Implementing the quality framework for the Istat Integrated System of Statistical Registers: challenges and solutions

Cecilia Casagrande, Sara Giavante, Fabiana Rocci and Giorgia Simeoni
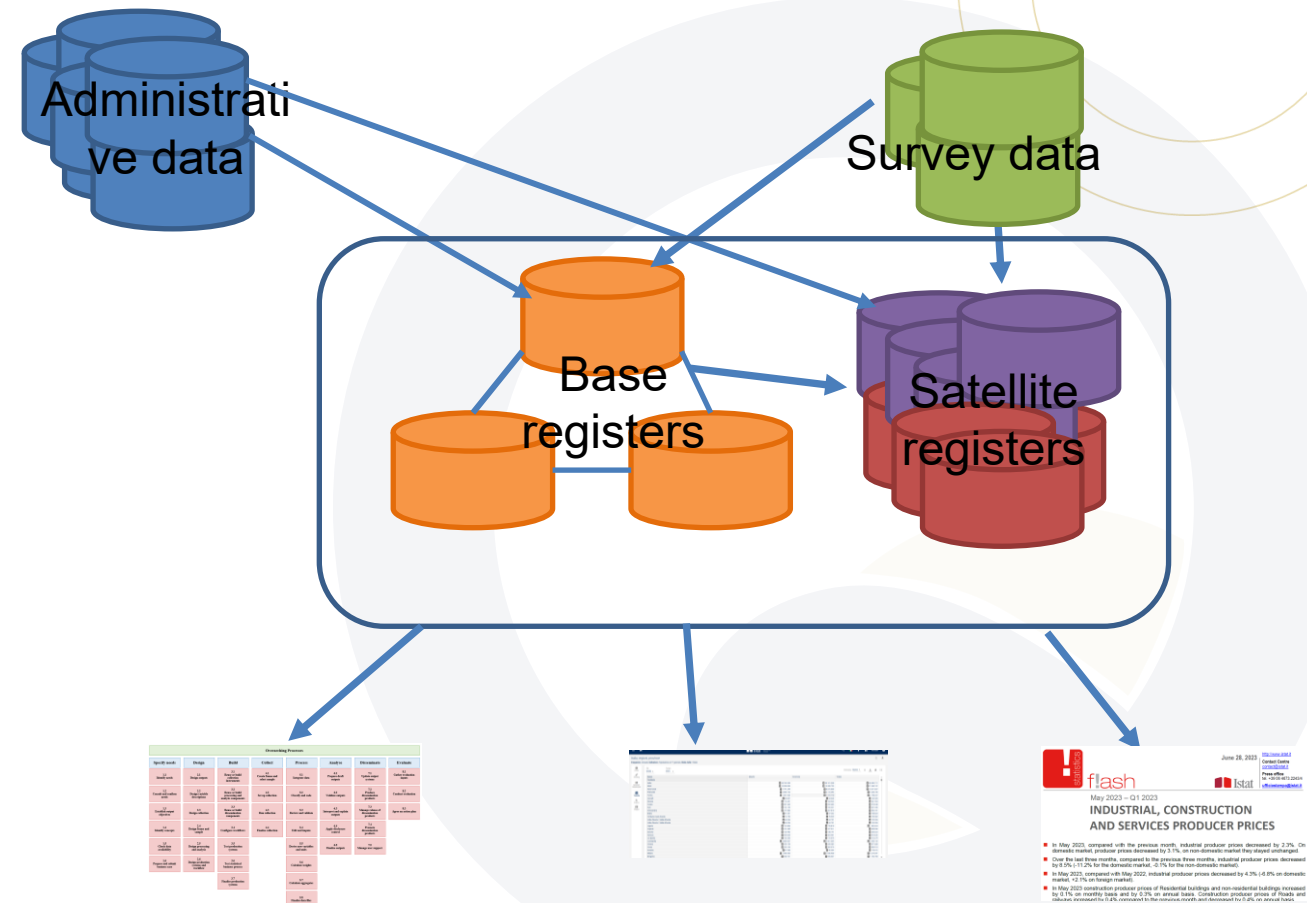
Istat, Italy

Presenter: Giorgia Simeoni

Istat Integrated System of Statistical Registers

- **Cornerstone of Istat modernisation programme started in 2016**

- **Changed the paradigm of statistical production**

- **Base/extended/thematic statistical registers**

- **Complex multisource statistical processes**

The QSIR quality framework

Development

- **Why?**

Need to develop new harmonised framework to monitor, document and evaluate the quality of Statistical Registers of the ISSR

- **Who?**

Interdipartimental Working groups with different expertise (thematic, quality, metadata, methodology, IT)

- **When?**

From 2019 to 2022

- **What?**

Design of the framework, test on 2 statistical registers, finalisation

The QSIR quality framework

- **Identification of relevant sub-processes**

QSIR Sub-processes	GSBPM corresponding Sub-processes
Check data availability	1.4 Check data availability
Acquire data	4.3 Run Collection
Conduct preliminary evaluation	8.2 Conduct evaluation
Integrate data	5.1 Integrate data
Classify and code	5.2 Classify and code
Edit and impute	5.3 Review and validate; 5.4 Edit and impute
Derive new variables and units	5.4 Derive new variables and units
Calculate aggregates	5.5 Calculate weights; 5.6 Calculate aggregates
Validate outputs	6.2 Validate outputs

The QSIR quality framework

- **Identification of relevant sub-processes**

QSIR Sub-processes	GSBPM corresponding Sub-processes
Check data availability	1.4 Check data availability
Acquire data	4.3 Run Collection
Conduct preliminary evaluation	8.?
Integrate data	5
Classify and code	
Edit and impute	
Derive new variables and units	
Calculate aggregates	
Validate outputs	6.2 Validate outputs

Rationale:
- Focus on current data processing phases not on the design phases
- Monitoring of data acquisition and pseudonymisation out of scope

The QSIR quality framework

For each sub-process:

- Textual description of what is included

- Metadata template

- Standard set of quality indicators

The QSIR quality framework

Textual description of what is included: example

Conduct preliminary evaluation

In this sub-process **preliminary checks and evaluations** are carried out on a dataset, which can be coming from **individual sources** or be an **integrated dataset**. **Deduplication** is performed and **not usable records** are identified and deleted. **Missing values** are checked. If an appropriate auxiliary source is available (benchmark), the **coverage** of the dataset under consideration can be estimated, both on the basis of aggregate comparisons and through micro-level matching. The coverage estimate is accompanied by the evaluation of the representativeness of the dataset with respect not only to the number but also to the characteristics of the units contained in it, for example: presence or absence of large companies, of universities with unique or rare degree courses.

The indicators calculated in this phase on the individual sources may be a useful feedback for the unit identification phase, in particular those relating to deduplication, given that these errors may be specific to the source or due to the pseudonym attribution process.

The QSIR quality framework

Metadata template: *example*

Conduct preliminary evaluation

Macro Item	GSIM Object	Possible values
Input	Core input	Data-set to be evaluated (data structure: units and variables): it can be one of the source dataset or the integrated data set
	Parameter input	Key linkage variables
	Process support input	Reference/benchmark data-set Definition of the population of the data-set to be evaluated
Sub-process	Business Function	Deduplicating data-set, checking for missing values, evaluating the coverage of the dataset
	Business Process (GSBPM phase)	8. Evaluate
	Process Step (GSBPM sub-process)	8.2 Conduct evaluation
	Process Method	Duplications identification Identification of missing or not usable data Coverage evaluation through aggregate comparisons Coverage evaluation through microdata matching
	Rule	Rules to identify the deduplication Rules to identify missing values Rules to identify not usable data Integration model, relationship 1-1, n-1, n-n
	Software Agent	Relais, Statmatch, Ad hoc procedures
Output	Core output	Data set without duplicates or unusable data
	Process Metric (Quality indicators)	See next slide
	Process Execution Log	Processing time

The QSIR quality framework

Standard set of quality indicators: example

Conduct preliminary evaluation

Deduplication indicators
3.1. Percentage of duplicates records
3.2. Percentage of duplicates records on the key variable
3.3. Percentage of duplicates records on a set of relevant variables
3.4 Discrepancies between information present in duplicate records
Missing values indicators
3.5 Missing value rate for the main variables
3.6 Percentage of not usable records
Coverage and representativeness indicators
3.7 Coverage rate of the evaluated dataset with respect to the benchmark dataset
3.8 Comparison between statistics (average, totals, …) and distribution of variables between the evaluated dataset and the related sub-population in the benchmark dataset
3.9 Comparison between statistics (average, totals, …) and distribution of variables between the evaluated dataset and the total population in the benchmark dataset

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

INSTITUTO NACIONAL DE ESTATÍSTICA
Statistics Portugal

eurostat

The conference is partly financed by the European Union

The QSIR quality framework

Implementation

The QSIR framework is being applied in 4 different statistical registers through different working groups. Applying the QSIR framework in a statistical register is demanding!!!

| describing the process through the metadata templates | defining the workflow that links the sub-processes | identifying which of quality indicators proposed is applicable | test quality indicators (QI) | provide input for the IT sector | automatise the QI in register monitoring system | implement metadata and QI in new metadata system METAstat |

Implementation issues

A group of QSIR expert coordinating the implementation discuss challenges and identify common solutions that sometimes lead to an improvement of the framework.

Examples:

- *Check data availability* metadata template

- Adaptive implementation strategy

Implementation issues

Check data availability *metadata template*

- The template was intended to be applied only in case of **lack of one source**, the **availability of a potential new source**, a **change within a source**.

- In practice, it was experienced it could be used also to deliver a **preliminary structured and exhaustive mapping of each source**, that could be very useful to eventually discover other problems than the ones listed above.

- An additional **quality indicator on presence/absence of the data source** is also proposed, together with the evaluation of what is the impact in the case of absence.
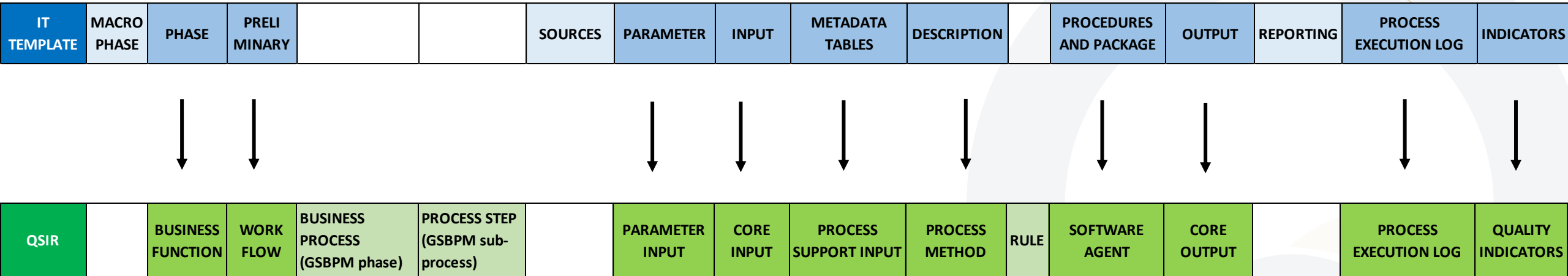
Implementation issues

Adaptive implementation strategy

- Different registers, different stages of implementation, harmonized but not identical technical implementation approaches.

- In some registers a template to document the process step from an IT point of view is already used.

- Even if the IT process step could have a meaning and a granularity different from QSIR sub-processes, some information are common (e.g. input, output, parameters, etc.).

- To facilitate implementation, QSIR template has been integrated with the existing one, asking to provide only missing information.

Implementation issues

IT TEMPLATE	MACRO PHASE	PHASE	PRELI MINARY			SOURCES	PARAMETER	INPUT	METADATA TABLES	DESCRIPTION		PROCEDURES AND PACKAGE	OUTPUT	REPORTING	PROCESS EXECUTION LOG	INDICATORS

QSIR		BUSINESS FUNCTION	WORK FLOW	BUSINESS PROCESS (GSBPM phase)	PROCESS STEP (GSBPM sub-process)		PARAMETER INPUT	CORE INPUT	PROCESS SUPPORT INPUT	PROCESS METHOD	RULE	SOFTWARE AGENT	CORE OUTPUT		PROCESS EXECUTION LOG	QUALITY INDICATORS

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

INSTITUTO NACIONAL DE ESTATÍSTICA
Statistics Portugal

eurostat

The conference is partly financed by the European Union

Concluding remarks

- Developing monitoring systems for multisource processes is complex, QSIR framework aims at facilitating the task, and contemporarily defines structured and standard solutions for documentation and ex-post quality evaluation.

- Beside the framework proposed, we wanted to share the collaborative approach that we are following for coordinating the implementation, highlighting the importance of not considering the work «closed» after its theorical definition, since further improvements are always possible.