

28 April 2023

Modernizing management systems: modeling processes using ModernStats models

ModernStats models GSBPM and GSIM as tools for quality monitoring, documentation and assessment

Giorgia Simeoni, Istat Quality Manager

Quality management at Istat

- Activities on quality management started more than 20 years ago at Istat
- Documentation and process orientation have always been among the pillars of Istat approach to quality
- SIQual, the Istat Information System for Quality of Statistical processes, documents statistical processes, including their quality assurance system and their quality indicators since the early 2000s.
- Audit and self-assessment on Istat statistical processes started in 2010 and the compliance with *Quality guidelines for statistical processes* was verified

SIQual model for the proc

Phases of the production process



- Face to face interviewing
 - Computer Assisted Personal Interviewing (CAPI) (since 28/03/2004)
 - Control on interviewers
 - Prerequisites for control on interviewers
 - Interviewer-questionnaire matching by identification codes (since 28/03/2004)
 - Collection of information on interviewer characteristics (since 28/03/2004)
 - Availability of a database on interviewers (since 28/03/2004)
 - Activities for preventing interviewer errors
 - Training course for interviewers (since 28/03/2004)
 - Drafting an interviewer instruction manual (since 28/03/2004)
 - Training course for managing staff of data collecting bodies/institutions provided by the section in charge of survey (since 28/03/2004)
 - Control on interviewers during field operations
 - Debriefing with interviewers on data collection problems (since 28/03/2004)
 - Supervision of interviewers by observing interviews (since 28/03/2004)
 - Monitoring response rates per interviewer during data collection (since 28/03/2004)
 - Methods to estimate interviewer effects
 - Ex post evaluation of interviewers performance based on indicators (since 28/03/2004)
 - Control on unit nonresponse
 - Activities for preventing unit nonresponse
 - Survey presentation letter signed by Istat President (since 28/03/2004)
 - Guarantees on statistical confidentiality (since 28/03/2004)
 - Interviewer identification badge (since 28/03/2004)
 - Description of survey objectives by interviewers (since 28/03/2004)
 - Telephone contacts to make an appointment for the interview (since 28/03/2004)
 - Collecting information on nonrespondent units or bodies/institutions
 - Collecting variables on nonrespondent units during data collection (since 28/03/2004)
 - Collecting interviewer identification codes of nonrespondent units (since 28/03/2004)
 - Collecting nonresponse causes during data collection (since 28/03/2004)
 - Collecting identification codes of nonrespondent units (since 28/03/2004)
 - Use of adjustment methods to reduce the unit nonresponse effects (weights, duplications of records, etc.)
 - Use of weighting methods (since 28/03/2004)
 - Control on the responses provided by units different from those included in the primary list
 - Actions to increase the number of respondents
 - Nonrespondent unit substitution (since 28/03/2004)
 - Use of proxy respondents (since 28/03/2004)
 - Prerequisites for control on responses provided by units different from the target units
 - Collecting identification codes of substituting units (since 28/03/2004)
 - Collecting proxy respondent identification codes (since 28/03/2004)
 - Questions on proxy respondents aiming at knowing their characteristics (since 28/03/2004)
 - Methods to estimate effects of responses provided by units different from the target units
 - Studies to assess substitution effects on data quality (since 28/03/2004)
 - BLAISE for data capturing (since 28/03/2004)
 - Use of SIGIF (Sistema di Gestione delle Indagini sulle Famiglie) (since 28/03/2004)
- Telephone interviewing
 - Computer Assisted Telephone Interviewing (CATI) (since 28/03/2004)

Quality guidelines for statistical processes

- Istat developed several quality guidelines, they contain the basic principles for planning, implementing and assessing statistical processes and a description of methods for ensuring compliance with these principles.
- The firsts to be developed were the Quality guidelines for statistical processes.
- They were oriented to surveys, the first section was devoted to the process quality and organised according to survey phases
- Also the questionnaire supporting audits and sel-assessment was developed following the same structure

Section I: Process Quality

A. Survey Objectives

- A.1. Information needs, users and uses

B. Survey Design

- B.1. Response burden
- B.2. Survey design for total surveys
- B.3. Sample design
- B.4. Questionnaire design and testing

C. Survey Frames

- C.1. Frame updates and coverage

D. Data collection

- D.1. Use of administrative data
- D.2. Direct data collection

E. Data Processing

- E.1. Coding
- E.2. Data capture
- E.3. Error detection and treatment
- E.4. Integration of data sources
- E.5. Estimation
- E.6. Seasonal adjustment
- E.7. Revision policy
- E.8. Data validation

F. Data storage, dissemination and documentation

- F.1. Data storage, dissemination and documentation

... and then GSBPM was released

- After its release, GSBPM suddenly started to be widely used in Official Statistics and it was clear that we should take it into consideration
- At the beginning we were a bit worried, but then we realised that we can map our SIQual model to GSBPM and we identified pros and cons of the 2 approaches:
 - SIQual model is more detailed and more quality oriented
 - GSBPM is more flexible and less «survey oriented»
- The mapping between SIQual Phases and operations and GSBPM phases and subprocesses was stored in the database, so that, on request, we can provide the SIQual process documentation according to GSBPM

Overarching Processes							
Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Reuse or build collection instruments	4.1 Create frame and select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult and confirm needs	2.2 Design variable descriptions	3.2 Reuse or build processing and analysis components	4.2 Set up collection	5.2 Classify and 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 Reuse or build dissemination components	4.3 Run collection	5.3 Review and validate	6.3 Interpret and explain outputs	7.3 Manage release of dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame and sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit and impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing and analysis	3.5 Test production systems		5.5 Derive new variables and units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare and submit business case	2.6 Design production systems and workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production systems		5.7 Calculate aggregates			
				5.8 Finalise data files			

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	Design frame and sample	Frame implementation planning
			Design processing and analysis
	Design processing and analysis	Planning data analysis and tabulation procedures	
		Planning editing and imputation methods	
	Collect	Create frame and select sample	Definition procedures for automatic editing and imputation
			Classification and coding planning
	Collect	Run collection	Drawing the list of reporting units from a non-Istat frame (e.g. Chambers of Commerce, Industry and Crafts - C.C.I.A.A.)
			Computer Assisted Telephone Interviewing (CATI)
	Process	Classify and code	Computer Assisted Personal Interviewing (CAPI)
			Computer assisted coding supported by interactive software (expert systems)
	Process	Review and validate	Review based on constraints among records of the same survey
			Error detection based on consistency edits
	Process	Edit and impute	Automatic error detection during data entry based on consistency edits
			Probabilistic error and outlier detection and nearest-neighbour donor imputation
	Process	Calculate aggregates	Deterministic error and outlier detection and imputation based on deterministic rules (IF-THEN)
			Estimate calculation
Analyse	Finalise data files	Setting up final macrodata files for Eurostat	
		Setting up final microdata files for Eurostat	
Analyse	Prepare draft outputs	Setting up final microdata files for Istat National Accounts Division	
		Processing final indexes	
Disseminate	Validate outputs	Use of seasonal adjustment procedures	
		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.)
Disseminate	Manage release of dissemination products	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.)	
Evaluate	Gather evaluation inputs	Release of File Standard	
		Microdata File for Research (MFR) release	
Over arching	Data management	Disseminating quality indicators supporting statistical information	
		Developing electronic tools for quality indicators computation	
Over arching	Data management	Data storage in a local repository	
		Microdata file storage in Istat Validated microdata repository (ARMIDA)	

GSBPM in the Istat quality tools

- Gradually, GSBPM become our starting point when developing quality tools related to statistical processes.
- As an example, let's see the Quality guidelines for the statistics produced by the National Statistical System (available in Italian)
- We used these guidelines as a reference for the audits to the Other National Authorities (ONAs) producing European Statistics

Tabella 1. Fasi e/o sotto-processi del GSBPM e Sezioni della Parte II delle Linee Guida

Fasi e sotto-processi GSBPM	Sezioni del manuale
Specify needs (1.1. – 1.2. – 1.3. – 1.4. – 1.5.), Design outputs (2.1.), Design variable description (2.2.), Gather evaluation inputs (8.1.)	A. Identificazione delle esigenze degli utenti, definizione dei concetti scelta delle fonti e valutazione della soddisfazione
Design frame & sample (2.4.), Build or enhance process components (3.2.), Create frame & select sampe (4.1.), Calculate weights (5.6.), Calculate aggregates (5.7.), Gather evaluation inputs (8.1.)	B. Scelta del disegno, lista di riferimento, campionamento e stima
Design collection (2.3.), Build collection instrument (3.1.), Build or enhance process components (3.2.), Test production system (3.5.), Set up collection (4.2.), Run collection (4.3.), Gather evaluation inputs (8.1.)	C. Acquisizione dei dati
Design collection (2.3.), Build collection instrument (3.1.), Build or enhance process components (3.2.), Test production system (3.5.), Finalise collection (4.4.), Gather evaluation inputs (8.1.)	D. Conversione in formato elettronico (registrazione)
Design processing and analysis (2.5.), Test production system (3.5.), Integrate data (5.1.), Gather evaluation inputs (8.1.)	E. Integrazione
Design processing and analysis (2.5.), Test production system (3.5.), Classify & code (5.2.), Gather evaluation inputs (8.1.)	F. Codifica e classificazioni
Design processing and analysis (2.5.), Test production system (3.5.), Review & validate (5.3.), Edit & impute (5.4.), Gather evaluation inputs (8.1.)	G. Identificazione e trattamento degli errori
Design processing and analysis (2.5.), Test production system (3.5.), Derive new variables and units (5.5.), Gather evaluation inputs (8.1.)	H. Derivazione delle unità
Design processing and analysis (2.5.), Test production system (3.5.), Derive new variables and units (5.5.), Gather evaluation inputs (8.1.)	I. Derivazione delle variabili
6.1. Prepare draft outputs	J. Destagionalizzazione
From Run collection (4.3.) to Finalise outputs (6.5.)	K. Politica delle revisioni
Design processing and analysis (2.5.), Test production system (3.5.), Validate outputs (6.2.), Gather evaluation inputs (8.1.)	L. Validazione dei risultati
Design processing and analysis (2.5.), Build or enhance dissemination components (3.3.), Apply disclosure control (6.4.), Disseminate (7.1. – 7.2. – 7.3. – 7.4. – 7.5.), Gather evaluation inputs (8.1.)	M. Diffusione dei dati e tutela della riservatezza, archiviazione, documentazione

GSBPM in the quality tools

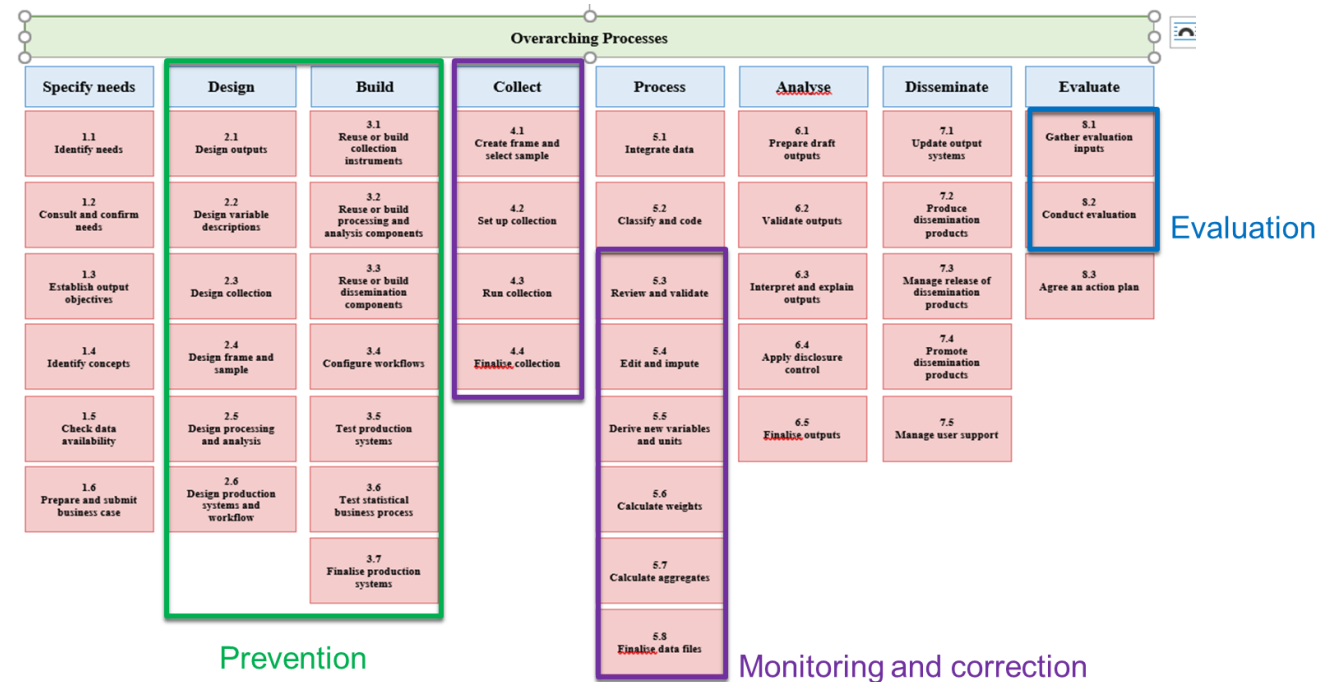
- More recently, we developed a checklist to verify the conformity of Istat traditional processes (like surveys) to standard methodologies and procedures, with the aim of assigning an internal quality label to compliant processes. Also in this case the starting point was GSBPM



GSBPM Phase	GSBPM sub process	Section Istat checklist
1.Specify Needs	1.1 Identify needs	Section A – User needs/concepts definition
	1.2 Consult and confirm needs	
	1.4 Identify concepts	
	1.5 Check data available	
2.Design 4.Collect	2.4 Design frame and sample methodology	Section B - Frame and sample design
	4.1 Create frame and select sample	
3.Build 4.Collect	3.1 Reuse or build collection instruments	Section C – Data Collection
	4.2 Set up collection	
	4.3 Run collection	
5.Process	5.1 Integrate data	Section D- Integration
	5.2 Classify and code	Section E – Coding and classification
	5.3 Review and validate	Section F –Investigation and treatment of non-sampling errors
	5.4 Edit and impute	Section G – Derivation of units
	5.5 Derive new variables and units	
	5.5 Derive new variables and units	Section H –Derivation of variables
	5.6 Calculate weights	Section I – Calculation of weights and aggregates
5.7 Calculate aggregates		
6.Analyse	6.1 prepare draft outputs	Section L –Validation of outputs
	6.4 Apply disclosure control	Section M- Disclosure control
7.Disseminate		Section N - Dissemination
8.Evaluate		Section O - Documentation and evaluation
Evaluate checklist		Section P – Checklist evaluation

GSBPM in the quality tools

- GSBPM is also used in internal training course on quality at Istat.
- Indeed, in the introductory course we have a session on process quality and quality assurance system and we use GSBPM to describe it



Facing quality aspect of modernised processes

- Till now, we saw how we used GSBPM to model mainly traditional processes
- Since 2016 Istat started a modernisation programme. One of the pillars of the programme is the building of the Integrated System of Statistical Registers (ISSR)
- ISSR consists in a number of coherent registers to produce several types of statistical outputs.
- Each statistical register is obtained by integrating sources of different typology, mainly administrative data, but also survey results or other registers, such as to create new processes that can vary a lot in complexity.
- The quality framework for monitoring, assessing and documenting these processes needed to be defined and two internal working groups worked to this aim.

Quality framework for ISSR

- The main objective of the working groups was to define a system of quality indicators, but Istat approach has always been to accompany quality indicators with the metadata needed to correctly interpret them.
- Thus we defined a general metadata model able to describe the complex multisource processes that are carried out currently to create every edition of a register and their quality.
- GSBPM was immediately considered as the reference model: the processes of different registers were analysed to identify the most relevant GSBPM subprocesses
- The work was not so easy, some quality assurance activities are integrated in the ISSR processes and the lack of specification of GSBPM overarching quality management made it sometimes difficult to map and describe these activities with GSBPM

Definition of the main GSBPM subprocesses to be considered

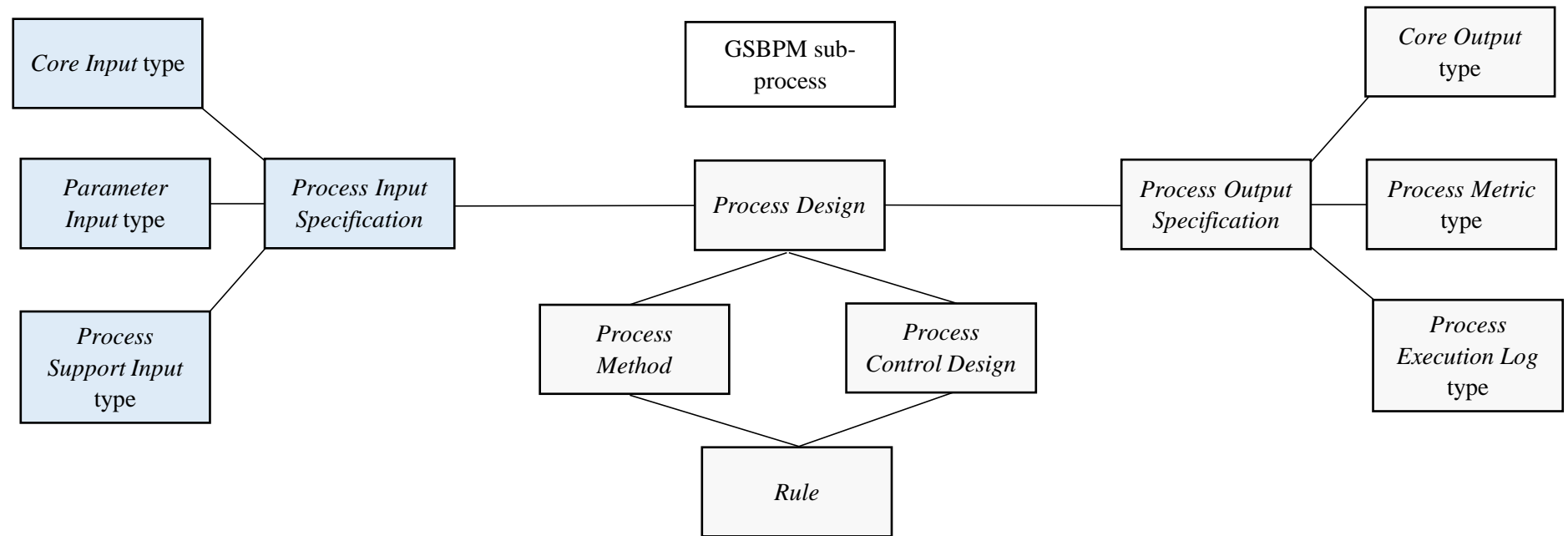
Overarching Processes							
Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Reuse or build collection instruments	4.1 Create frame and select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult and confirm needs	2.2 Design variable descriptions	3.2 Reuse or build processing and analysis components	4.2 Set up collection	5.2 Classify and 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 Reuse or build dissemination components	4.3 Run collection	5.3 Review and validate	6.3 Interpret and explain outputs	7.3 Manage release of dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame and sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit and impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing and analysis	3.5 Test production systems		5.5 Derive new variables and units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare and submit business case	2.6 Design production systems and workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production systems		5.7 Calculate aggregates			
				5.8 Finalise data files			

Definition of the main GSBPM subprocesses to be considered

Overarching Processes							
Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Reuse or build collection instruments	4.1 Create frame and select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult and confirm needs	2.2 Design variable descriptions	3.2 Reuse or build processing and analysis components	4.2 Set up collection	5.2 Classify and 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 Reuse or build dissemination components	4.3 Run collection	5.3 Review and validate	6.3 Interpret and explain outputs	7.3 Manage release of dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame and sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit and impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing and analysis	3.5 Test production systems		5.5 Derive new variables and units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare and submit business case	2.6 Design production systems and workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production systems		5.7 Calculate aggregates			
				5.8 Finalise data files			

Quality framework for ISSR

- Then for each subprocess the metadata elements were identified according to the GSIM Information Objects of Business Group useful to describe the Process step.
- Great inspiration was taken from the activities of the UNECE Task team Linking GSBPM and GSIM



Metadata model for each GSBPM sub-process

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

In order to make the model more easily usable the description of each GSIM IO 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 set of possible values for each Item of each subprocess has been identified, including quality indicators as process metrics

The framework has been tested on 2 registers. Refinements were then made to the framework on the basis of the test. The framework will be gradually applied on the different registers of the ISSR.

Model for Data Integration

Macro Item	GSIM Object	Possible values
Input	Transformable input	Data-set 1, Data-set 2, ... (data structure: units and variables)
	Parameter	Threshold, Linkage keys, blocking variables
	Process support input	Further variables useful for identification other than the keys or to control the matching
GSBPM subprocess	Business Function	Increasing units, increasing variables, increasing both
	Business process (GSBPM phase)	5. Process
	Process step (GSBPM sub-process)	5.1. Integrate data
	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 source prioritisation
	Rule	Integration model, Rules for the hierarchical selection of the sources, transformation rules
	<i>Software</i>	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

Pros and cons in the use of GSBPM and GSIM

- The possibility to refer to GSBPM and GSIM simplified the development of the metadata model and allowed to concentrate resources on quality indicators
- Nonetheless the definition of the model and its application are complex and resource demanding:
 - The lack of subprocesses for the overarching quality management caused a lot of discussion on how to model some parts of the process
 - Presenting GSIM IO is necessary to make several examples because the definitions rarely results clear to non-experts
 - To apply the model to the process of a statistical register several sessions involving metadata and quality experts as well as experts from the statistical register process and IT experts were needed.

... and finally METAstat

- The quality framework for ISSR will be not only applied to different registers, but also implemented in the new Istat metadata system, METAstat, that is currently being designed.
- The system will substitute SIQual and also the Istat system for structural metadata SUM, that is already implementing GSIM IO from Concept and (partly) Structure Group
- METAstat will manage structural and reference metadata for Istat statistical processes in an integrated way and we are taking inspiration from both GSBPM and GSIM for its design. The project is long-term and we are still at the beginning.
- No more mapping with GSBPM will be needed!!!

Concluding remarks

- We have seen some examples of the use of Modernstats models in quality activities. There are other uses of the models, in particular GSBPM, at Istat.
- At the moment, we can say that GSBPM has become a reference widely known and used at Istat, GSIM is still seen as something complex and the pros in using it are clear mainly to metadata expert.
- Sometimes definitions are not straightforward, examples are useful.
- Adaptation are to be promoted?

References

- Brancato and Simeoni (2012) Istat statistical process modelling and the Generic Statistical Business Process Model: a comparison. European Conference on Quality in Official Statistics (Q2012)
- Istat (2012) Quality guidelines for statistical processes https://www.istat.it/it/files//2013/04/QualityGuidelines_EngVers_1.11.pdf
- Istat (2018) Linee guida per la qualità delle statistiche del Sistema statistico nazionale <https://www.istat.it/it/files/2018/08/Linee-Guida-2.5-agosto-2018.pdf>
- Simeoni et al (2022) A model for documenting and monitoring quality of statistical registers according to GSBPM and GSIM. ModernStats World Workshop 2022 27-29 June 2022, Belgrade, Serbia <https://unece.org/statistics/events/MWWW2022>
- SIQual system: <https://siqual.istat.it>
- UNECE (2019) Generic Statistical Business Process Model GSBPM (Version 5.1, January 2019) <https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1>
- UNECE (2019) Generic Statistical Information Model GSIM Version 1.2 <https://statswiki.unece.org/display/gsim/GSIM+v1.2+main+changes>
- UNECE (2022) Linking GSBPM and GSIM (Version 1.0 January 2022) <https://statswiki.unece.org/display/GSBPM/Information+flow+within+GSBPM+using+GSIM>

Thank you for your
attention!

Giorgia Simeoni | simeoni@istat.it