GSBPM

Giorgia Simeoni

Directorate for Methodology and Statistical Process Design
Istat

ESTP Training Course "Enterprise Architecture (EA) for Official Statistics"

Rome, 18 – 21 October 2022

GSBPM

"The Generic Statistical Business Process Model (GSBPM) describes and defines the set of business processes needed to produce official statistics.

It provides a standard framework and harmonised terminology to help statistical organisations to modernise their statistical production processes, as well as to share methods and components. The GSBPM can also be used for integrating data and metadata standards, as a template for process documentation, for harmonising statistical computing infrastructures, and to provide a framework for process quality assessment and improvement." (GSBPM 5.1)

GSBPM Background

- Developed by the Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS)
- First version released on April 2009 (4.0), then revised twice:
  - in 2013 (version 5.0 released in December 2013): to incorporate feedback based on practical implementation and to improve consistency with Generic Statistical Information Model (GSIM)
  - In 2018-9 (version 5.1 released in January 2019): to incorporate feedback based on further practical implementation and to improve consistency with other HLG-MOS models, the GSIM and the Generic Activity Model for Statistical Oganisations (GAMSO).
- Current version GSBPM 5.1

https://statswiki.unece.org/display/GSBPM/Generic+Statistical+Business+Process+Model

GSBPM Background

- Widely adopted by the global official statistics community
- A cornerstone of the UNECE High-Level Group for the Modernisation of Official Statistics (HLG-MOS) vision and strategy for standards-based modernisation
- Now Managed by the "Supporting Standards Group", under the HLG
- **GSBPM has been approved as ESS Standard by the ESSC in February 2017**

UNECE High-Level Group for the Modernisation of Official Statistics

- UNECE is the United Nations Economic Commission for Europe. It is one of five regional commissions of the United Nations.
- High-Level Group for the Modernisation of Official Statistics (HLG-MOS) setup in 2010 under the Conference of European Statisticians (CES).
- HLG-MOS is composed by the Chief statistician (Presidents or General Directors) of National and International Statistical Organisations.
- Since 2016 the community is branded as:

modernistats BY HLG - MOS

- Australia
- Canada
- Ireland
- Italy
- Mexico
- Netherlands
- New Zealand
- Poland
- Republic of Korea
- Slovenia
- United Kingdom
- Eurostat
- OECD
- UNECE

UNECE HLG-MOS Mission

Mission: The HLG-MOS is a group of committed Chief Statisticians actively steering the modernisation of statistical organisations.
Their mission is to work collaboratively to identify trends, threats, and opportunities in modernising statistical organisations.
It provides a common platform for experts to develop solutions in a flexible and agile way.
It is a voluntary collaboration of willing and able.

HLG-MOS oversees the activities of «stable» modernisation groups and «annual» modernisation projects

Supporting Standards

It is one of the HLG- MOS modernisation groups. The goal of the group is to find ways how to develop, enhance, integrate, promote, support and facilitate implementation of the range of standards needed for statistical modernisation

It has the operational responsibility for the maintenance and development of:

GAMSO - Generic Activity Model for Statistical Organizations

GSBPM - Generic Statistical Business Process Model

GSIM - Generic Statistical Information Model

And the documentation of:

CSPA – Common Statistical Production Archiecture

European Commission

GSBPM

What is GSBPM?

GSBPM is a generic process model for production of official statistics. It consists of 8 phases and 44 sub-processes.

More than 40 NSOs and international statistical organizations are using GSBPM

GSBPM is a living model that evolves to reflect changing business landscape. The latest version came out in 2019

How does GSBPM help statistical organizations?

Provides a structure for documentation of statistical processes

Facilitates sharing statistical methods and software

Provides a framework for quality assessment

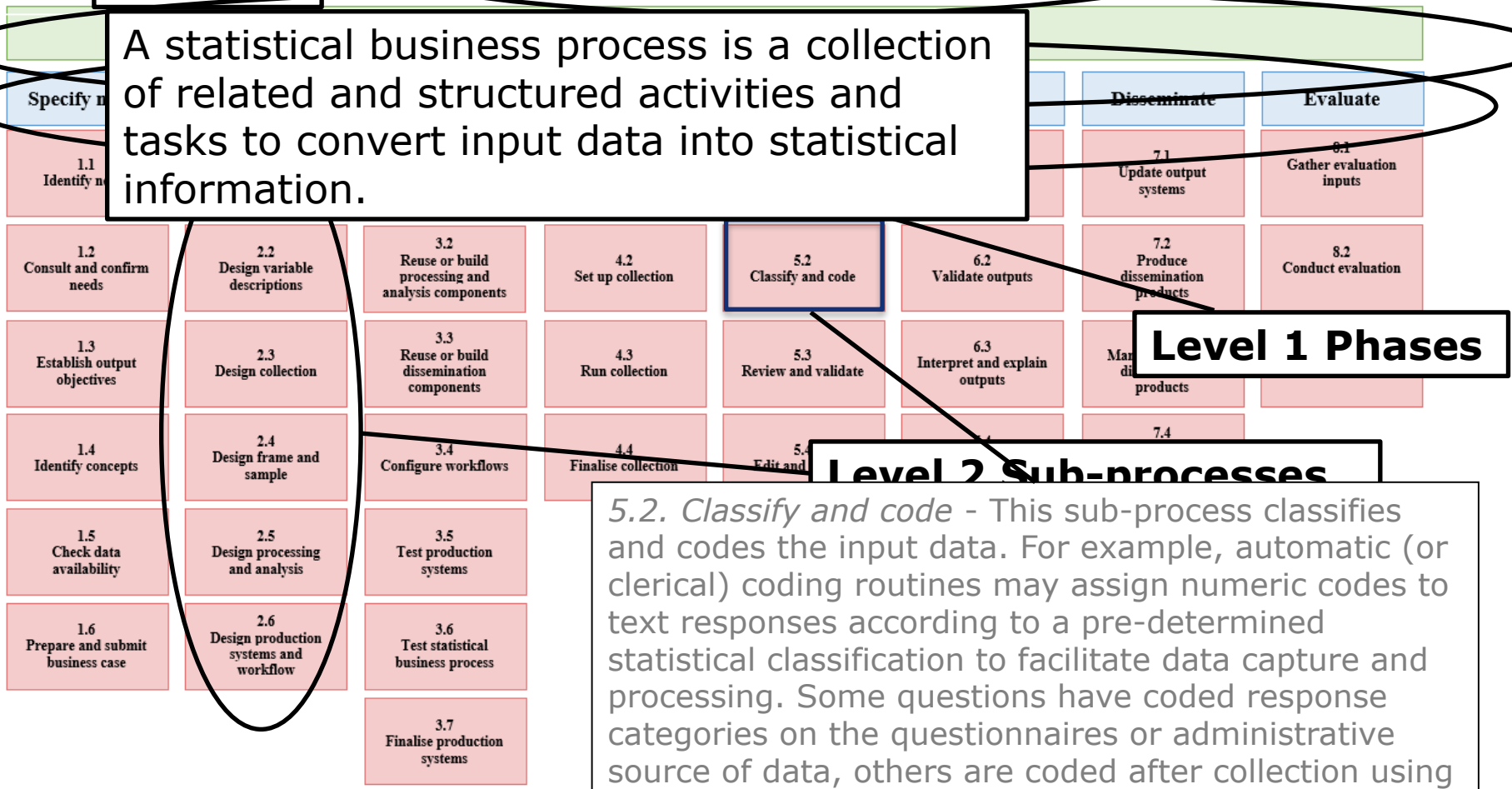
Helps to measure operational cost and system performance

For more details, visit UNECE GSBPM Wikis: https://statswiki.unece.org/display/GSBPM/

The Statistical Business Process Model

Level 0

A statistical business process is a collection of related and structured activities and tasks to convert input data into statistical information.



Level 1 Phases

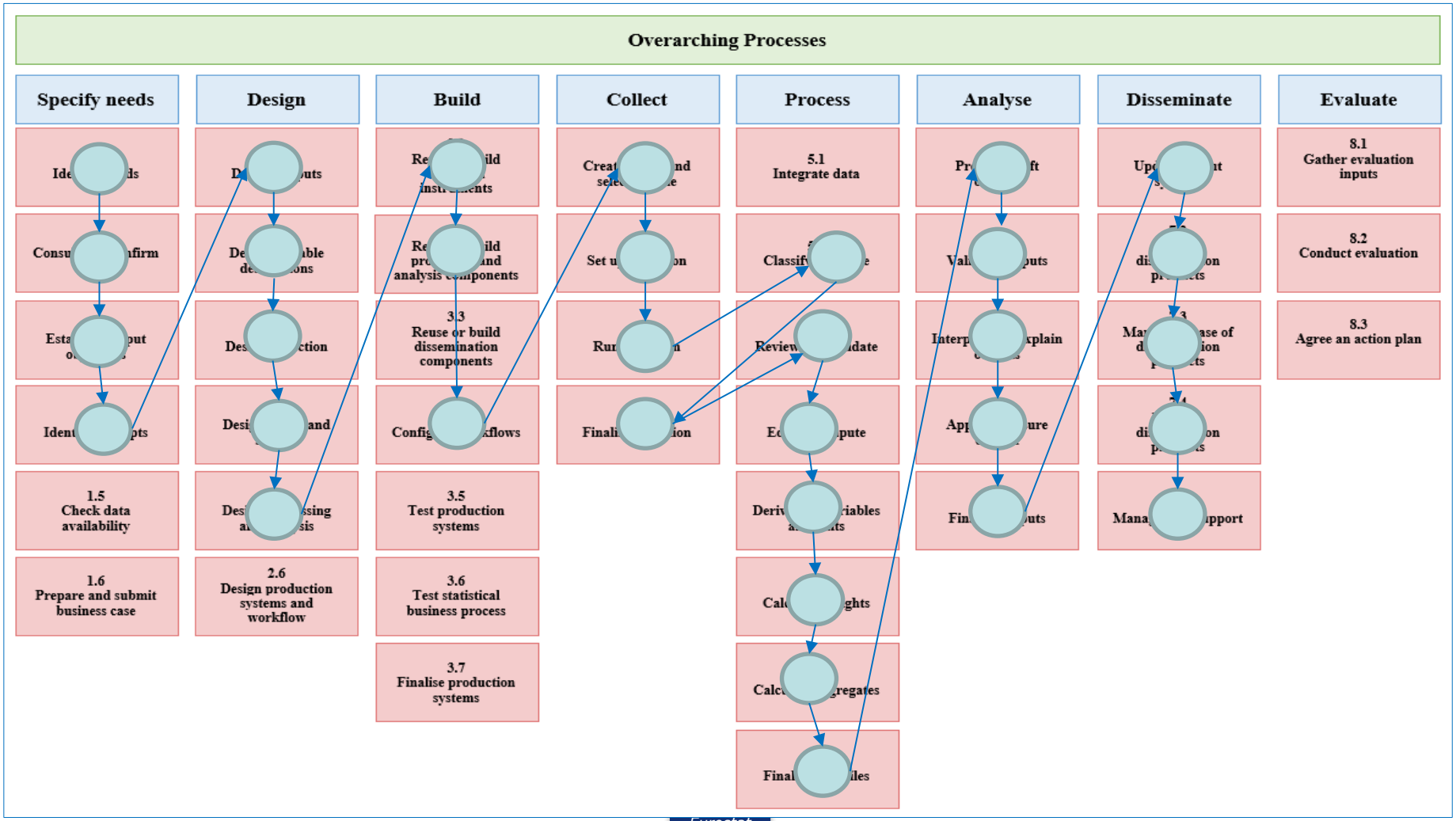
Level 2 Sub-processes

5.2. Classify and code - This sub-process classifies and codes the input data. For example, automatic (or clerical) coding routines may assign numeric codes to text responses according to a pre-determined statistical classification to facilitate data capture and processing. Some questions have coded response categories on the questionnaires or administrative source of data, others are coded after collection using an automated process (which may apply machine learning techniques) or an interactive, manual process.

Main features

- "The GSBPM is intended to apply to all activities undertaken by producers of official statistics, at both the national and international levels, which result in data outputs. "
- Organised in 3 levels with increasing level of detail
- Detailed description of what is included in each sub-process
- Extremely flexible:
  - Not all the subprocesses should be performed
  - The order of subprocesses can be different from the one presented
  - It is possible to repeat some steps more than once
  - It can be considered a matrix

Example of workflow in GSBPM



Overarching processes

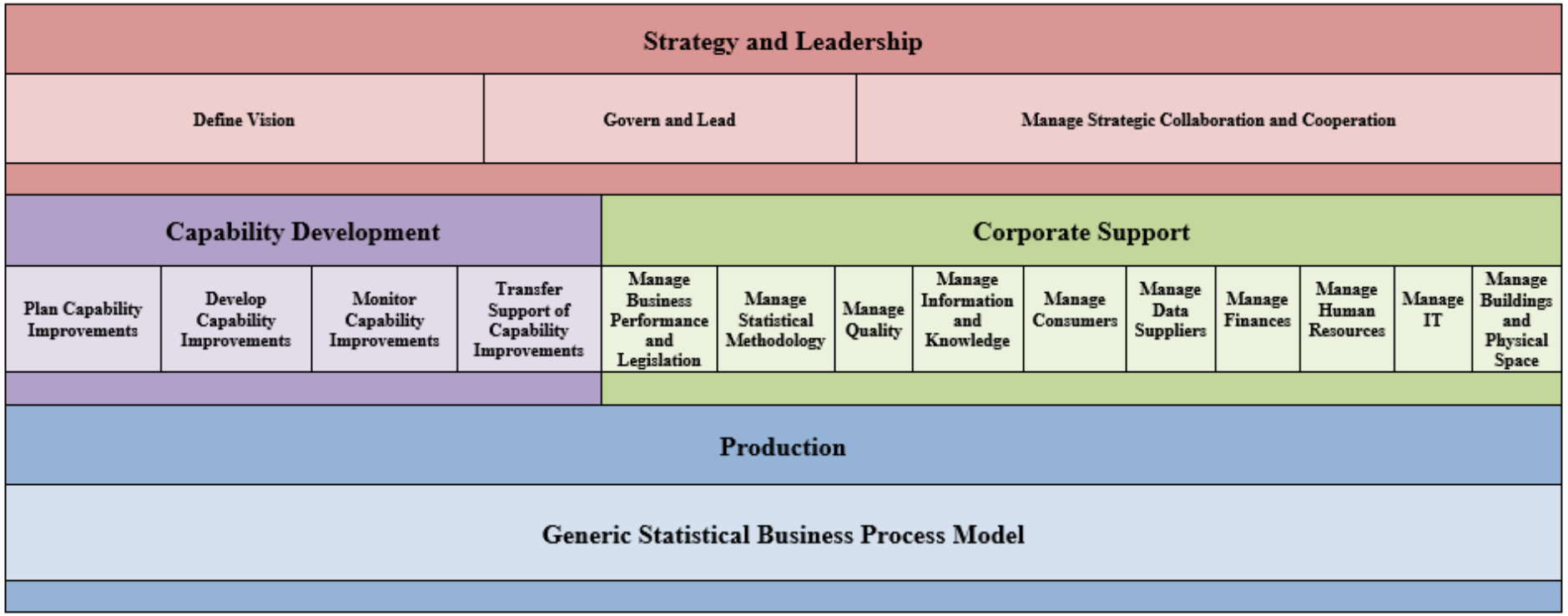
Processes with a strong statistical component that apply throughout the eight phases are included in GSBPM as Overarching processes. They are:

- Quality management
- Metadata management
- Data management
- Process data management
- Knowledge management
- Provider management

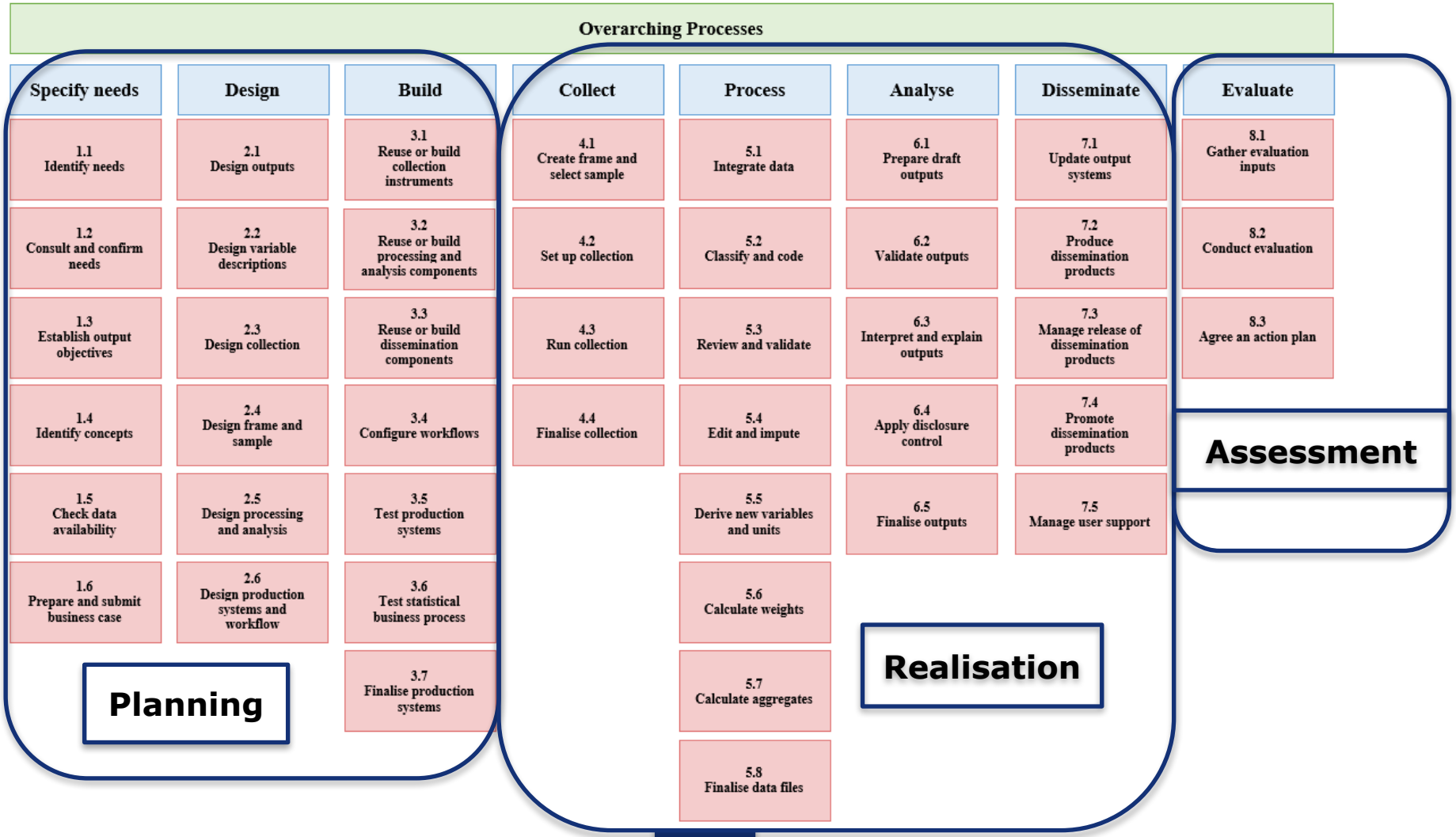
Activities that are carried out at the level of the organisation to support the statistical production are included in the GAMSO

GAMSO v.1.2

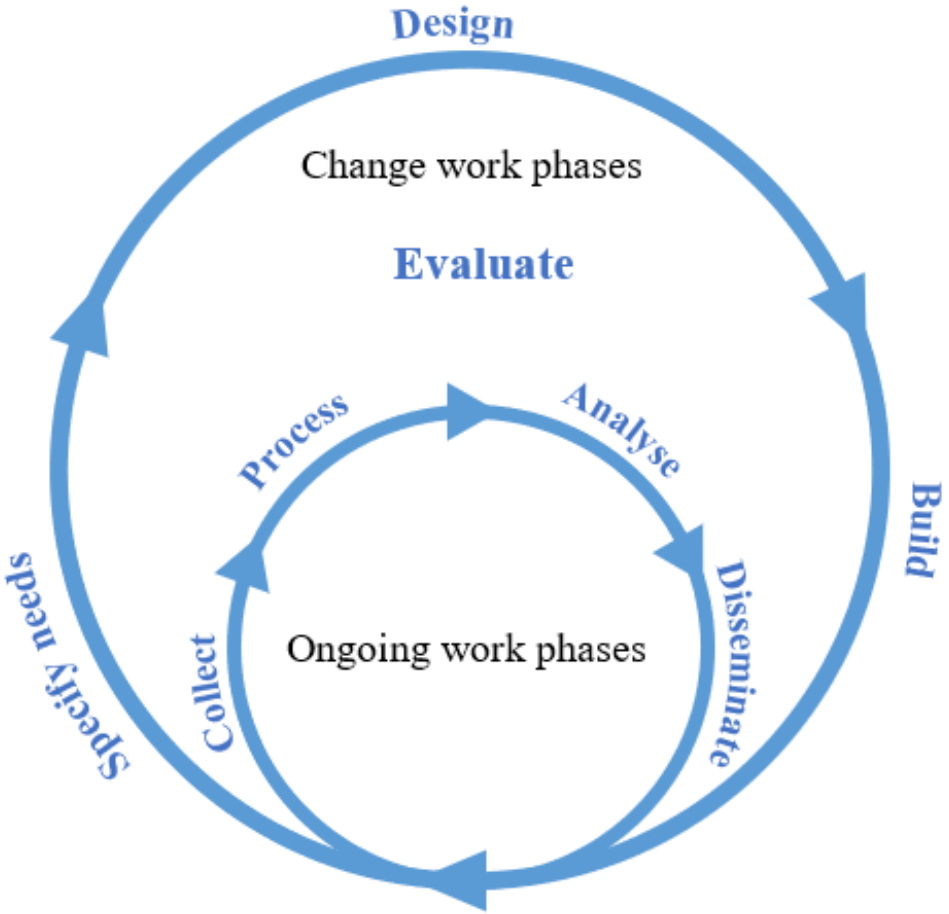
The Generic Activity Model for Statistical Organisations (GAMSO) describes and defines the activities that take place within a typical organisation that produces official statistics. It extends and complements GSBPM by adding additional activities needed to support statistical production.



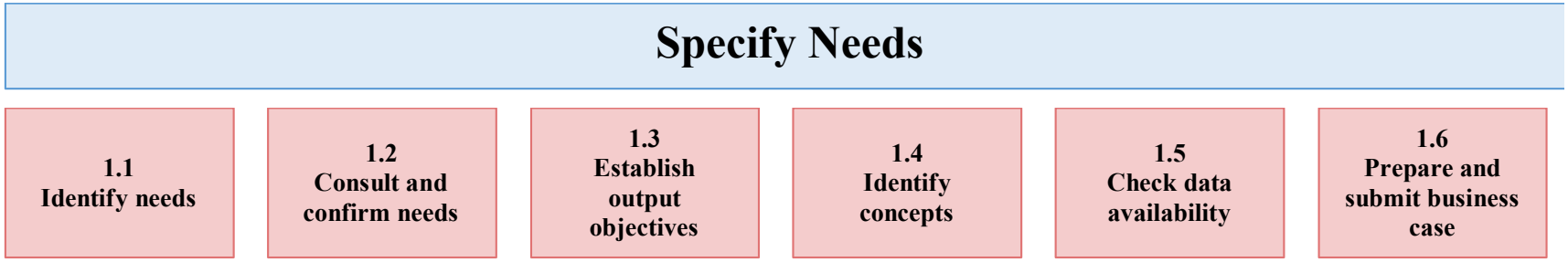
The Generic Statistical Business Process Model



GSBPM 5.1


Design
Change work phases
Evaluate
Process
Analyse
Build
Disseminate
Collect
Specify needs
Ongoing work phases

The planning sub-processes



- Dialog with users,
- Identification of needs (new or additional),
- Definition of high level solution,
- Get approval from senior management

The planning sub-processes

Design					
2.1 Design outputs	2.2 Design variable descriptions	2.3 Design collection	2.4 Design frame and sample	2.5 Design processing and analysis	2.6 Design production systems and workflow

- **Definition** of all methods and tools that will be used in the realisation of the statistical process

Build						
3.1 Reuse or build collection instruments	3.2 Reuse or build processing and analysis components	3.3 Reuse or build dissemination components	3.4 Configure workflows	3.5 Test production systems	3.6 Test statistical business process	3.7 Finalise production systems

- **Set up and test** of all methods and tools defined in the Design phase

The realisation sub-processes

Collect

4.1 Create frame and select sample	4.2 Set up collection	4.3 Run collection	4.4 Finalise collection

- The actual data acquisition, whatever the source or the method used. Data entry is included in «Finalise collection»

Process

5.1 Integrate data	5.2 Classify and code	5.3 Review and validate	5.4 Edit and impute	5.5 Derive new variables and units	5.6 Calculate weights	5.7 Calculate aggregates	5.8 Finalise data files

- The traditional phases of data treatment till the macrodata estimates are produced

The realisation sub-processes

Analyse				
6.1 Prepare draft outputs	6.2 Validate outputs	6.3 Interpret and explain outputs	6.4 Apply disclosure control	6.5 Finalise outputs

- It includes the production of complex statistics (e.g. indices), macrodata validation, confidentiality treatment

Disseminate				
7.1 Update output systems	7.2 Produce dissemination products	7.3 Manage release of dissemination products	7.4 Promote dissemination products	7.5 Manage user support

- The release of statistical outputs to users

The Assessment sub-process

Evaluate		
8.1 **Gather evaluation inputs**	**8.2** **Conduct evaluation**	**8.3** **Agree an action plan**

- The quality evaluation done at the end of a specific edition of a statistical business process

Main overarching processes

Quality Management

- The overarching process on Quality represents quality assurance system implemented across the business process
- Quality at an institutional level is considered in the GAMSO

Metadata Management

- Metadata are present in every phase, either created, updated or carried forward from a previous phase or reused from another business process.
- When metadata management is considered at corporate or strategic level it should be considered in the framework of the GAMSO

Data Management

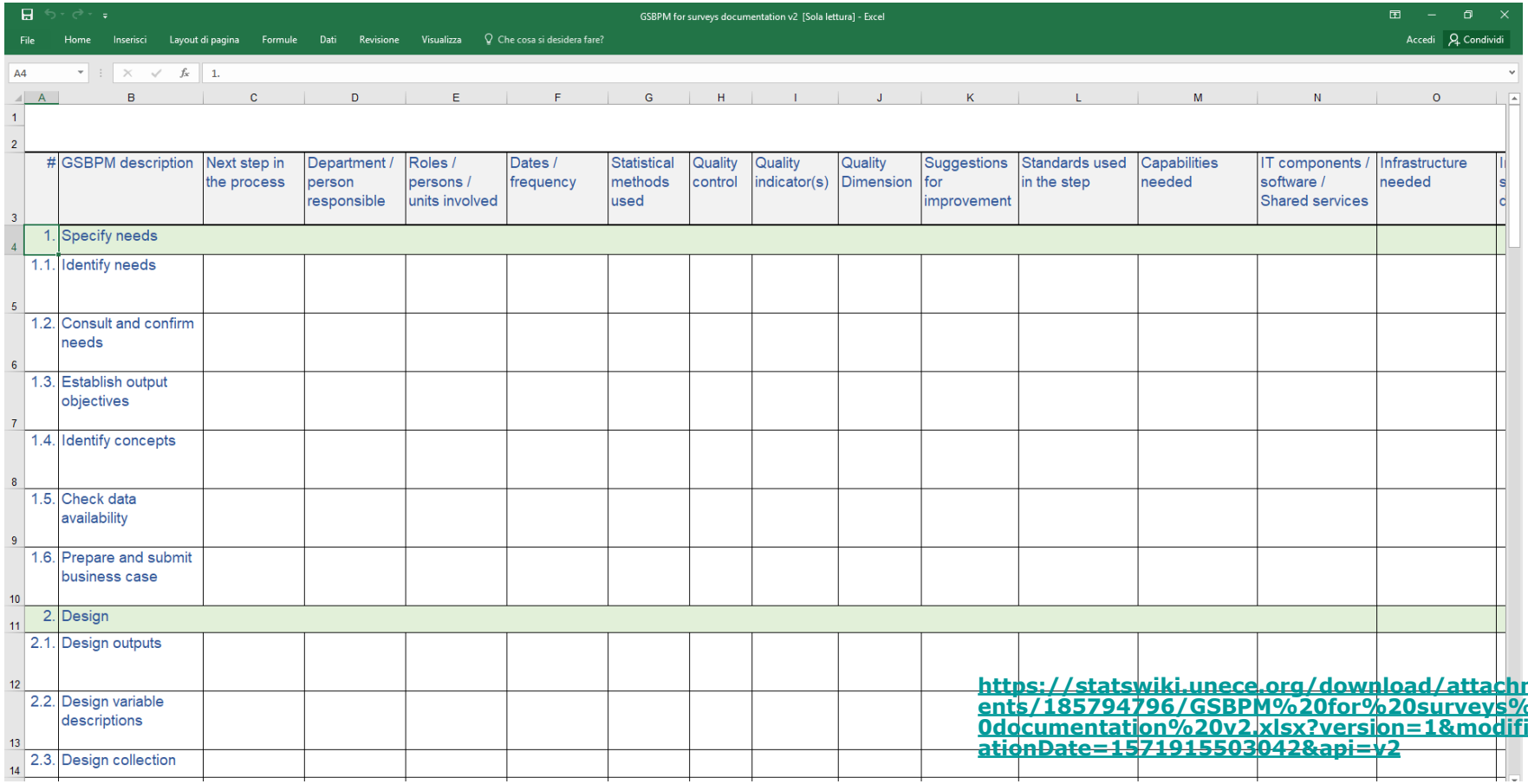
Uses of GSBPM

- Standardisation of terminology in international context
- Support to statistical process documentation
- Analyse processes in order to identify common subprocesses
- Make inventory of available IT tools and application to rationalise and identify gaps
- Make inventory of available methodological tools to rationalise and identify gaps
- Reference model to support audit and self assessment procedures
- …

Example 1. Business Process Model at Statistics Sweden

http://www1.unece.org/stat/platform/display/GSBPM/Implementing+the+GSBPM

Example 2. Excel sheet for process/survey documentation



GSBPM for surveys documentation v2 [Sola lettura] - Excel

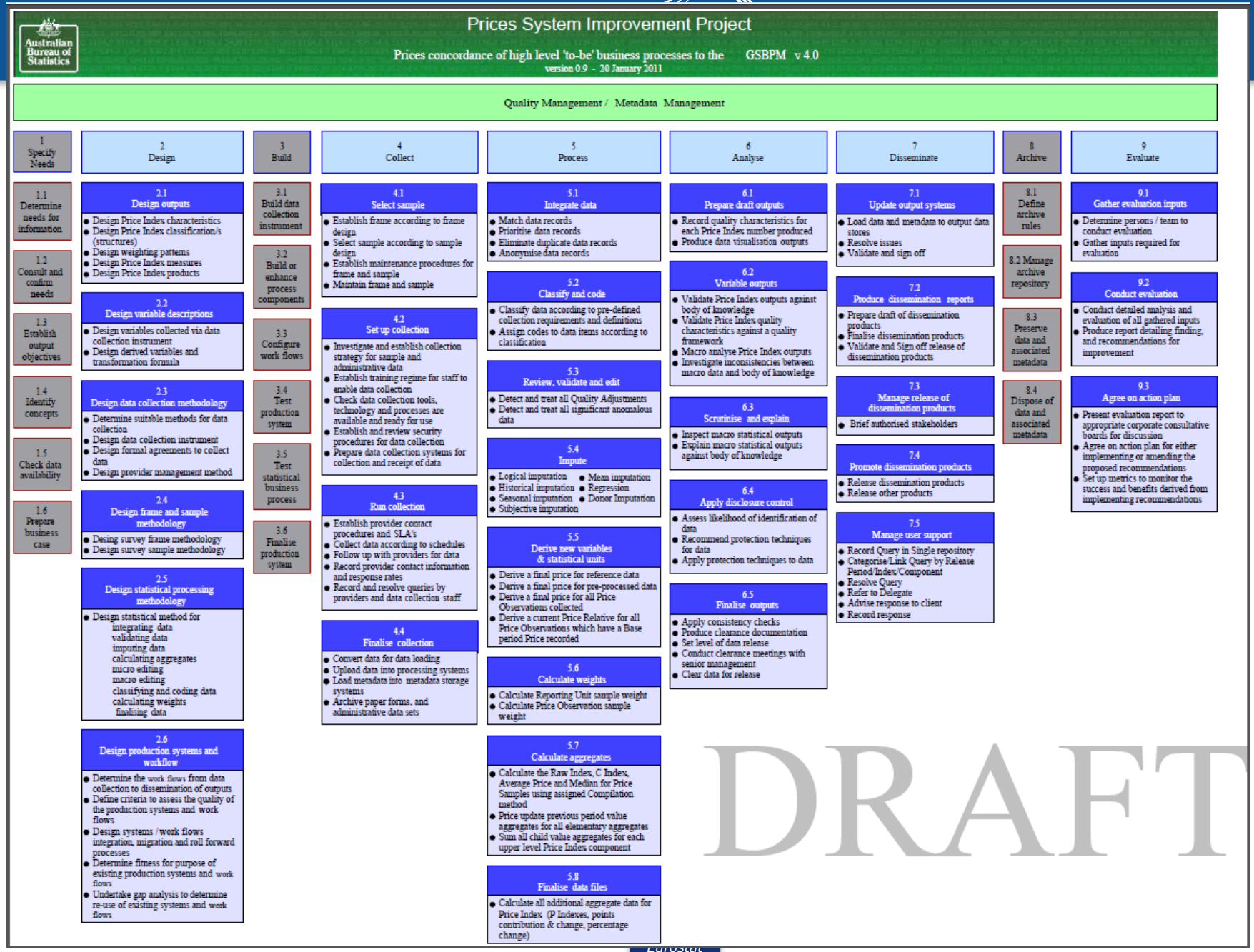
File Home Inserisci Layout di pagina Formule Dati Revisione Visualizza Che cosa si desidera fare? Accedi Condividi

A4 1.

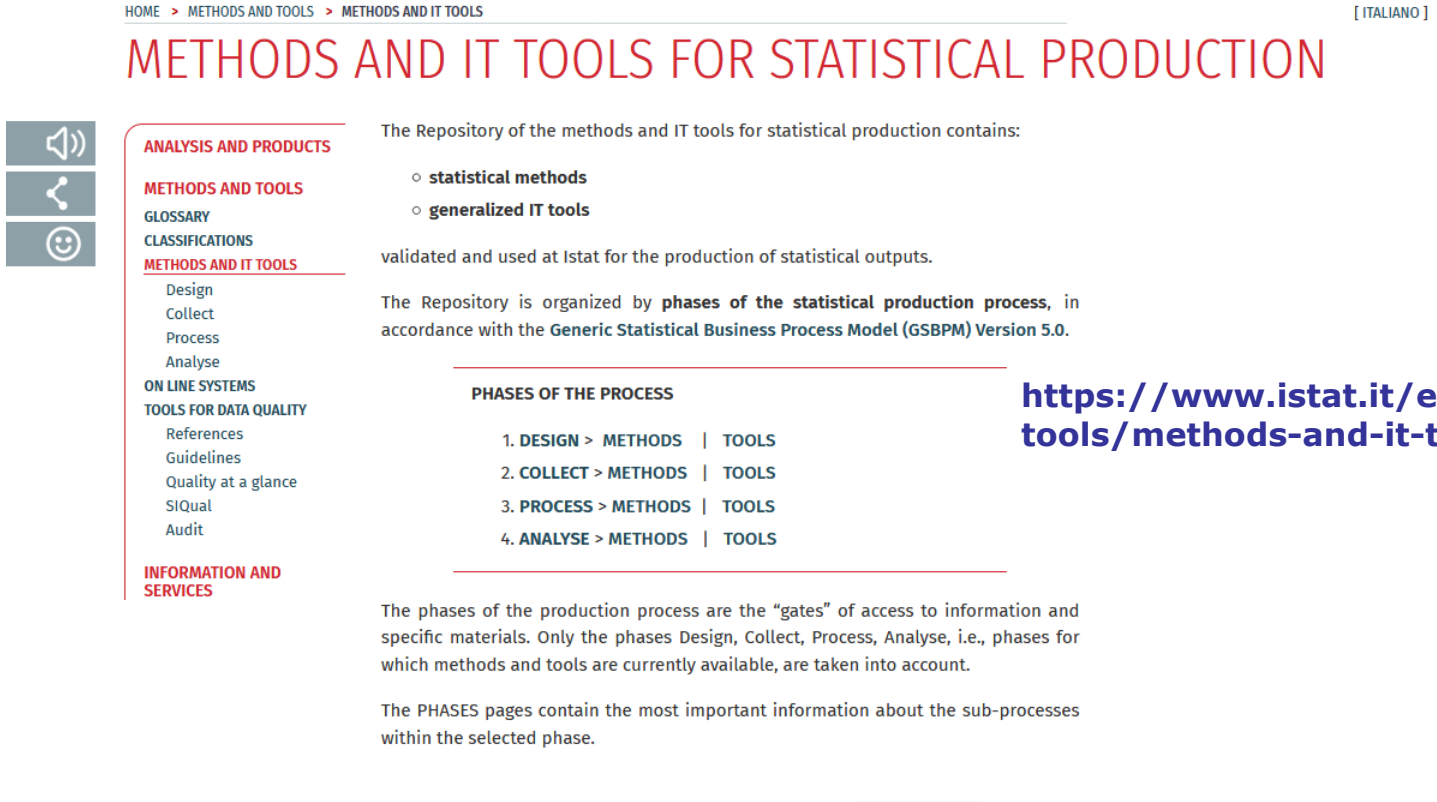
#	GSBPM description	Next step in the process	Department / person responsible	Roles / persons / units involved	Dates / frequency	Statistical methods used	Quality control	Quality indicator(s)	Quality Dimension	Suggestions for improvement	Standards used in the step	Capabilities needed	IT components / software / Shared services	Infrastructure needed
1.	Specify needs													
1.1.	Identify needs													
1.2.	Consult and confirm needs													
1.3.	Establish output objectives													
1.4.	Identify concepts													
1.5.	Check data availability													
1.6.	Prepare and submit business case													
2.	Design													
2.1.	Design outputs													
2.2.	Design variable descriptions													
2.3.	Design collection													

https://statswiki.unece.org/download/attachments/185794796/GSBPM%20for%20surveys%20documentation%20v2.xlsx?version=1&modificationDate=1571915503042&api=v2

Example 3. ABS Prices processes mapped to GSBPM (V4.0)



Example 4. Istat repository of methods and tools organised by GSBPM phases



https://www.istat.it/en/methods-and-tools/methods-and-it-tools

Implementing GSBPM

GSBPM can be applied in several ways

- Adopting it as it is to document statistical processes
- Personalising it for the needs of the Institute
- As a reference for metadata-driven system that should process data
- To rationalize tools and methods
- To map activities and assign responsibilities
- …

Implementation steps

Some steps should be conducted sequentially other could be in parallel

- Clarify the objectives: Documentation? Traceability? Quality assessment? Reporting? Processing data?
- Define the contents: sub-process description, input, output, workflow, tool, need of subtasks… (see excel file). Test contents and refine.
- Decide the governance: who does what? First implementation/maintainance. Training. Need for coordination
- Develop the tool: complexity of the tool is connected with objectives
- Start implementation: start collecting information

Quality indicators for GSBPM

- A wide set of quality indicators (partly qualitative and partly quantitative) has been mapped to each sub-process of GSBPM version 5.0 with the aim of expanding the quality management layer for the GSBPM

- UNECE (2017) 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/download/attachments/185794796/Quality%20Indicators%20for%20the%20GSBPM%20-%20For%20Statistics%20derived%20from%20Surveys%20and%20Administrative%20Data%20Sources_Final.pdf?api=v2

Quality indicators for GSBPM

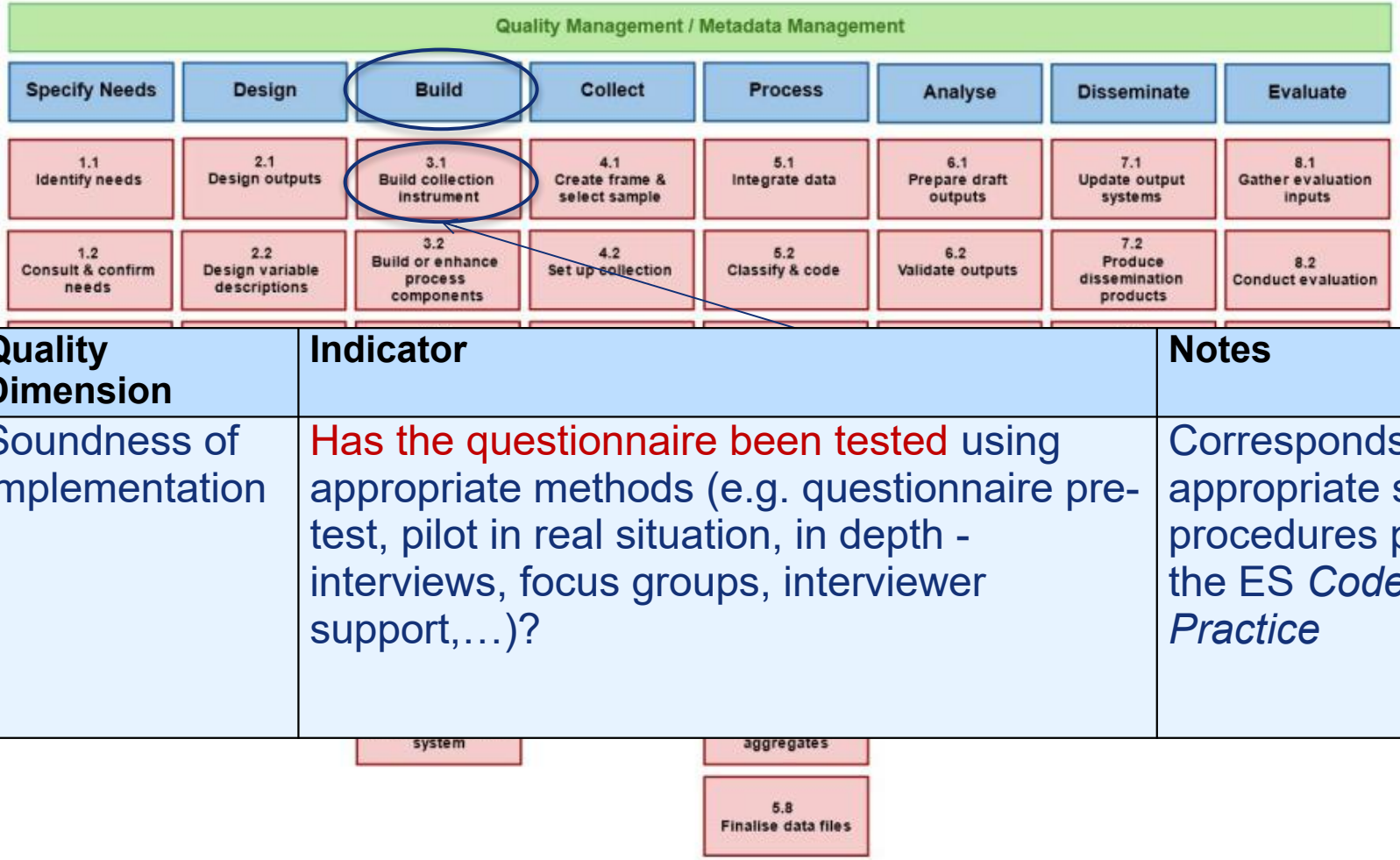
- Generic indicators as GSBPM is;
- Consistent with existing quality assurance frameworks;
- No formulas, only descriptions or explanations;
- Quantitative indicators whenever possible;
- Qualitative indicators in the form of yes/no or large/medium/low when appropriate;
- Map indicators to the phase they measure even if they might be calculated at a later stage;
- Allow for a certain degree of redundancy by mentioning the same indicators in different phases or sub-processes

→ Personalisation of the quality indicators left to NSIs

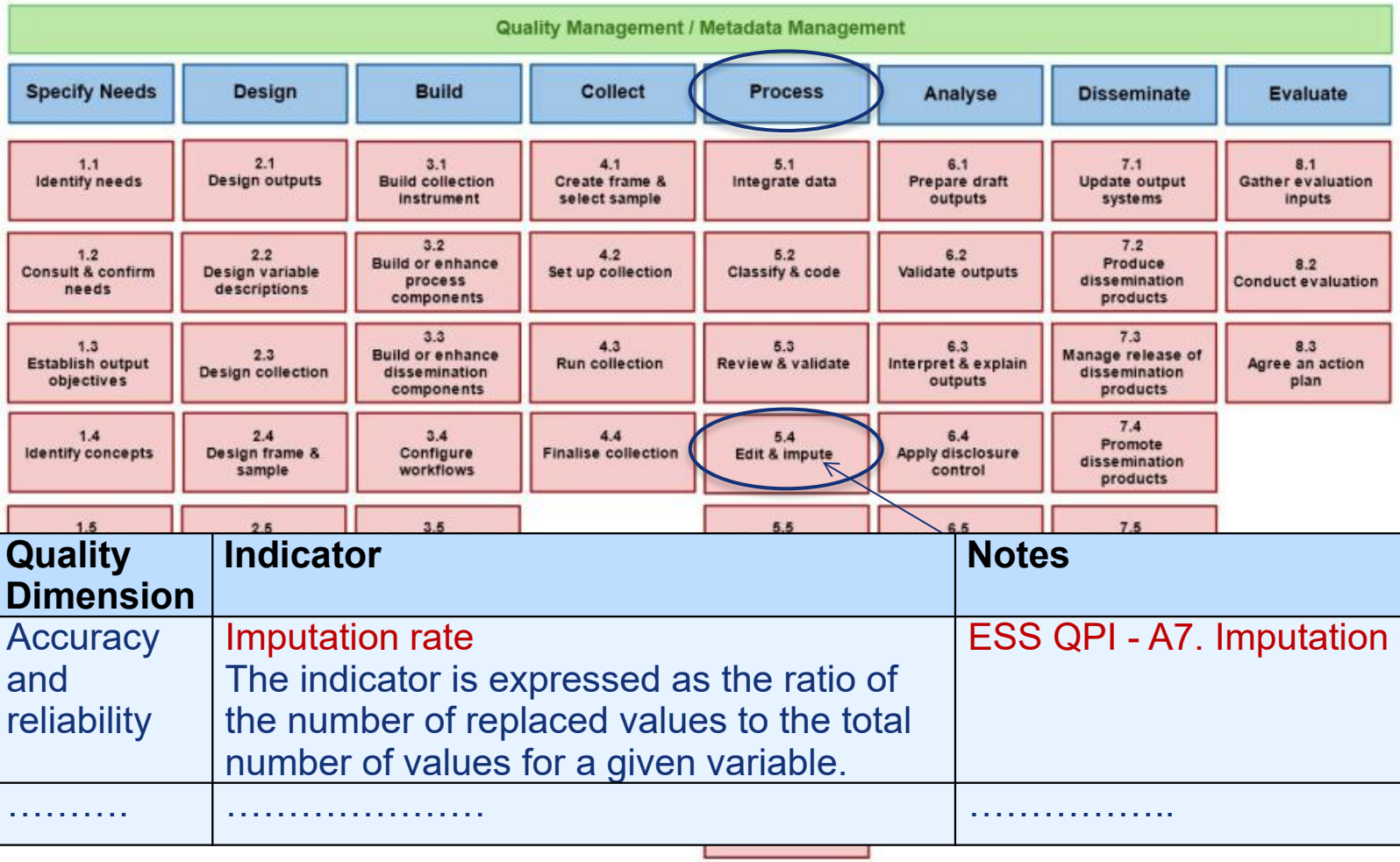
Structure of Quality Indicators

Quality Dimension	Indicator	Notes
.....	.....	.....

The (UN) National Quality Assurance Framework dimensions – NQAF (2012) were taken as reference for relating Quality Indicators to the corresponding quality dimension but mapping to the ES *Code of Practice* was indicated only in case of discrepancies.



Quality Dimension	Indicator	Notes
Soundness of implementation	Has the questionnaire been tested using appropriate methods (e.g. questionnaire pre-test, pilot in real situation, in depth - interviews, focus groups, interviewer support,…)?	Corresponds to the appropriate statistical procedures principle in the ES *Code of Practice*



Quality Dimension	Indicator	Notes
Accuracy and reliability	Imputation rate The indicator is expressed as the ratio of the number of replaced values to the total number of values for a given variable.	ESS QPI - A7. Imputation - rate
…….	……	……..

Recent developments on GSBPM

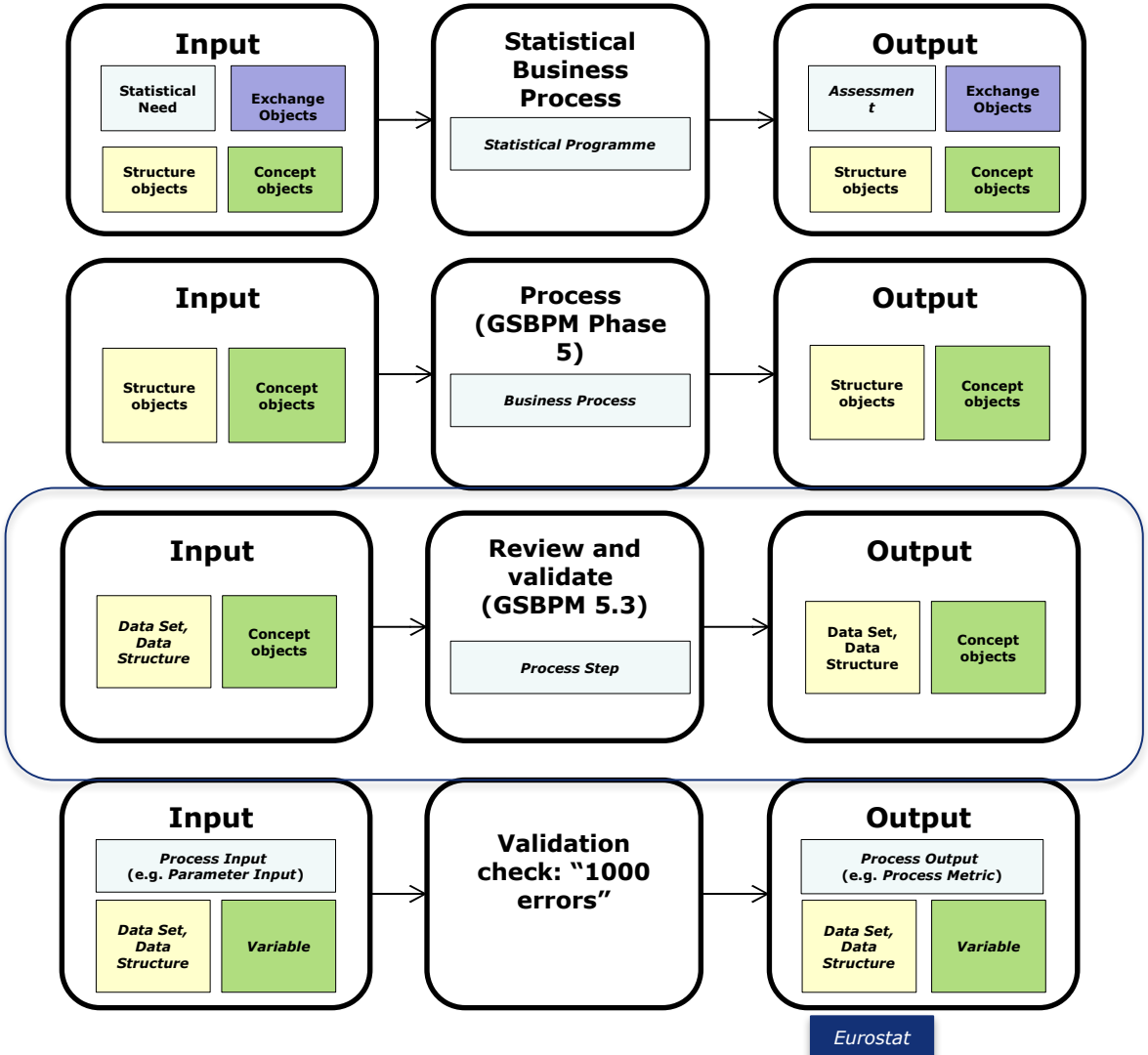
- "Linking GSBPM-GSIM"
- "Geospatial view of GSBPM"
- "GSBPM tasks"

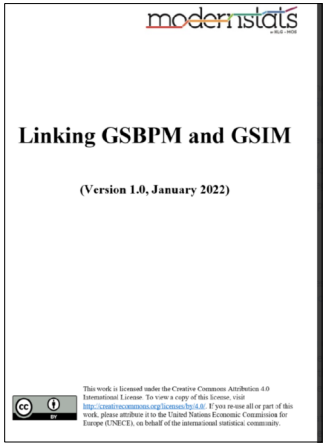
GSIM 1.2

- The Generic Statistical Information Model (GSIM) is a reference framework of internationally agreed definitions, attributes and relationships that describe the *pieces* of information used in the production of official statistics (information objects).
- This framework enables generic descriptions of the definition, management and use of data and metadata throughout the statistical production process, such as the ones described by GSBPM.
- GSIM is a conceptual model and does not prescribe how the information should be implemented.
- Current version of GSIM is 1.2

GSIM and GSBPM should be complementary

GSIM & GSBPM





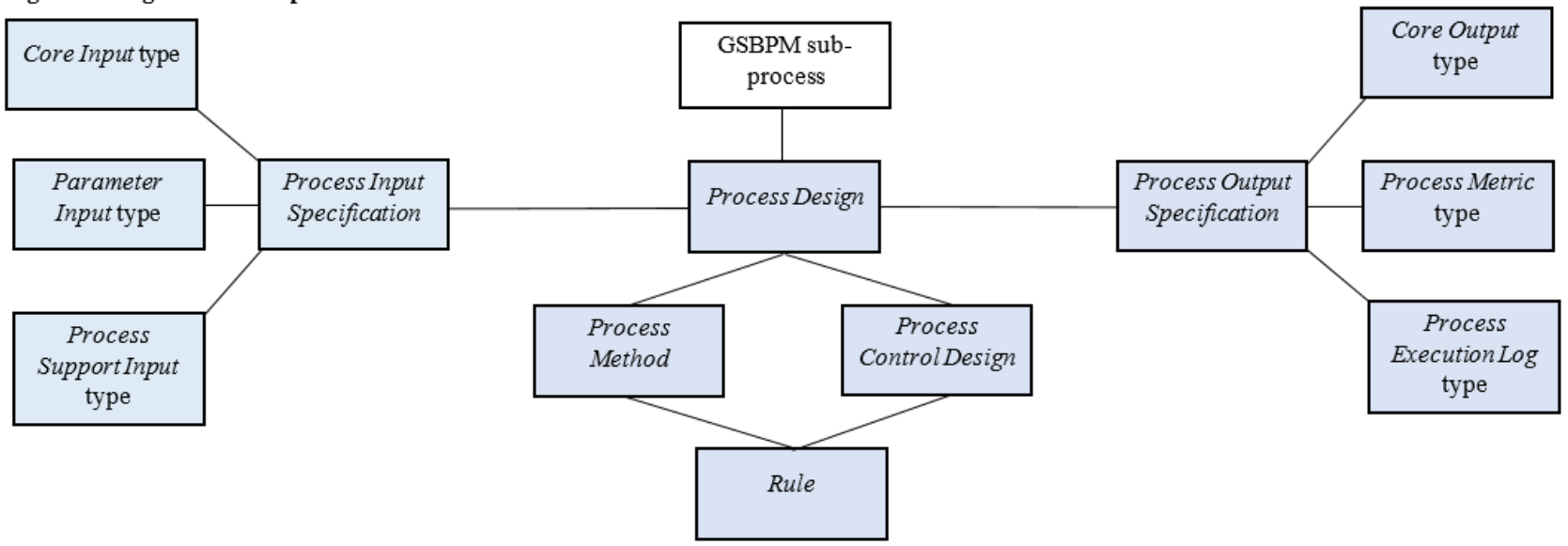
Linking GSBPM GSIM task team worked at this level https://statswiki.unece.org/display/GSBPM/Information+flow+within+GSBPM+using+GSIM

Linking GSIM & GSBPM

Table 1. Specification level (design time)

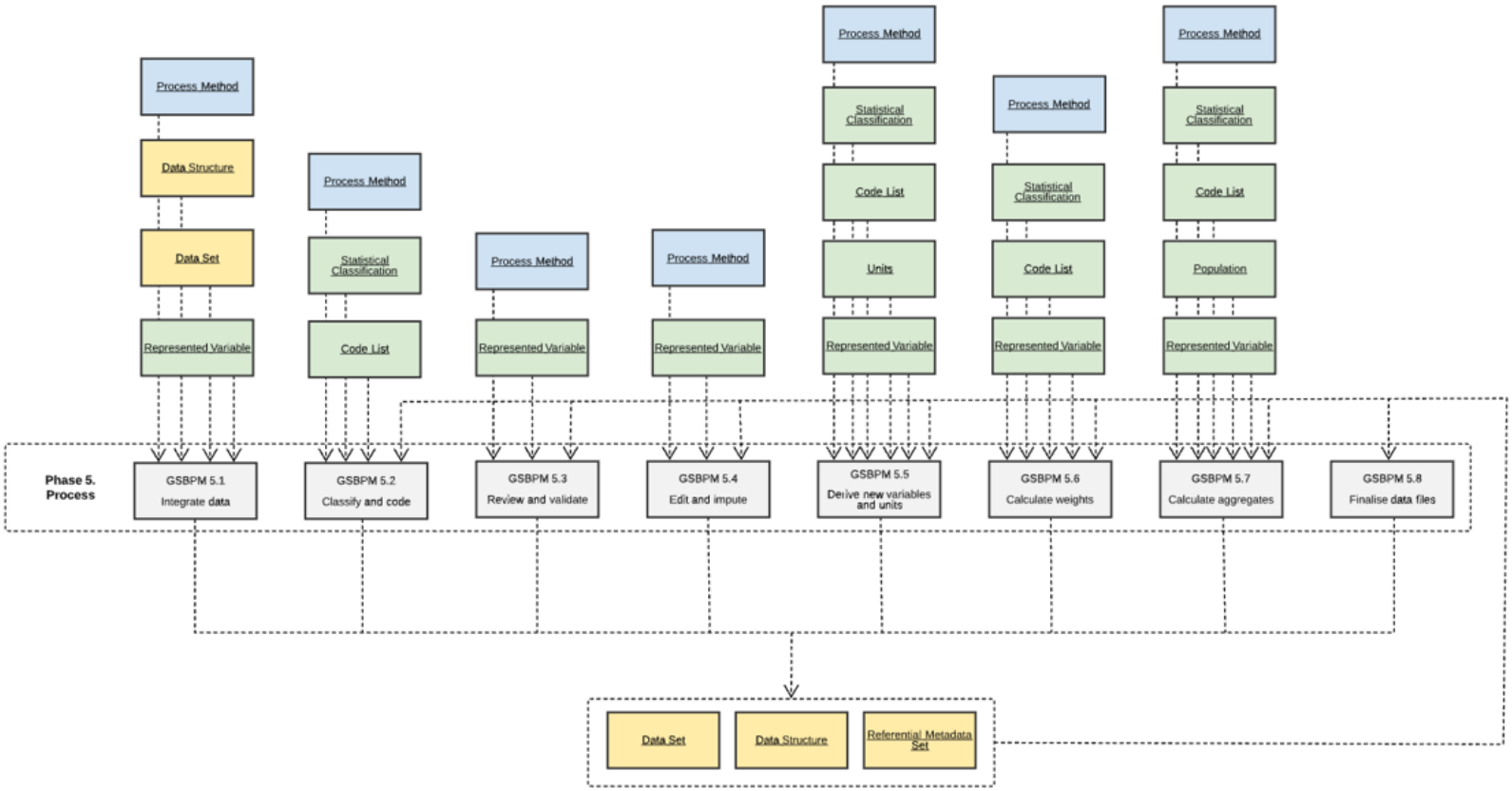
Process Input Specification	*Process Design*	*Process Output Specification*
***Core Input* type:** ***Parameter Input* type:** ***Process Support Input* type:**	***Process Method:***	***Core Output* type:** ***Process Metric* type:** ***Process Execution Log* type:**
***Process Control Design*:**		

Figure 1. Diagram for the specification level



Linking GSIM & GSBPM

Phase 5. Process



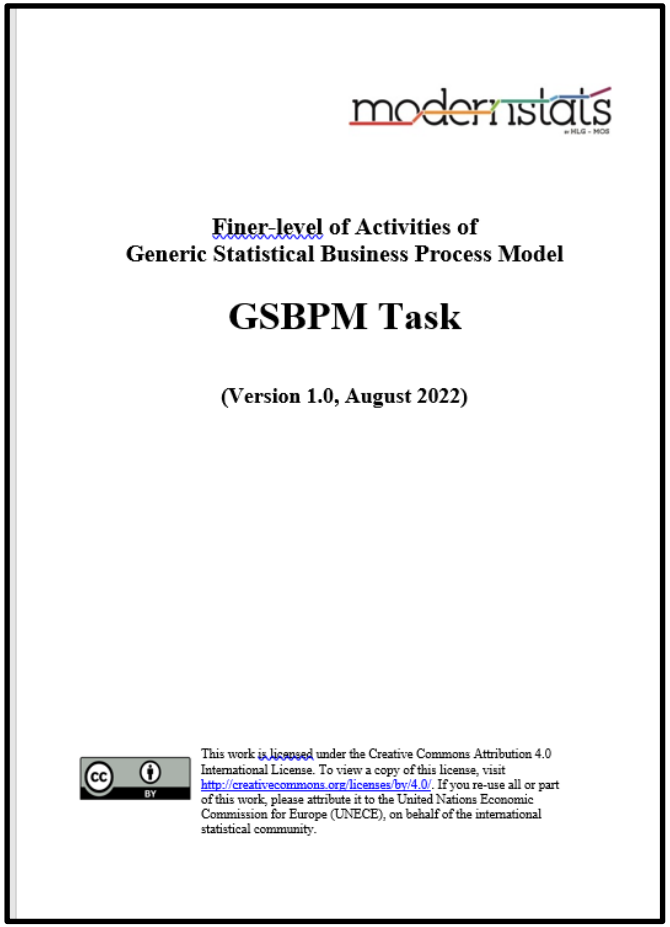
Linking GSIM & GSBPM

Sub-process 5.4 Edit and impute

Specification level

Process Input Specification	***Process Design***	***Process Output Specification***
***Core Input* type** *Data Sets* to be edited and imputed *Data Structures* specifying the structure of *Data Sets* *Represented Variables* to be edited and imputed *Process Methods* specifying methodology and associated *Rules* to conduct editing & imputation (E&I) ***Parameter Input* type** Parameter values to be used for editing and imputation methodologies as specified in *Process Method* such as: • Editing/imputation model parameter • Choice of edit rule (e.g., hard vs. soft) • Choice of model assumptions or hypotheses (e.g., parametric vs. non-parametric; linear relationship vs. non-linear relationship) • Threshold values ***Process Support Input* type** *Data Set:* auxiliary *Data Set* *Data Structure* specifying the structure of the auxiliary *Data Set* Technical / methodological handbooks, policies or guidelines to be followed regarding E&I as well as quality management	***Process Method*** Review *Data Sets* and *Process Methods* Apply *Process Methods* and *Rules* to edit and impute *Represented Variables* in *Data Sets* Calculate quality measures specified by *Process Methods* Update *Data Sets* and associated element in *Data Structure* with results from review and validation	***Core Output* type** *Data Set* (*Unit Data Set*) edited and imputed *Data Structure* specifying the structure of *Data Set* (e.g., if there are flags that correspond to additional *Represented Variables*) *Referential Metadata Set*: descriptions of the *Process Methods* used, quality information summarising *Process Metrics* or any other relevant information to be passed along with *Data Sets* ***Process Metric* type** Quality measures related to E&I such as: • Edit failure rate • Imputation rate Quality measures of *Process Step* such as: • Time spent to complete the *Process Step* (derived from *Process Execution Log)* • Cost spent to complete the *Process Step* ***Process Execution Log* type** Execution log such as • Time that *Process Step* started • Time that *Process Step* ended • Any message or event log generated from software used for E&I (e.g., completion of E&I algorithms)

GSBPM Tasks



modernistats

Finer-level of Activities of
Generic Statistical Business Process Model

GSBPM Task

(Version 1.0, August 2022)

"The purpose of this document is to compile examples from countries that have added lower-level activities to their national version of GSBPM, and to develop a **list of common tasks for all GSBPM sub-processes** without losing the generic nature of the model under a set of principles (e.g., coding, granularity, minimality, description format) with reference to the current version of the GSBPM (version 5.1)."

"The list of tasks was developed based mainly on the description of each sub-process in the GSBPM, by adding country examples as well as feedback received from the community. Each task uses the format of "Action verb + an object" as well as with hierarchical coding (e.g., 1.1.1, 1.1.2, etc.)."

GSBPM Tasks: example

2.1. Design outputs

24. This sub-process contains the detailed design of the statistical outputs, products and services to be produced, including the related development work and preparation of the systems and tools used in the "Disseminate" phase. Processes governing access to any confidential outputs are also designed here. Outputs should be designed to follow existing standards wherever possible, so inputs to this process may include metadata from similar or previous collections (including extractions from statistical, administrative, geospatial and other non-statistical registers and databases), international standards, and information about practices in other statistical organisations from sub-process 1.1 (Identify needs). Outputs may also be designed in partnership with other interested bodies, particularly if they are considered to be joint outputs, or they will be disseminated by another organisation.

2.1.1 Design statistical concepts
2.1.2 Design statistical outputs, products and services and necessary components
2.1.3 Design dissemination procedure
2.1.4 Design systems and tools for dissemination

Note: 2.1.1: Sub-process 1.4 (Identify concepts) touches on concepts, but it does not seem that it is where we decide on concepts. Design of concepts is crucial for design of output.

Note: 2.1.2: e.g., it could include quality level, dissemination calendar, access to confidential outputs and examples on necessary components are table, chart and metadata.

The Istat experience on business process documentation: SIDI/SIQual

Istat official information system for documenting quality and reference metadata of the statistical production processes

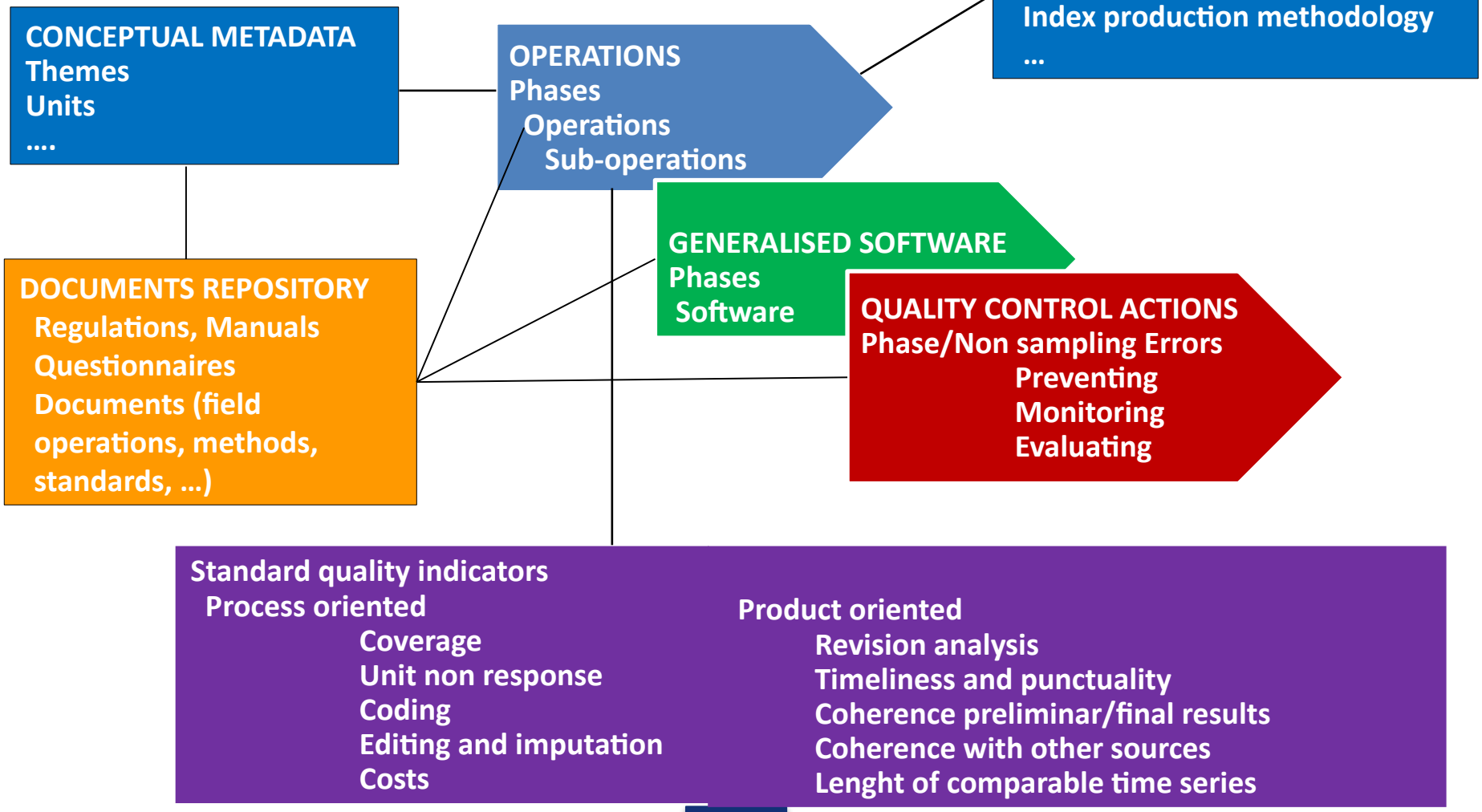
SIDI «input» system, SIQual for consultation

SIDI first implementation in 2001, later converted in a web-based architecture

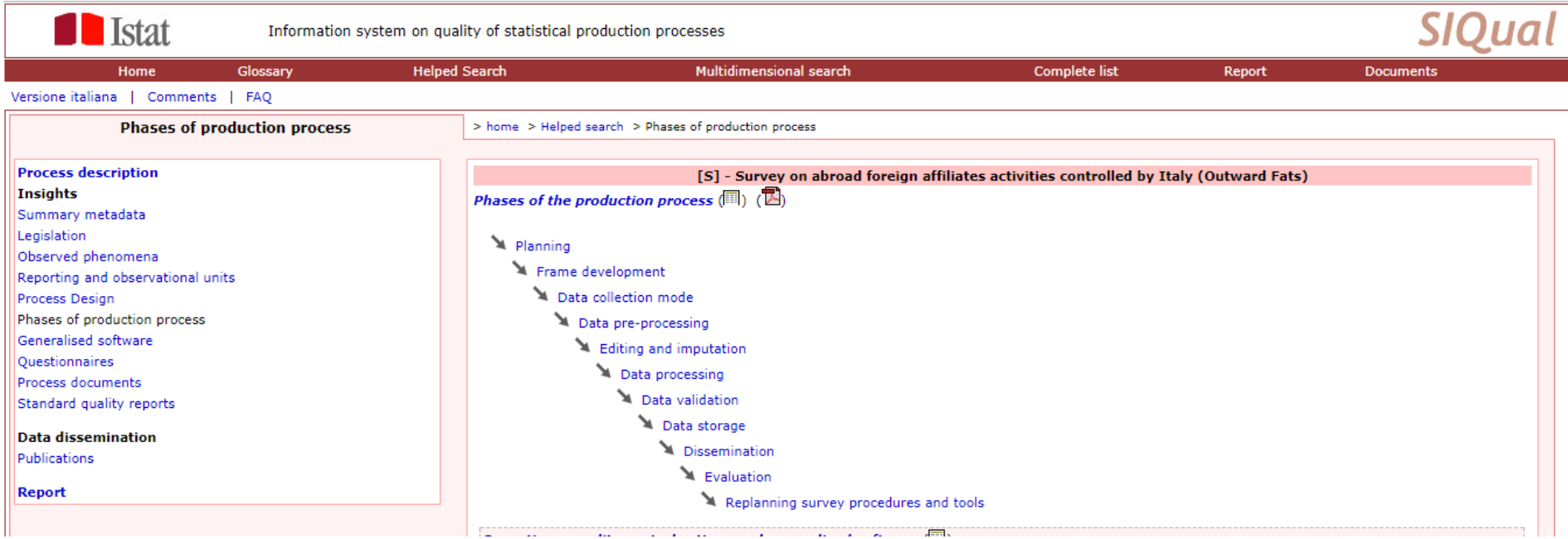
SIQual first release in 2005, English version released in 2008

SIDI/SIQual approach to documentation is highly structured and standardised, descriptive additional fields are available to better describe standard items

SIDI/SIQual contents



SIDI/SIQual example of documentation



SIDI/SIQual example of documentation

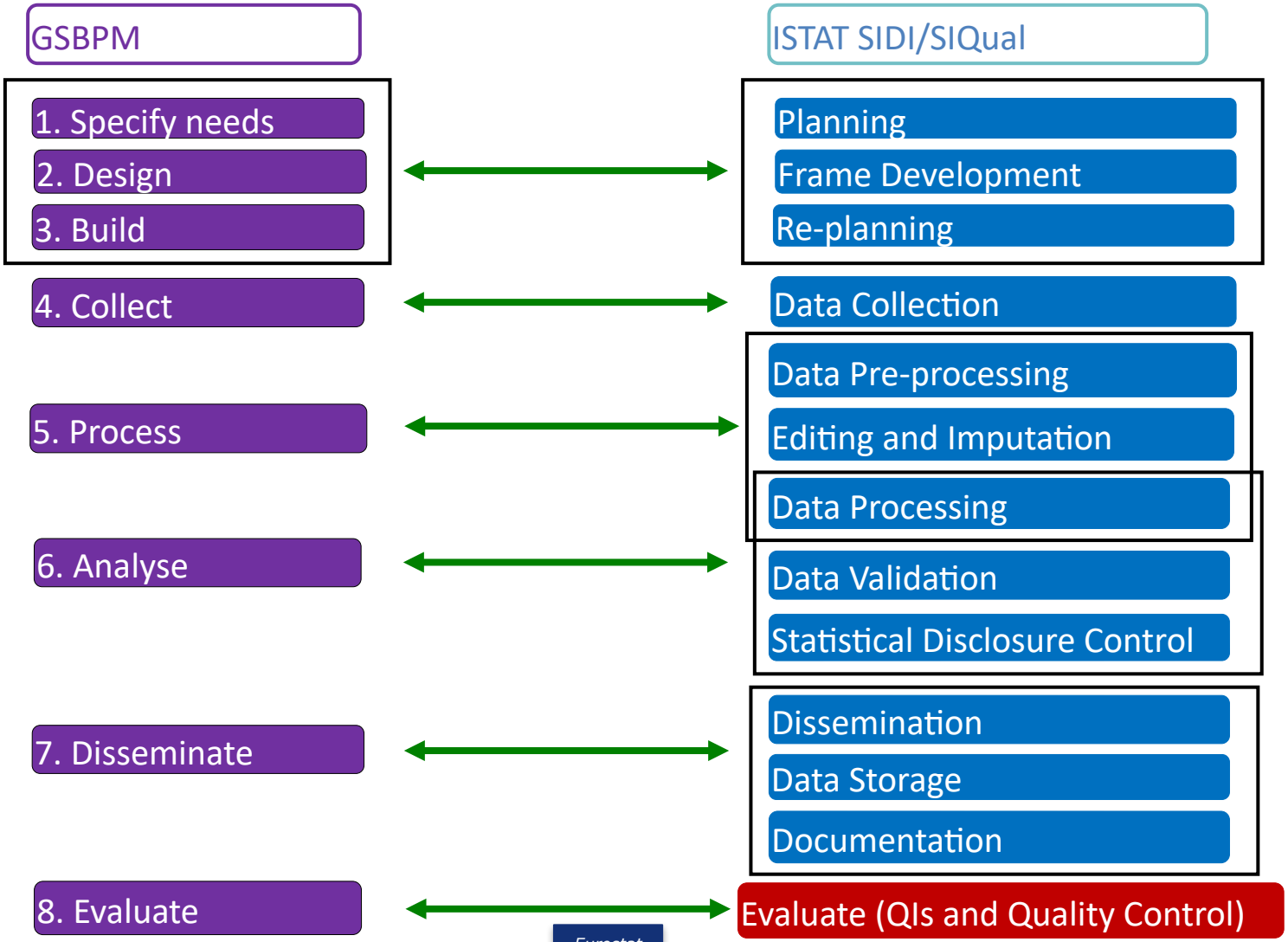
Operations, quality control actions and generalised software (📄)

Data collection mode

Collapse

- Self-administered data collection via electronic transmission
  - Self-administered data collection by e-questionnaire (Computer Assisted Web Interviewing-CAWI) or through upload of datasets on Istat web site INDATA *(since 31/12/2007)*
    - **Control on unit nonresponse**
      - **Prerequisites for control on unit nonresponse**
        - **Recording and coding of contact outcomes for e-management of follow-ups *(since 31/12/2007)***
          - Use of the monitoring system for structural business surveys (PERVENUTO) *(since 31/12/2007)*
        - Control on interview outcome codes (completed, refused, noncontacted, duplicated, etc.) *(since 31/12/2007)*
      - **Activities for preventing unit nonresponse**
        - Survey presentation letter signed by Istat President *(since 31/12/2007)*
        - Guarantees on statistical confidentiality *(since 31/12/2007)*
        - Written description of survey objectives *(since 31/12/2007)*
        - Special care in writing clear instructions to fill-in questionnaires *(since 31/12/2007)*
        - Establishing a toll free line or telephone number for further explanations *(since 31/12/2007)*
        - Administrative fines for nonrespondents *(since 31/12/2007)*
        - Special care in drafting clear instructions for database setting up *(since 31/12/2007)*
        - Establishing an e-mail address for information on the web procedure *(since 31/12/2007)*
      - **Follow-ups of nonrespondent units**
        - Telephone follow-ups *(since 31/12/2007)*
        - Mail follow-ups *(since 31/12/2007)*
        - Follow-ups by Postel *(since 31/12/2007)*
        - Follow-ups by e-mail *(since 31/12/2007)*

Mapping GSBPM and SIDI/SIQual model

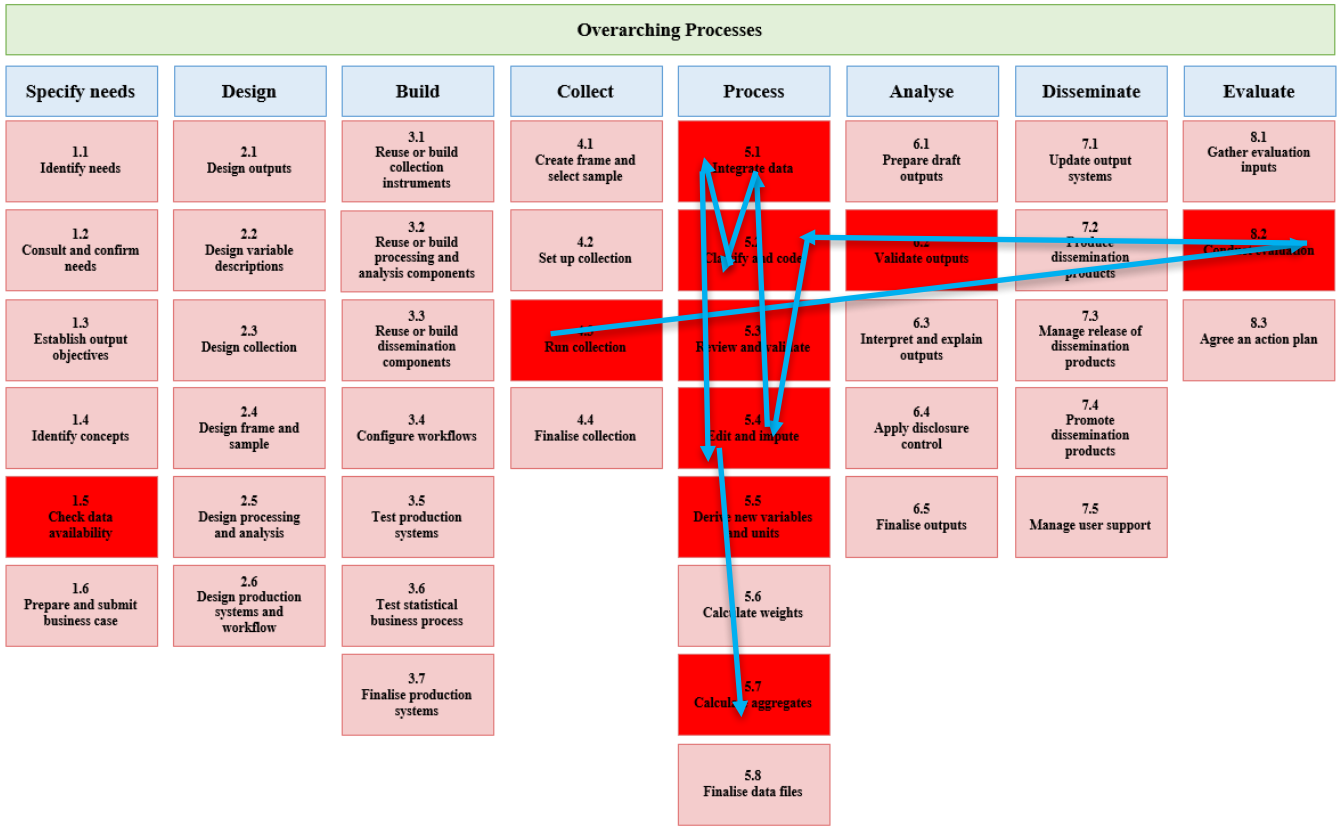


SURVEY	GSBPM1	GSBPM2	SIDI/SIQual
LFS	Design	Design collection	Planning analysis unit contact and observation: direct interview, telephone interview, photocopy of administrative documents, etc.
			Planning data collection modes
			Planning data entry modes
			Questionnaire design
		Design frame and sample	Frame implementation planning
		Design processing and analysis	Planning data analysis and tabulation procedures
			Planning editing and imputation methods
			Definition procedures for automatic editing and imputation
			Classification and coding planning
	Collect	Create frame and select sample	Drawing the list of reporting units from a non-Istat frame (e.g. Chambers of Commerce, Industry and Crafts - C.C.I.A.A.)
		Run collection	Computer Assisted Telephone Interviewing (CATI)
			Computer Assisted Personal Interviewing (CAPI)
	Process	Classify and code	Computer assisted coding supported by interactive software (expert systems)
		Review and validate	Review based on constraints among records of the same survey
			Error detection based on consistency edits
			Automatic error detection during data entry based on consistency edits
		Edit and impute	Probabilistic error and outlier detection and nearest-neighbour donor imputation
			Deterministic error and outlier detection and imputation based on deterministic rules (IF-THEN)
		Calculate aggregates	Estimate calculation
			Setting up final macrodata files for Eurostat
		Finalise data files	Setting up final microdata files for Eurostat
			Setting up final microdata files for Istat National Accounts Division
	Analyse	Prepare draft outputs	Processing final indexes
			Use of seasonal adjustment procedures
		Validate outputs	Coherence control with previous data of the same survey
	Disseminate	Update output systems	Dissemination in Istat data bases accessible at the Data shop, Regional Offices or other locations
		Produce dissemination products	Publication of insights or specific studies (Series "Argomenti", "Metodi e Norme", etc.)
			Publishing data in volumes of International Organisations (OECD, Eurostat, etc.)
			Publication of final data on survey-specific volumes (Series Yearbooks, "Informazioni", etc.)
			Press release dissemination ("Statistiche-Flash", "Statistiche-Report", "Statistiche-Focus", "Note informative")
			Publication of data on Istat general volumes (Statistical Yearbook, "Conoscere l'Italia", Italian Statistical Compendium, etc.)
		Manage release of dissemination products	Release of File Standard
			Microdata File for Research (MFR) release
	Evaluate	Gather evaluation inputs	Disseminating quality indicators supporting statistical information
			Developing electronic tools for quality indicators computation
	Overarching	Data management	Data storage in a local repository
			Microdata file storage in Istat Validated microdata repository (ARMIDA)

Applying GSBPM to multisource statistical processes: work in progress at Istat

- One of the pillars of Istat modernisation process has been the System of Integrated Statistical registers.
- Each Statistical register of the system is a complex multisource statistical process
- An internal Working Group designed a system of quality indicators to monitor and assess such complex processes, also with a view to feeding the new metadata system of the Institute (that is being designed, also)
- GSBPM was used to map this processes

Applying GSBPM to multisource statistical processes: work in progress at Istat



- Focus on a subset of GSBPM sub-processes
- The workflows can be different. The one reported is only an example

Metadata model for each GSBPM sub-process

Macro Item	GSIM Object
Input	Transformable input
	Parameter
	Process support input
GSBPM suprocess	Business Function
	Business process (GSBPM phase)
	Process step (GSBPM sub-process)
	Quality control actions
	Process Method
	Rule
	Software
Output	Transformed output
	Process Metric (Quality indicators)
	Process Execution Log

Each object is accompanied by a definition and an example in Italian. E.g.: Parameter
«*Oggetti forniti in input al sotto-processo per configurare il sotto-processo stesso.*
Es.: I parametri di un modello di stima»
«Objects provided as input to the sub-process, to configure the sbprocess itself
E.g.: the parameter of a model for estimating data»

The model has been developed for each GSBPM sub-process that was considered relevant for the process and tested on 2 statistical registers:

- the Base register of individuals and households RBI
- the extended register of principal economic variables FRAME-SBS

Model for Data Integration

Macro Item	GSIM Object	Possible values
Input	Transformable input	Data-set1, Data-set2, …(data structure: units and variables)
	Parameter	Thresold, Linkage keys, blocking variables
	Process support input	Furher variables useful for identification other than the keys or to control the matching
GSBPM suprocess	Business Function	Increasing units, increasing variables, increasing both
	Business process (GSBPM phase)	5. Process
	Process step (GSBPM sub-process)	5.1. Integrate data
	Quality control actions	Actions for preventing, monitoring, reducing errors due to integration
	Process Method	Record linkage (deterministic, hierarchical, probabilistic, privacy preserving and predictive linkages (classification or regression techniques); Statistical matching; Appending procedures; Data pooling; Integration base on data surce prioritisation
	Rule	Integration model, Rules for the hyerarchical selection of the sources, transofrmation rules
	Software	Relais, Statmatch, Ad hoc procedures
Output	Transformed output	Integrated Data set, Non linked records data sets
	Process Metric (Quality indicators)	SEE NEXT SLIDE
	Process Execution Log	Integration time

Quality indicators for data integration

Indicators on data integration performance
4.1. Missing values or errors in linkage variable
4.2. Match rate
4.3. False link rate
4.4. False non-link rate
Indicators on units
4.5. Percentage of units from different datasets on unit total
Indicators on variables
4.6 Percentage of variables from different input datasets on total number of variables in the integrated dataset
4.7 Distances between variable distributions on the integrated dataset and on the input datasets
4.8 Number of variables derived at the end of integration

Application to RBI – variable education level last integration step

Macro Item	GSIM Object	Values
Input	Transformable input	Dataset RBI2019 (AGE>=9 e residente=1), dataset output step 6, dataset APR4, Master sample census
	Parameter	CODICE_INDIVIDUO
	Process support input	-
GSBPM suproces s	Business Function	ncreasing variables (add education level to RBI)
	Business process (GSBPM phase)	5. Process
	Process step (GSBPM sub-process)	5.1. Integrate data
	Quality control actions	-
	Process Method	Deterministic Record linkage
	Rule	Left join with RBI as reference; pop_abc =A if individual is in BIT, pop_abc=B if individual is in CENS11 and not in BIT, pop_abc=C if individual is not in BIT and not in CENS11
	Software	Oracle procedure
Output	Transformed output	Integrated Data set with all RBI units and with variables G_ISTR, tit_stu, pop_abc
	Process Metric (Quality indicators)	SEE NEXT SLIDE
	Process Execution Log	-

Quality indicators on data integration: test on RBI

Application to integration step of variable education level

Data source	4.1: missing key	4.2: Match rate	4.5: Hyerarchical coverage
MS 2019	0,195%	92,882%	4,711%
BIT 2017	0%	88,404%	22,213%
CENS 2011	0,001%	88,645%	68,345%
RBI 2019	0%	n.c.	n.c.

Reference

UNECE (2019) Generic Statistical Business Process Model GSBPM (Version 5.1, December 2013)

https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1

UNECE (2020) Generic Statistical Information Model (GSIM): Communication Paper for a General Statistical Audience (Version 1.2, October 2020)
https://statswiki.unece.org/display/gsim/GSIM+v1.2+Communication+Paper

UNECE (2019) Generic Activity Model for Statistical Organisations

GAMSO (Version 1.2, January 2019) https://statswiki.unece.org/display/GAMSO/GAMSO+v1.2

UNECE (2017) 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+Home

UNECE (2022) Linking GSBPM & GSIM Version 1.0, January 2022

https://statswiki.unece.org/display/GSBPM/Information+flow+within+GSBPM+using+GSIM

UNECE (2022) Finer-level of Activities of Generic Statistical Business Process Model - GSBPM Task Version 1.0, August 2022 https://statswiki.unece.org/display/GSBPM/GSBPM+Tasks