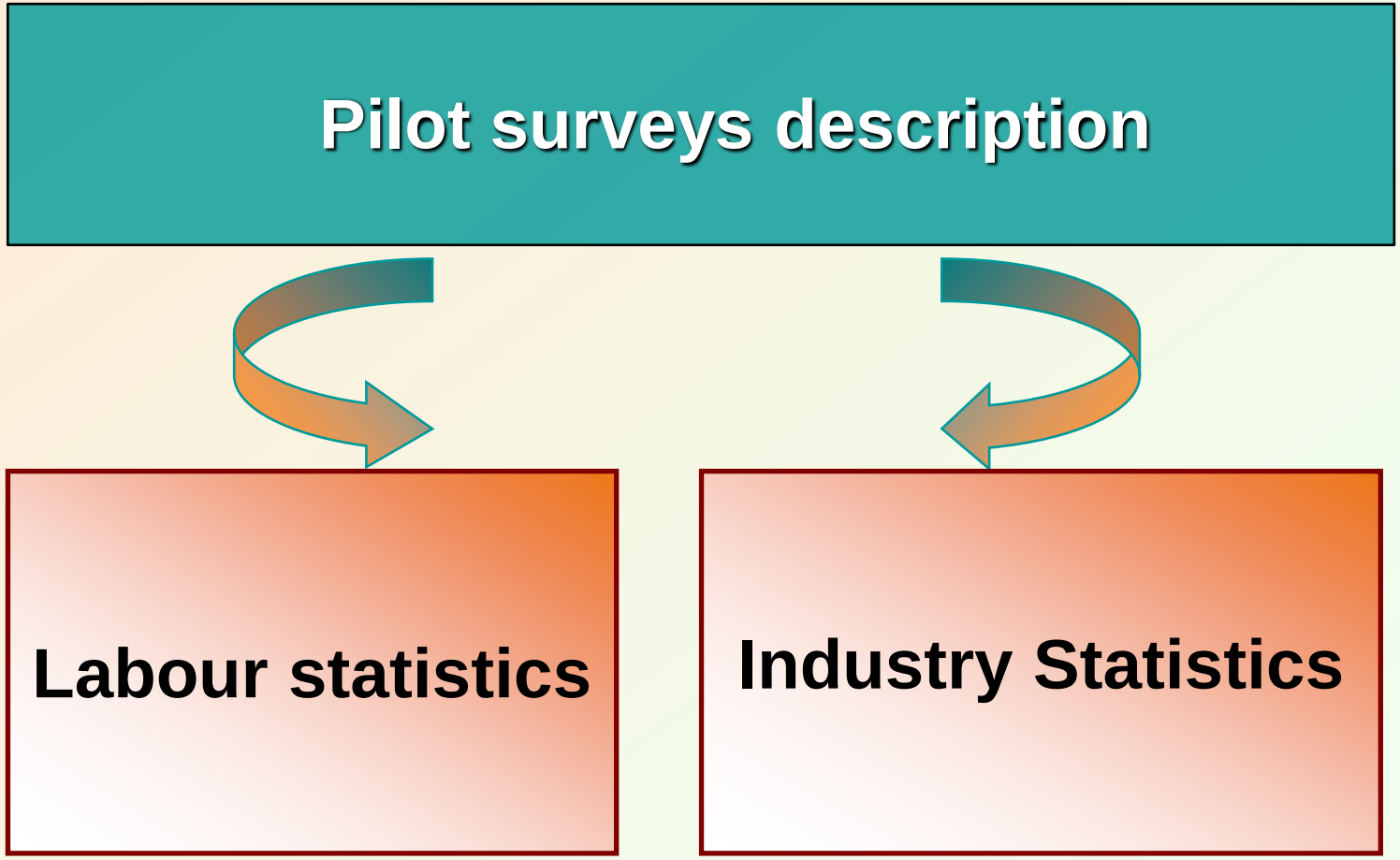
On the reason that in Armenia the census data are collected from one source, there is no necessity to integrate data collected from different sources. The data integration is performed after the data coding and data entry. As the result of data entry, the text files are created, each of them corresponds to one enumeration area. After the data entry completion, all the files are combined, and the database is created for further cleaning and analysis. The data integration can be performed also before the data entry completion for the processing software testing. For the data entry as well as integration the software CSPro is used. Duration of the data integration is very few seconds. Responsible for this sub-process is the Division of information technology and programming of the Population census department. The sub-process of data integration is also used in other processes of data production.

5.2. Classify and code

The works on classification and coding were launched in parallel with the information acceptance at the Department of census. Each portfolio corresponds to one

Similar approaches are now widely used in many other countries

Belarus: Using GSBPM 5.0 to describe the existing statistical production processes



Results:

Identification of **gaps** in the existing processes

Lack of necessary documentation

Existence of **unsettled** processes

Purpose of documentation

- Needs to be agreed before work starts!
- Examples:
  - Knowledge management
  - Succession planning
  - Standardisation – understanding the starting point
  - Metadata / quality management
  - International reporting

What to document?

- For the whole process:
  - Summary of purpose
  - Sources, outputs and users
  - Links to other processes
  - Costs?

What to document?

- For GSBPM sub-processes:
  - Purpose
  - Inputs and outputs (use GSIM?)
  - Tools and methods
  - Quality criteria
    - When to move on to next sub-process
  - Costs?

Documentation

Comparison

Efficiency!

How to begin?

- Read the GSBPM sub process descriptions:

2.3. Design collection

This sub-process determines the most appropriate collection method(s) and instrument(s). The actual activities in this sub-process will vary according to the type of collection instruments required, which can include computer assisted interviewing, paper questionnaires, administrative data interfaces and data integration techniques. This sub-process includes the design of collection instruments, questions and response templates (in conjunction with the variables and statistical classifications designed in sub-process 2.2 (Design variable descriptions)). It also includes the design of any formal agreements relating to data supply, such as memoranda of understanding, and confirmation of the legal basis for the data collection. This sub-process is enabled by tools such as question libraries (to facilitate the reuse of questions and related attributes), questionnaire tools (to enable the quick and easy compilation of questions into formats suitable for cognitive testing) and agreement templates (to help standardise terms and conditions). This sub-process also includes the design of process-specific provider management systems.

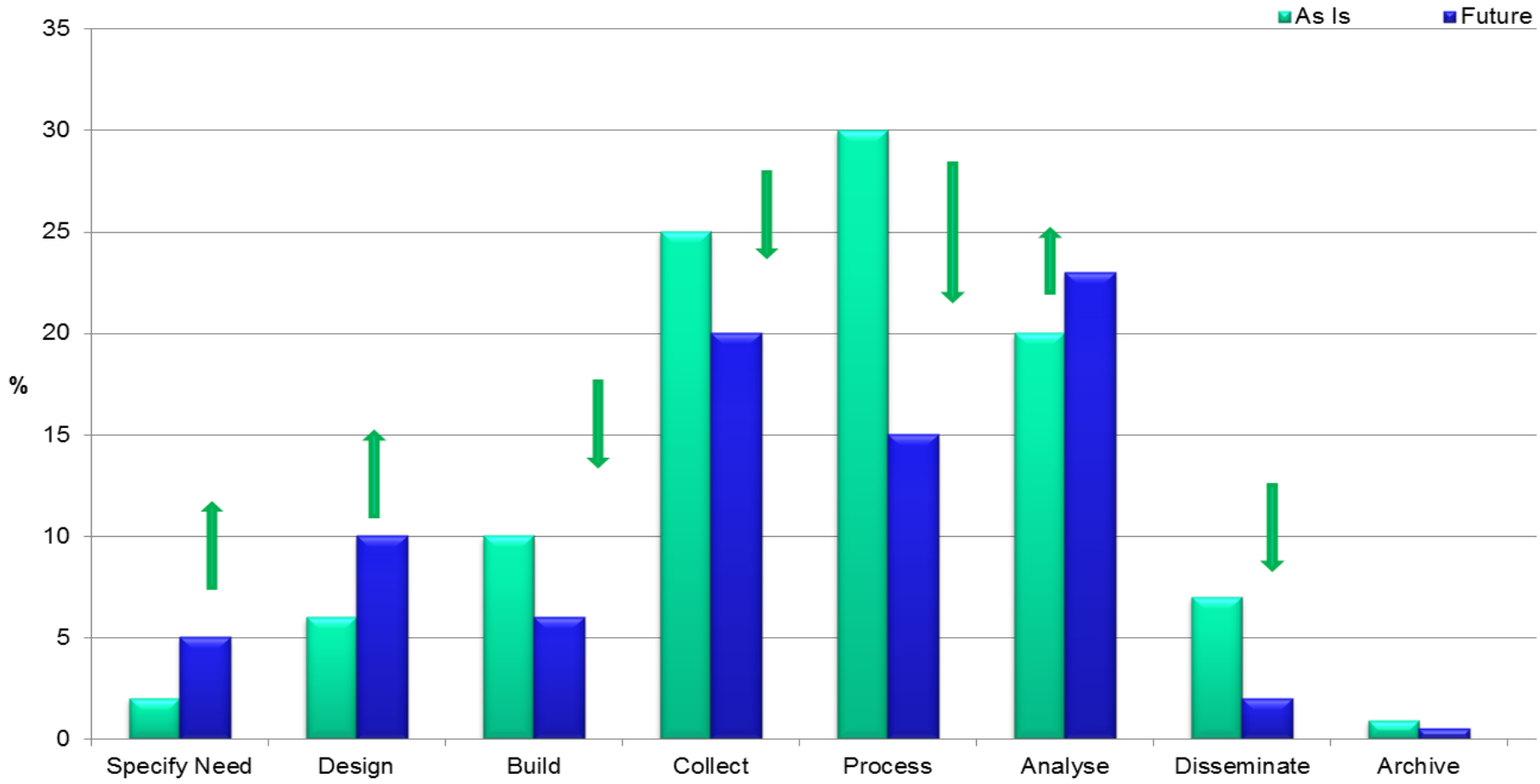
How to begin?

- Bring together all colleagues who are involved in the process
- Agree and describe the steps needed to complete the process
- Document them!
- Benefits include increased transparency and identifying areas for improvement
- Discussions often bring new ideas

Resource planning

UNECE

Indicative Current and Future Effort Distribution



GSBPM quality indicators

- Developed by task team
  - Canada, Hungary, Italy, Turkey, Eurostat, UNECE
- Generic quality indicators for each GSBPM sub-process
  - Processes based on surveys: 2016
  - Expanded to include administrative data: 2017
- Consistent with existing frameworks:
  - UN NQAF, ESS Code of Practice / QAF

UNECE

modernstats
by HLG - MOS

Quality Indicators for the Generic Statistical Business Process Model (GSBPM) - For Statistics derived from Surveys and Administrative Data Sources

(Version 2.0, October 2017)

https://statswiki.unece.org/display/GSBPM/Quality+Indicators

Uses of the Quality Indicators

- To provide a standard framework / common terminology to support a process-oriented approach to Quality Management
- To rationalise quality work within an NSO
- To define a mid-term quality policy
  - Set quality targets for a 3-5 year period

Example: 4.4 Finalise Collection

Quality Dimension	Indicator
Cost-effectiveness	Discrepancy between planned versus actual collection costs Percentage of collection activities that met requirements (assessed through analysis of paradata)
Accuracy and reliability	Outgoing error rates; estimate of non-sampling error
Accuracy and reliability	The rate of over-coverage: The proportion of units accessible via the frame that do not belong to the target population (are out-of-scope).

Quality and Metadata Management

- Needed at many different levels:
  - Process / sub-process level – GSBPM
  - Organisation level – GAMSO

GSBPM implementation information

GSBPM Wiki https://statswiki.unece.org/display/GSBPM

UNECE

Generic Statistical Business Process Model

PAGE TREE

- GSBPM v5.1
- Clickable GSBPM v5.1
- GSBPM v5.1 main changes
- GSBPM Resources Repository
- Uses of GSBPM
- GSBPM Training Materials
- Old Versions of the GSBPM
- Quality Indicators
- GeoGSBPM
- GSBPM Discussion Forum

Space tools

Pages 61,217 views Edit Save for later Share ...

Generic Statistical Business Process Model

Created by Steven Vale, last modified by InKyung Choi on 08 Sep, 2021

modernstats by HLG - MOS

GSBPM
Current version of GSBPM v5.1 available here

Geospatial view of GSBPM
New Geospatial-related activities and considerations needed for the production process

Clickable GSBPM
Clickable GSBPM v5.1

Learning GSBPM
Are you new to GSBPM and want to learn about it? Check out here for introductory presentations from past training and workshops

Uses of GSBPM
Use cases of GSBPM (mapping specific processed to GSBPM, using GSBPM for managing statistical programmes and etc.)

Quality Indicators for GSBPM
Quality indicators for each GSBPM subprocess

Previous versions of GSBPM

Implementations and case studies

GSBPM Resources Repository

Created by InKyung Choi, last modified on 08 Sep, 2021

Title	Resource type	Date	Relevant GSBPM version	Language	Author name	Author organisation	Category	Note
Mapping data production processes to GSBPM	Presentation	2014-09	5.0	English	Steven Vale	UNECE	USE CASE INTRODUCTION	GSBPM Workshop, NSO Malta, September 2014
Communicating the GSBPM – How GSBPM's other uses can play a role	Paper			English	Joe Treacy	Central Statistics Office, Ireland	USE CASE COMMUNICATION	
Example using Statistics Norway's Business Process Model v1.1	Excel spread sheet			English		Statistics Norway	USE CASE IMPLEMENTATION	
A five-stage data quality compliance framework	Presentation	2019-06	5.0	English	Eduardo Jallath	INEGI, Mexico	USE CASE	Presented at ModernStats World Workshop 2019, 2019-06

Discussion forum

UNECE

GSBPM Discussion Forum

Created by Steven Vale, last modified by Tetyana Kolomiyets on 06 Apr, 2021

Comments or questions about the GSBPM? Ideas for revision? Experiences to share?

Use this discussion forum to share your thoughts.

You need to be logged in to participate in the forum. To request a username and password contact support.stat@un.org

Search for Topics... + New Topic 84 topics found

Show Advanced Filters

Sticky topics on top

Status	Topic	Author	# of Replies	Last Activity Date
	Issue #5: Mapping GSBPM Overarching processes to GAMSO (from Franck Cotton) Draft mapping produced during the last GAMSO revision	Chris Jones	0	Sep 08, 2021

Other papers on the web