

Belgrade, Serbia, 27-29 June 2022

ModernStats World Workshop 2022

A model for documenting and monitoring quality of statistical registers according to GSBPM and GSIM

Istat Working Group on Quality for Integrated System of Statistical Registers

Presenter: Giorgia Simeoni

Istat | Directorate for methodology and statistical process design

Background: documentation and quality monitoring at Istat

- Istat has a long tradition on quality monitoring, documentation and assessment of statistical processes.
- SIQual system, since 2001, allows the documentation of statistical processes adopting a model that is mappable with GSBPM, collects and stores time series of quality indicators on different phases of the statistical process. The system is well tailored for «traditional» processes, e.g. surveys

Information system on quality of statistical production processes

SIQual

Home Glossary Helped Search Multidimensional search Complete list Report Documents Comparing processes

Versione italiana | Comments | FAQ | User's guide

Phases of production process > home > Comparing processes > Phases of production process

[S] - European Union Statistics on Income and Living Conditions (EU-SILC)

Phases of the production process (📄) (📊)

- Planning
- Frame development
- Flows of activities through bodies/institutions prior to data collection
- Data collection mode
- Flows of activities through bodies/institutions after data collection
- Data pre-processing
- Editing and imputation
- Data processing
- Data validation
- Data storage
- Dissemination
- Documentation
- Evaluation

Webinfo Privacy policy Contact us Copyright

Information system on quality of statistical production processes

SIQual

Home Glossary Helped Search Multidimensional search Complete list Report Documents Comparing processes

Versione italiana | Comments | FAQ | User's guide

Quality indicators > home > Comparing processes > Quality indicators

[S] - European Union Statistics on Income and Living Conditions (EU-SILC)

INDICATORS ON UNIT NON RESPONSE From 31/12/2004 to 31/12/2019

Italy Geographical detail Time series - data Time series - charts Back

Comparison with averages Geographical detail

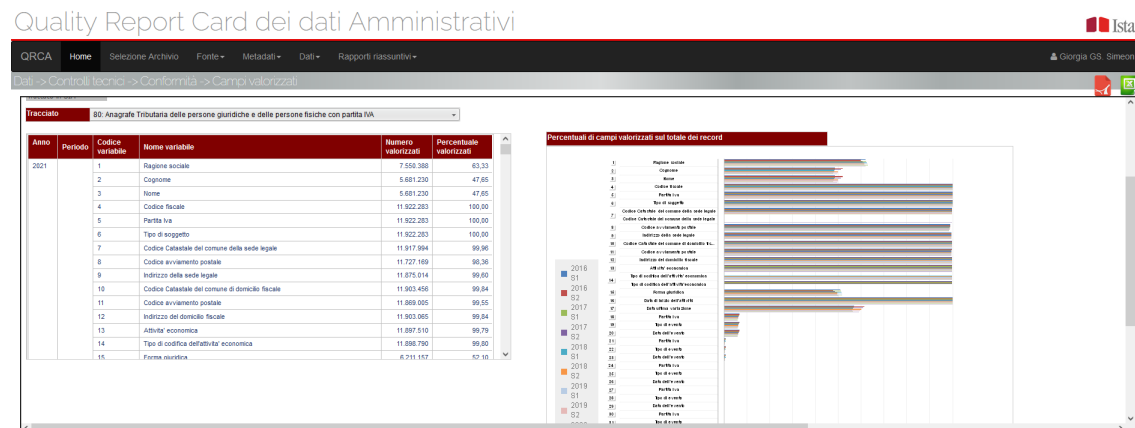
Unit Nonresponse Rate
Unit of Measure: %

Year	Unit Nonresponse Rate (%)
2004	20.0
2005	15.0
2006	14.5
2007	14.8
2008	15.0
2009	15.2
2010	13.5
2011	17.0
2012	20.5
2013	22.5
2014	28.5
2015	24.0
2016	26.5
2017	26.0
2018	25.0
2019	21.5
2020	29.5

Webinfo Privacy policy Contact us Copyright

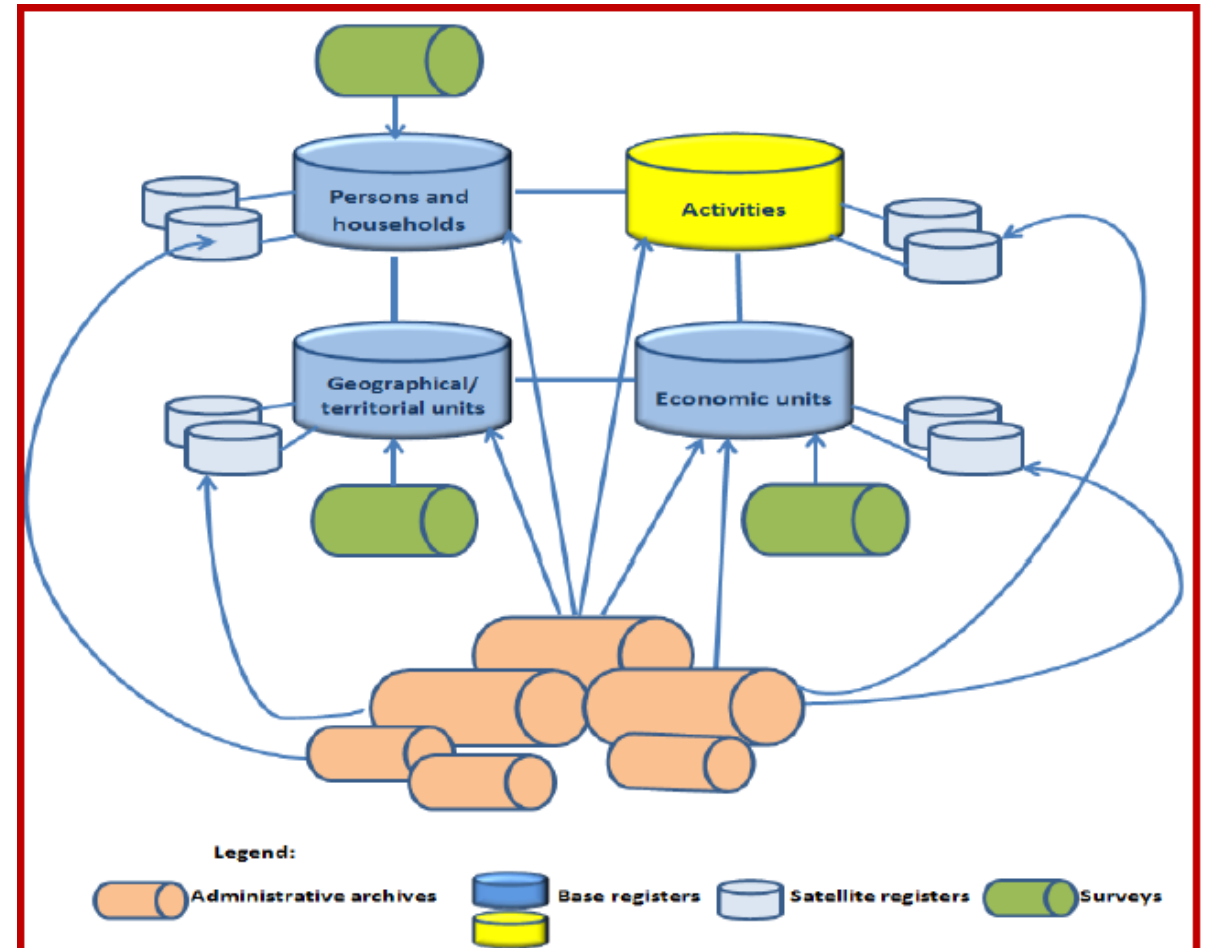
Background: documentation and quality monitoring at Istat

- Istat has a long tradition on quality monitoring, documentation and assessment of statistical processes.
- SIQual system, since 2001, allows the documentation of statistical processes adopting a model that is mappable with GSBPM, collects and stores time series of quality indicators on different phases of the statistical process. The system is well tailored for «traditional» processes, e.g. surveys
- Monitoring systems are integrated to statistical processes, e.g. the web portal for business surveys include several monitoring functionalities
- QRCA is the documentation system of administrative data acquired by Istat and used as input of the statistical production processes. It includes quality indicators on datasets, organized according to three hyperdimensions of Source, Data and Metadata.



Background: the Integrated System of Statistical Registers

- 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 issue: ISSR quality monitoring, documentation and assessment

- When statistical registers of ISSR started to be deployed, the issue of the *systematic, standardised and comparable* documentation and quality assessment of their multisource complex processes arose.
- SIQual system was/is not flexible enough to well describe such processes.
- A first Istat Working Group was set up in 2019 with the task of **defining a system for the documentation of the production processes of the ISSR registers that allows the monitoring and evaluation of the quality**. It developed a first proposal for the documentation model and a first set of quality indicators.
- A second Istat Working Group was set up in 2020 with the task of **designing and implementing the system of quality indicators for monitoring and evaluating the products and processes of the ISSR registers, also considering the feeding the new metadata documentation and process monitoring system of the Institute**.
- In practice, the second group should start from the metadata model and quality indicators defined by the first group and:
 - integrating them taking into account the quality among registers and the quality of registers outputs,
 - test them
 - defining the IT architecture on how to implement them

Some general considerations

- The main objective was to define 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**. Some considerations:
 1. GSBPM was immediately considered as the reference model
 2. The «Design» phase was not in scope, the identification of unit and pseudonymisation are included in data collection
 3. Even if the focus was on the definition of quality indicators, Istat approach has always been to accompany quality indicators with the metadata needed to correctly interpret them.
 4. In the same time, Istat was participating to the UNECE Linking GSIM-GSBPM task team, and the metadata model was largely inspired to the task team work
 5. Quality indicators mainly on accuracy

ISSR Documentation structure

- 1. General characteristics of a register:** identification information, general description and objectives.
- 2. Process quality:** Documentation and quality of the process of a single register
- 3. Output quality:** Quality of the product of a single register
- 4. Coherence:** measures of quality of the ISSR

General characteristics of a statistical register of the ISSR

General information	
Identification information	Name
	Acronym
	Code in the National Statistical Programme
	Responsible
	Structure
	Type (Base/Extended/Thematic)
	First year of release
	First reference year
	Type of temporal reference [punctual/interval]
	Frequency of update
	Frequency of release
	European regulations
Main Objectives	Description
	Target population
	Main target variables
Data Sources	For each source:
	Name
	Provider [Istat/Name of provider]
	Source type [Administrative data, Survey data, Other statistical register]
	Frquency of delivery
	Acquisition mode
State of data source [preliminary, final]	

General characteristics of a statistical register of the ISSR

General information		Example
Identification information	Name	Base register of individuals and households
	Acronym	RBI
	Code in the National Statistical Programme	IST-02721
	Responsible	Sabrina Prati
	Structure	DIPS-DCDC-DCA
	Type (Base/Extended/Thematic)	Base
	First year of release	2018
	First reference year	2015
	Type of temporal reference [punctual/interval]	Punctual (01/01/XXXX)
	Frequency of update	Annual
	Frequency of release	Annual (preliminary version in June T, final in January T+1)
European regulations	EU Reg.No. 1260/2012, DPCM n. 179/2012	
Main Objectives	Description	«The main objective of RBI is...»
	Target population	«Population with signs of presence in Italy...»
	Main target variables	Sex, civil status, date of birth, education level...
Data Sources	For each source:	
	Name	...
	Provider [Istat/Name of provider]	
	Source type [Administrative data, Survey data, Other statistical register]	
	Frequency of delivery	
	Acquisition mode	
State of data source [preliminary, final]		

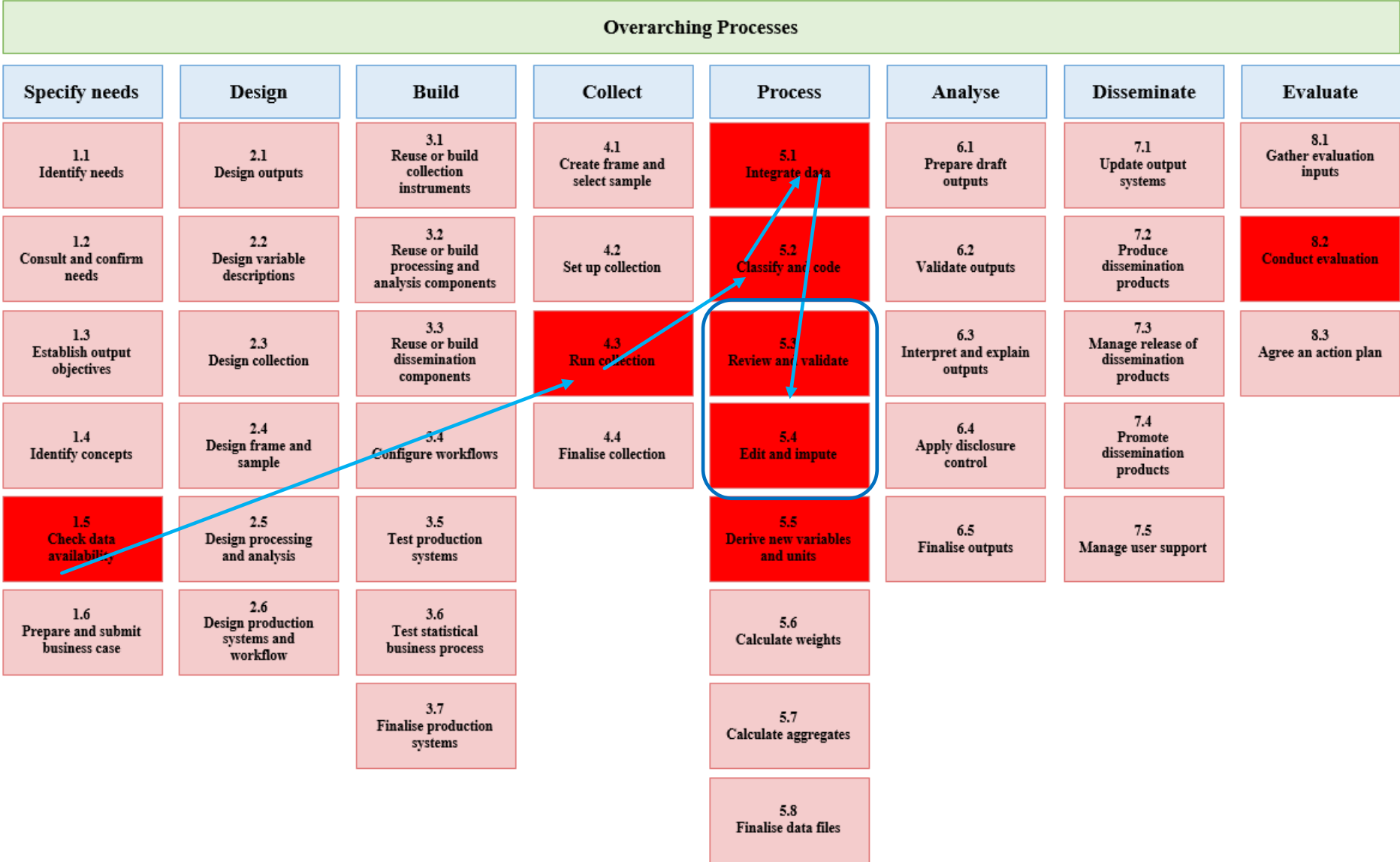
Process quality: Definition of the main GSBPM subprocesses to be considered

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			

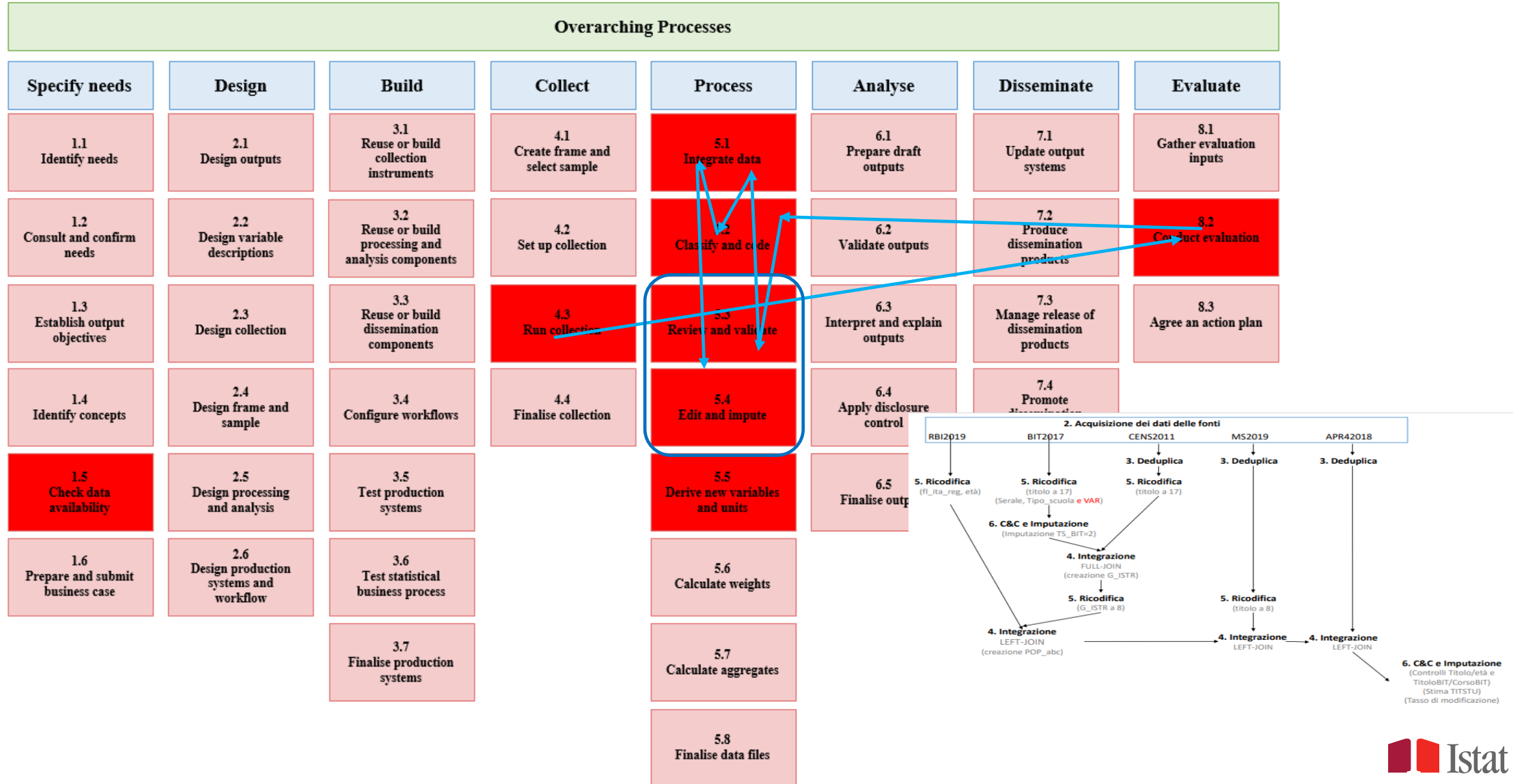
Definition of the main GSBPM subprocesses to be considered

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			

The workflow for RBI – variable sex



The workflow for RBI – variable education level (simplified)



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)
	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-set 1, Data-set 2, ... (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
	<i>Quality control actions</i>	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
	<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

Application to RBI – variable education level last integration step

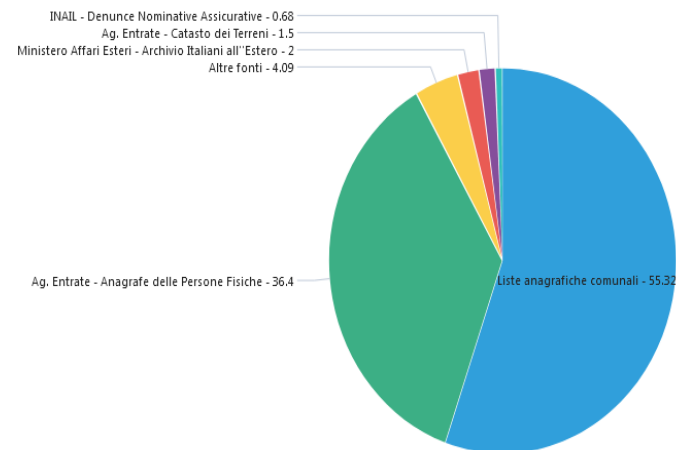
Macro Item	GSIM Object	Values
Input	Transformable input	Dataset RBI2019 (AGE>=9 e residente=1), dataset output step 6, dataset APR4, Master sample census
	Parameter	CODICE_INDIVIDUO
	Process support input	-
GSBPM suprocess	Business Function	Increasing variables (add education level to RBI)
	Business process (GSBPM phase)	5. Process
	Process step (GSBPM sub-process)	5.1. Integrate data
	<i>Quality control actions</i>	-
	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 CENSI I and not in BIT, pop_abc=C if individual is not in BIT and not in CENSI I
	<i>Software</i>	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: Hierarchical 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.

Application to integration step of variable sex, indicator 4.5



Output quality and coherence

Definition of the output for a statistical register: macrodata or microdata

Literary review

Two main proposal and related experimental applications:

1. Estimation of measurement error in case it is available the information from a sample
2. Bootstrap method for the estimation of the variance of the estimates that takes into account different error sources.

Review on keys to link different registers and relationships among registers of ISSR

Proposals and experimental application on quality indicators on:

1. coherence between the same variable in different registers
2. coherence among variables functionally connected in different registers

Additional 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			

Architecture and development

- The Working group activities are not limited to the design and test of the metadata model and quality indicators.
- A detailed analysis of has been done on possible sources of metadata and quality information, such as already existing systems. As expected, QRCA resulted a possible source for several items, but it was also noted that some additional quality indicators calculated on statistical registers could be of interest for QRCA, e.g.
 - 4.5bis = Target unit missing in the administrative data /unit in the (base) register
 - 4.5ter = Non target unit in the administrative dataset / unit in the administrative dataset
- A reflection has been done also on the architecture for the system that should collect and store metadata and quality indicators: the best solution seemed to be to calculate quality indicators directly on statistical registers management systems for monitoring purposes, while to wait for the new metadata system that Istat is starting to design (METAstat) for the metadata and form the collection of quality indicators for documentation and assessment purposes
- Some of the quality indicator have also been implemented in the RBI monitoring system, mapping the IT steps with the GSBPM sub-processes

Concluding remarks on the use of GSBPM and GSIM

- The possibility to refer to standard models simplified the development of the metadata model and allowed to concentrate resources on quality indicators
- The application of the model was not always straightforward:
 - the development of the overarching quality management process of GSBPM could help
 - the reference to GSIM object is challenging in the dialogue with methodologists as well as with thematic expert: to apply the model to actual register different expertises were needed
 - .register process expert
 - .GSIM-metadata expert
 - .IT expert

Aknowledgements

- The results presented are the product of 2 subsequent Istat Working groups that run in the period 2019-2020 and 2020-2021 and involved more than 40 people from different Directorates.
- The coordinators of the first group were Fabiana Rocci and Giovanna Brancato.
- The coordinators of the second group were Fabiana Rocci and Giorgia Simeoni.
- The coordinators of the Work packages of the second group are Giovanna Brancato and Marco Di Zio (WP1: Methodologies for measuring and monitoring quality); Francesca Brait, Romina Filippini and Roberto Sanzo (WP2: Experimental applications); Michele Riccio and Giulia Vaste (WP3: Architecture and development)

Thank you for your
attention!

Giorgia Simeoni | simeoni@istat.it