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.
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.
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.
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

<table>
<thead>
<tr>
<th>Identification information</th>
<th>Name</th>
<th>Acronym</th>
<th>Code in the National Statistical Programme</th>
<th>Responsible</th>
<th>Structure</th>
<th>Type (Base/Extended/Thematic)</th>
<th>First year of release</th>
<th>First reference year</th>
<th>Type of temporal reference [punctual/interval]</th>
<th>Frequency of update</th>
<th>Frequency of release</th>
<th>European regulations</th>
</tr>
</thead>
</table>

## Main Objectives

<table>
<thead>
<tr>
<th>Main Objectives</th>
<th>Description</th>
<th>Target population</th>
<th>Main target variables</th>
</tr>
</thead>
</table>

## Data Sources

<table>
<thead>
<tr>
<th>Data Sources</th>
<th>For each source:</th>
<th>Name</th>
<th>Provider [Istat/Name of provider]</th>
<th>Source type [Administrative data, Survey data, Other statistical register]</th>
<th>Frequency of delivery</th>
<th>Acquisition mode</th>
<th>State of data source [preliminary, final]</th>
</tr>
</thead>
</table>
# General characteristics of a statistical register of the ISSR

<table>
<thead>
<tr>
<th>General information</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Identification information</strong></td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Base register of individuals and households</td>
</tr>
<tr>
<td>Acronym</td>
<td>RBI</td>
</tr>
<tr>
<td>Code in the National Statistical Programme</td>
<td>IST-02721</td>
</tr>
<tr>
<td>Responsible</td>
<td>Sabrina Prati</td>
</tr>
<tr>
<td>Structure</td>
<td>DIPS-DCDC-DCA</td>
</tr>
<tr>
<td>Type (Base/Extended/Thematic)</td>
<td>Base</td>
</tr>
<tr>
<td>First year of release</td>
<td>2018</td>
</tr>
<tr>
<td>First reference year</td>
<td>2015</td>
</tr>
<tr>
<td>Type of temporal reference [punctual/interval]</td>
<td>Punctual (01/01/XXXX)</td>
</tr>
<tr>
<td>Frequency of update</td>
<td>Annual</td>
</tr>
<tr>
<td>Frequency of release</td>
<td>Annual (preliminary version in June T, final in January T+1)</td>
</tr>
<tr>
<td>European regulations</td>
<td>EU Reg.No. 1260/2012, DPCM n. 179/2012</td>
</tr>
</tbody>
</table>

| **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

<table>
<thead>
<tr>
<th>Overarching Processes</th>
<th>Specify needs</th>
<th>Design</th>
<th>Build</th>
<th>Collect</th>
<th>Process</th>
<th>Analyse</th>
<th>Disseminate</th>
<th>Evaluate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specify needs</td>
<td>1.1 Identify needs</td>
<td>2.1 Design outputs</td>
<td>3.1 Reuse or build collection instruments</td>
<td>4.1 Create frame and select sample</td>
<td>5.1 Integrate data</td>
<td>6.1 Prepare draft outputs</td>
<td>7.1 Update output systems</td>
<td>8.1 Gather evaluation inputs</td>
</tr>
<tr>
<td></td>
<td>1.2 Consult and confirm needs</td>
<td>2.2 Design variable descriptions</td>
<td>3.2 Reuse or build processing and analysis components</td>
<td>4.2 Set up collection</td>
<td>5.2 Classify and code</td>
<td>6.2 Validate outputs</td>
<td>7.2 Produce dissemination products</td>
<td>8.2 Conduct evaluation</td>
</tr>
<tr>
<td></td>
<td>1.3 Establish output objectives</td>
<td>2.3 Design collection</td>
<td>3.3 Reuse or build dissemination components</td>
<td>4.3 Run collection</td>
<td>5.3 Review and validate</td>
<td>6.3 Interpret and explain outputs</td>
<td>7.3 Manage release of dissemination products</td>
<td>8.3 Agree an action plan</td>
</tr>
<tr>
<td></td>
<td>1.4 Identify concepts</td>
<td>2.4 Design frame and sample</td>
<td>3.4 Configure workflows</td>
<td>4.4 Finalise collection</td>
<td>5.4 Edit and improve</td>
<td>6.4 Apply disclosure control</td>
<td>7.4 Promote dissemination product</td>
<td>7.5 Manage user support</td>
</tr>
<tr>
<td></td>
<td>1.5 Check data availability</td>
<td>2.5 Design processing and analysis</td>
<td>3.5 Test production systems</td>
<td></td>
<td>5.5 Derive new variables and units</td>
<td>6.5 Finalise outputs</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.6 Prepare and submit business case</td>
<td>2.6 Design production systems and workflow</td>
<td>3.6 Test statistical business process</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Definition of the main GSBPM subprocesses to be considered

<table>
<thead>
<tr>
<th>Overarching Processes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specify needs</td>
</tr>
<tr>
<td>Design</td>
</tr>
<tr>
<td>Build</td>
</tr>
<tr>
<td>Collect</td>
</tr>
<tr>
<td>Process</td>
</tr>
<tr>
<td>Analyse</td>
</tr>
<tr>
<td>Disseminate</td>
</tr>
<tr>
<td>Evaluate</td>
</tr>
</tbody>
</table>

### 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

### 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

### 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

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

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

### Analyse
- 6.1 Prepare draft outputs
- 6.2 Validate outputs
- 6.3 Interpret and explain output
- 6.4 Apply dissemination control
- 6.5 Finalise outputs

### 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

### Evaluate
- 8.1 Gather evaluation inputs
- 8.2 Conduct evaluation
- 8.3 Agree an action plan
### The workflow for RBI – variable sex

<table>
<thead>
<tr>
<th>Overarching Processes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specify needs</td>
</tr>
<tr>
<td>Design</td>
</tr>
<tr>
<td>Build</td>
</tr>
<tr>
<td>Collect</td>
</tr>
<tr>
<td>Process</td>
</tr>
<tr>
<td>Analyse</td>
</tr>
<tr>
<td>Disseminate</td>
</tr>
<tr>
<td>Evaluate</td>
</tr>
</tbody>
</table>

#### 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

#### 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

#### 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

#### Collect
- 4.1 Create frame and select sample
- 4.2 Set up collection
- 4.3 Run collection
- 4.4 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

#### 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

#### 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

#### Evaluate
- 8.1 Gather evaluation inputs
- 8.2 Conduct evaluation
- 8.3 Agree an action plan
The workflow for RBI – variable education level (simplified)

<table>
<thead>
<tr>
<th>Overarching Processes</th>
<th>Specify needs</th>
<th>Design</th>
<th>Build</th>
<th>Collect</th>
<th>Process</th>
<th>Analyse</th>
<th>Disseminate</th>
<th>Evaluate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Identify needs</td>
<td>2.1 Design outputs</td>
<td>3.1 Design build collection instruments</td>
<td>4.1 Create frame and select sample</td>
<td>5.1 Integrate data</td>
<td>6.1 Prepare draft outputs</td>
<td>7.1 Update output systems</td>
<td>8.1 Gather evaluation inputs</td>
<td></td>
</tr>
<tr>
<td>1.2 Consult and confirm needs</td>
<td>2.2 Design variable descriptions</td>
<td>3.2 Design build processing and analysis components</td>
<td>4.2 Set up collection</td>
<td>5.2 Classify and code</td>
<td>6.2 Validate outputs</td>
<td>7.2 Produce dissemination products</td>
<td>8.2 Agree on an action plan</td>
<td></td>
</tr>
<tr>
<td>1.3 Establish output objectives</td>
<td>2.3 Design collection</td>
<td>3.3 Design build dissemination components</td>
<td>4.3 Run collection</td>
<td>5.3 Review and validate</td>
<td>6.3 Interpret and explain outputs</td>
<td>7.3 Manage release of dissemination products</td>
<td>8.3 Agree on an action plan</td>
<td></td>
</tr>
<tr>
<td>1.4 Identify concepts</td>
<td>2.4 Design frame and sample</td>
<td>3.4 Design build dissemination components</td>
<td>4.4 Finalise collection</td>
<td>5.4 Edit and update</td>
<td>6.4 Apply dissemination control</td>
<td>7.4 Promote use</td>
<td>8.4 Agree on an action plan</td>
<td></td>
</tr>
<tr>
<td>1.5 Check data availability</td>
<td>2.5 Design processing and analysis</td>
<td>3.5 Design build dissemination components</td>
<td>4.5 Finalise collection</td>
<td>5.5 Derive new variables and units</td>
<td>6.5 Finalise output</td>
<td>7.5 Promote use</td>
<td>8.5 Agree on an action plan</td>
<td></td>
</tr>
<tr>
<td>1.6 Prepare and submit business case</td>
<td>2.6 Design production systems and workflow</td>
<td>3.6 Design build dissemination components</td>
<td>4.6 Finalise collection</td>
<td>5.6 Calculate weights</td>
<td>6.6 Finalise output</td>
<td>7.6 Promote use</td>
<td>8.6 Agree on an action plan</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.7 Design build dissemination components</td>
<td>4.7 Finalise collection</td>
<td>5.7 Calculate aggregates</td>
<td>6.7 Finalise output</td>
<td>7.7 Promote use</td>
<td>8.7 Agree on an action plan</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.8 Finalise production systems</td>
<td>4.8 Finalise collection</td>
<td>5.8 Finalise data files</td>
<td>6.8 Finalise output</td>
<td>7.8 Promote use</td>
<td>8.8 Agree on an action plan</td>
<td></td>
</tr>
</tbody>
</table>
Metadata model for each GSBPM sub-process

<table>
<thead>
<tr>
<th>Macro Item</th>
<th>GSIM Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
<td>Transformable input</td>
</tr>
<tr>
<td></td>
<td>Parameter</td>
</tr>
<tr>
<td></td>
<td>Process support input</td>
</tr>
<tr>
<td>GSBPM sub-process</td>
<td>Business Function</td>
</tr>
<tr>
<td></td>
<td>Business process (GSBPM phase)</td>
</tr>
<tr>
<td></td>
<td>Process step (GSBPM sub-process)</td>
</tr>
<tr>
<td></td>
<td>Quality control actions</td>
</tr>
<tr>
<td></td>
<td>Process Method</td>
</tr>
<tr>
<td></td>
<td>Rule</td>
</tr>
<tr>
<td></td>
<td>Software</td>
</tr>
<tr>
<td>Output</td>
<td>Transformed output</td>
</tr>
<tr>
<td></td>
<td>Process Metric (Quality indicators)</td>
</tr>
<tr>
<td></td>
<td>Process Execution Log</td>
</tr>
</tbody>
</table>

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 sub-process 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

<table>
<thead>
<tr>
<th>Macro Item</th>
<th>GSIM Object</th>
<th>Possible values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input</strong></td>
<td><strong>Transformable input</strong></td>
<td>Data-set1, Data-set2, … (data structure: units and variables)</td>
</tr>
<tr>
<td></td>
<td><strong>Parameter</strong></td>
<td>Threshold, Linkage keys, blocking variables</td>
</tr>
<tr>
<td></td>
<td><strong>Process support input</strong></td>
<td>Further variables useful for identification other than the keys or to control the matching</td>
</tr>
<tr>
<td><strong>GSBPM</strong></td>
<td><strong>Business Function</strong></td>
<td>Increasing units, increasing variables, increasing both</td>
</tr>
<tr>
<td><strong>suprocess</strong></td>
<td><strong>Business process</strong> (GSBPM phase)</td>
<td>5. Process</td>
</tr>
<tr>
<td></td>
<td><strong>Process step</strong> (GSBPM sub-process)</td>
<td>5.1. Integrate data</td>
</tr>
<tr>
<td></td>
<td><strong>Quality control actions</strong></td>
<td>Actions for preventing, monitoring, reducing errors due to integration</td>
</tr>
<tr>
<td></td>
<td><strong>Process Method</strong></td>
<td>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</td>
</tr>
<tr>
<td></td>
<td><strong>Rule</strong></td>
<td>Integration model, Rules for the hierarchical selection of the sources, transformation rules</td>
</tr>
<tr>
<td></td>
<td><strong>Software</strong></td>
<td>Relais, Statmatch, Ad hoc procedures</td>
</tr>
<tr>
<td><strong>Output</strong></td>
<td><strong>Transformed output</strong></td>
<td>Integrated Data set, Non linked records data sets</td>
</tr>
<tr>
<td></td>
<td><strong>Process Metric</strong> (Quality indicators)</td>
<td>SEE NEXT SLIDE</td>
</tr>
<tr>
<td></td>
<td><strong>Process Execution Log</strong></td>
<td>Integration time</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Macro Item</th>
<th>GSIM Object</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input</strong></td>
<td>Transformable input</td>
<td>Dataset RBI2019 (AGE&gt;=9 e residente=1), dataset output step 6, dataset APR4, Master sample census</td>
</tr>
<tr>
<td></td>
<td>Parameter</td>
<td>CODICE_INDIVIDUO</td>
</tr>
<tr>
<td></td>
<td>Process support input</td>
<td>-</td>
</tr>
<tr>
<td><strong>GSBPM suprocess</strong></td>
<td>Business Function</td>
<td>Increasing variables (add education level to RBI)</td>
</tr>
<tr>
<td></td>
<td>Business process (GSBPM phase)</td>
<td>5. Process</td>
</tr>
<tr>
<td></td>
<td>Process step (GSBPM sub-process)</td>
<td>5.1. Integrate data</td>
</tr>
<tr>
<td></td>
<td>Quality control actions</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Process Method</td>
<td>Deterministic Record linkage</td>
</tr>
<tr>
<td></td>
<td>Rule</td>
<td>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</td>
</tr>
<tr>
<td></td>
<td>Software</td>
<td>Oracle procedure</td>
</tr>
<tr>
<td><strong>Output</strong></td>
<td>Transformed output</td>
<td>Integrated Data set with all RBI units and with variables G_ISTR, tit_stu, pop_abc</td>
</tr>
<tr>
<td></td>
<td>Process Metric (Quality indicators)</td>
<td>SEE NEXT SLIDE</td>
</tr>
<tr>
<td></td>
<td>Process Execution Log</td>
<td>-</td>
</tr>
</tbody>
</table>
## Application to integration step of variable education level

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

## Application to integration step of variable sex, indicator 4.5

![Diagram showing the application to integration step of variable sex]
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
- 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

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

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

Collect
- 4.1 Create frame and select sample
- 4.2 Set up collection
- 4.3 Run collection
- 4.4 Finalise collection
- 4.5 Derive new variables and units
- 4.6 Calculate weights
- 4.7 Calculate aggregates

Process
- 5.1 Integrate data
- 5.2 Classify and code
- 5.3 Review and validate
- 5.4 Edit and impute
- 5.5 Validate outputs
- 5.6 Finalise outputs

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

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

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