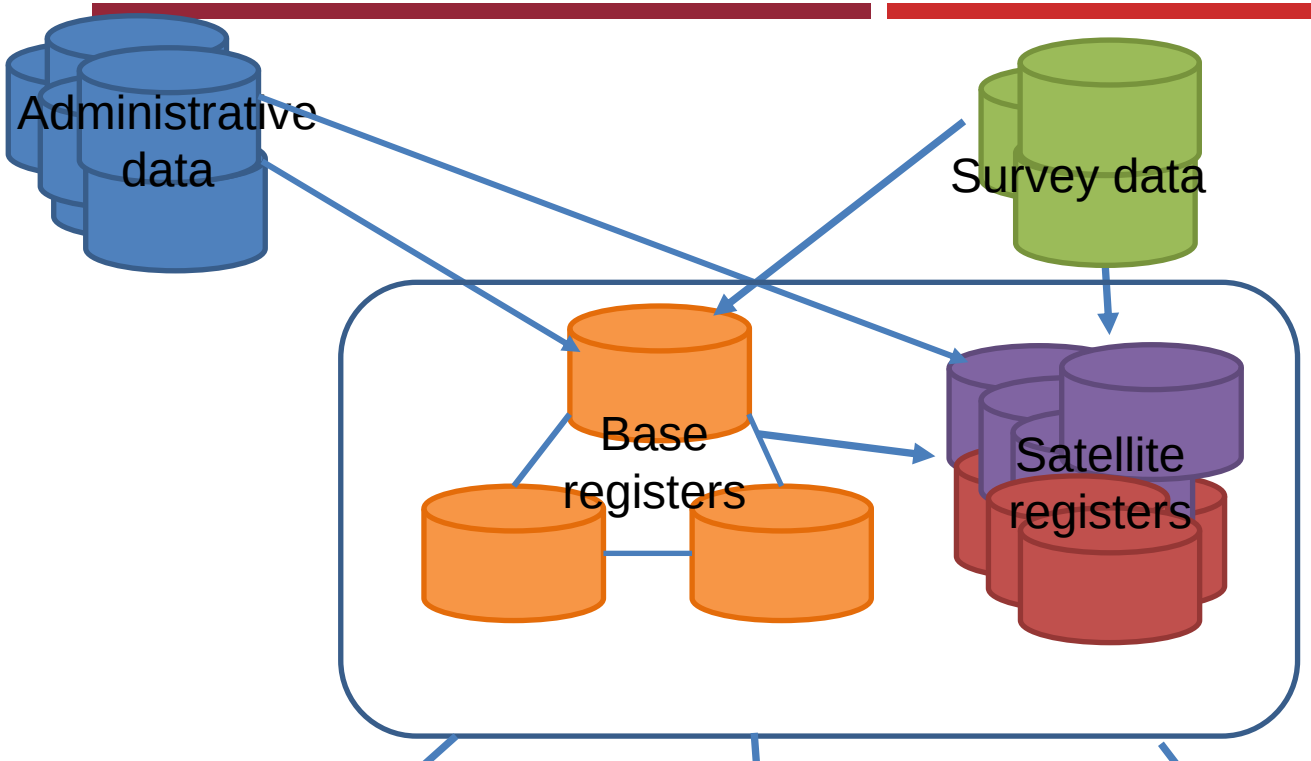
14 November 2023

Global seminar on quality framework for administrative and other data

Istat framework for monitoring, documenting and assessing the quality of the Integrated System of Statistical Registers

Presenter: Giorgia Simeoni, Istat, Italy

The Istat Integrated System of Statistical Registers (ISSR)



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 ISSR changed the paradigm of statistical production at Istat and is designed to cover and support a large part of the production of official statistics in a structured way.

The new quality framework QSIR

Starting from 2019, internal working groups, involving quality and metadata experts, methodologists, thematic and IT experts working on statistical registers developed, tested and finalized the new quality framework, internally called QSIR.

The framework includes

1. **Metadata to identify and describe the objectives and contents of each register**

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] Frequency 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	*Name Surname*
	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]	

The new quality framework QSIR

Starting from 2019, internal working groups, involving quality and metadata experts, methodologists, thematic and IT experts working on statistical registers developed, tested and finalized the new quality framework, internally called QSIR.

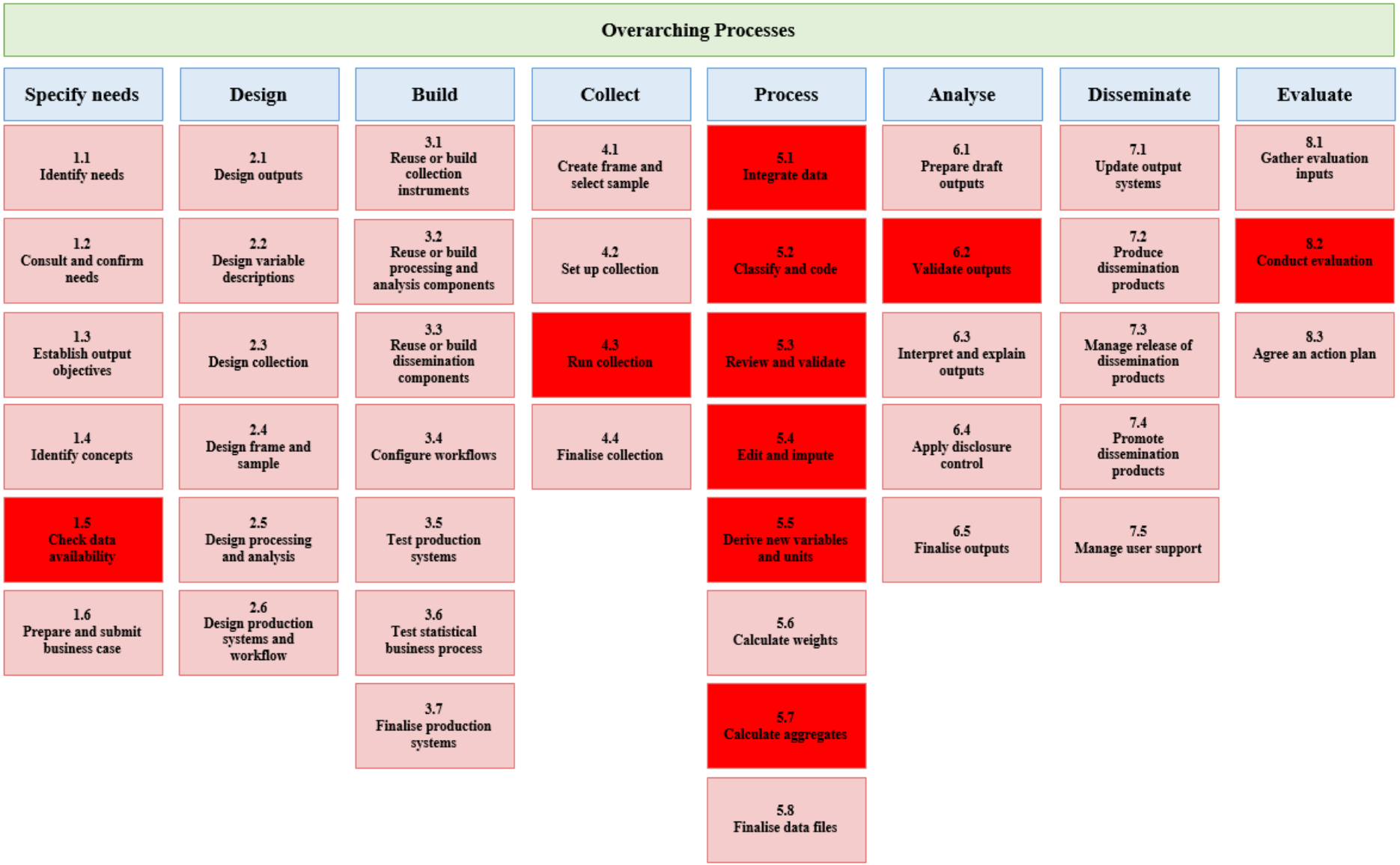
The framework includes

1. Metadata to identify and describe the objectives and contents of each register
2. **Identification of the most relevant GSBPM subprocesses of the ISSR processes, in order to define then standard quality indicators for each of it**

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



The new quality framework QSIR

Starting from 2019, internal working groups, involving quality and metadata experts, methodologists, thematic and IT experts working on statistical registers developed, tested and finalized the new quality framework, internally called QSIR.

The framework includes

1. Metadata to identify and describe the objectives and contents of each register
2. Identification of the most relevant GSBPM subprocesses of the ISSR processes, in order to define then standard quality indicators for each of it
3. **Definition of the generic set of metadata elements to be specified for each sub-process, according to GSIM**

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)
	Process Method
	Rule
	Software
Output	Transformed output
	Process Metric (Quality indicators)
	Process Execution Log

Metadata model from UNECE(2019) Linking GSBPM and GSIM

The new quality framework QSIR

Starting from 2019, internal working groups, involving quality and metadata experts, methodologists, thematic and IT experts working on statistical registers developed, tested and finalized the new quality framework, internally called QSIR.

The framework includes

1. metadata to identify and describe the objectives and contents of each register
2. Identification of the most relevant GSBPM subprocesses of the ISSR processes, in order to define then standard quality indicators for each of it
3. Definition of the generic set of metadata elements to be specified for each sub-process, according to GSIM
4. **Identify of the set of quality indicators specifically to serve the monitoring and the assessment of each sub-process and the metadata needed to calculate and correctly interpret them**

Model for «Integrate data» sub-process

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
	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, transformation 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
4.6 Under-coverage of administrative dataset
4.7 Over-coverage of administrative dataset
Indicators on variables
4.8 Percentage of variables from different input datasets on total number of variables in the integrated dataset
4.9 Distances between variable distributions on the integrated dataset and on the input datasets
4.10 Number of variables derived at the end of integration
4.11 Incoherence in the information present in the different sources on linked records

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
	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.

Next steps

The QSIR framework is currently being applied in 4 different statistical registers of the ISSR:

- ✓ Base register of individuals
- ✓ Thematic register of education
- ✓ Extended register of public administrations
- ✓ Thematic register of labour

The application is demanding, time and specific expertise are needed, but it is seen as an investment by the Statistical register managers

Quality indicators will be implemented (as dashboards) in each register monitoring system, while metadata will be collected and stored in the new metadata system Istat is designing METAstat

Thank you for your attention!

Giorgia Simeoni | simeoni@istat.it

Istat | Istituto Nazionale di Statistica