

ModernStats World Workshop 2019

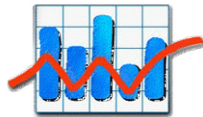
Geneva, Switzerland 26-28 June 2019

# Challenges in re-designing the national metadata system according to international standards

Giorgia Simeoni ([simeoni@istat.it](mailto:simeoni@istat.it)), Mauro Scanu ([scanu@istat.it](mailto:scanu@istat.it))

Istat

- Reference metadata and quality information: **SIDI/SIQual**



**ISTAT  
SIDI**

Sistema Informativo di documentazione delle indagini  
Ambiente di gestione

Username:   
Password:

Information system on quality of statistical production processes

Home Glossary Helped Search Multidimensional search Complete list Report Documents

[Versione italiana](#) | [Comments](#) | [FAQ](#)

**SIQual**

Helped Search  
Multidimensional search  
Complete list  
Documents

SIQual information system on quality contains information on the execution of Istat primary surveys and secondary studies and on activities developed to guarantee quality of the produced statistical information. The system is devoted to the navigation of metadata describing the production process and its characteristics: information content; phases and operations of the production process; activities to prevent, monitor and evaluate errors. Statistical information systems for data dissemination and their sources are also documented. In SIQual, primary surveys, secondary studies and statistical information systems are generically referred as statistical processes. Documentation of general interest or specific for a certain process is available in the "Documents" area of the system. SIQual also allows to consult statistical on-line data disseminated on the Web.

Visitor N° 1315247 Best resolution: 1024x768, Internet Explorer 5.X or superiori, Netscape 6.x o superiori. Copyright

- Structural metadata: **SUM-MS**

**SUM** SISTEMA UNITARIO METADATI

Autenticazione Operatore

Username:   
Password:

«OLD»

Deployed in the early 2000s and still currently updated by a net of quality pilots throughout the NSI

- Technologically old
- Not much flexible to document multisource statistics
- Not compliant with more recent international standards

RICH

About 400 statistical processes are documented, about 240 regularly carried out

COMPLEX

Composed by two main subsystems: SIDI for entering the information, SIQual for consulting. Integrated with other several Istat systems.

USEFUL

Includes minor subsystems that re-use the metadata and the quality indicators for different purposes (from statistical yearbook methodological notes to ESS standard quality reports transmitted to Eurostat)

NOT  
NEW

Under development since 2010

- Could be improved technologically

RICH

Contents:

- The definition of micro and macro data-structures produced along any statistical program undertaken by Istat.
- The necessary concepts for micro and macro-data structures definitions.

STANDARD

Metadata are modelled according to GSIM (Concept and Structures Groups)

Aims:

- Centralise the management of structural metadata
- Harmonise metadata between and along statistical programs
- Traceability of each data production process

USEFUL

In progress

# Requirements for the new Istat metadata system

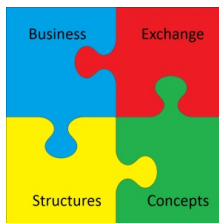
SIDI/SIQual and SUM-MS are currently being redesigned as a **new comprehensive metadata system**.

Requirements:

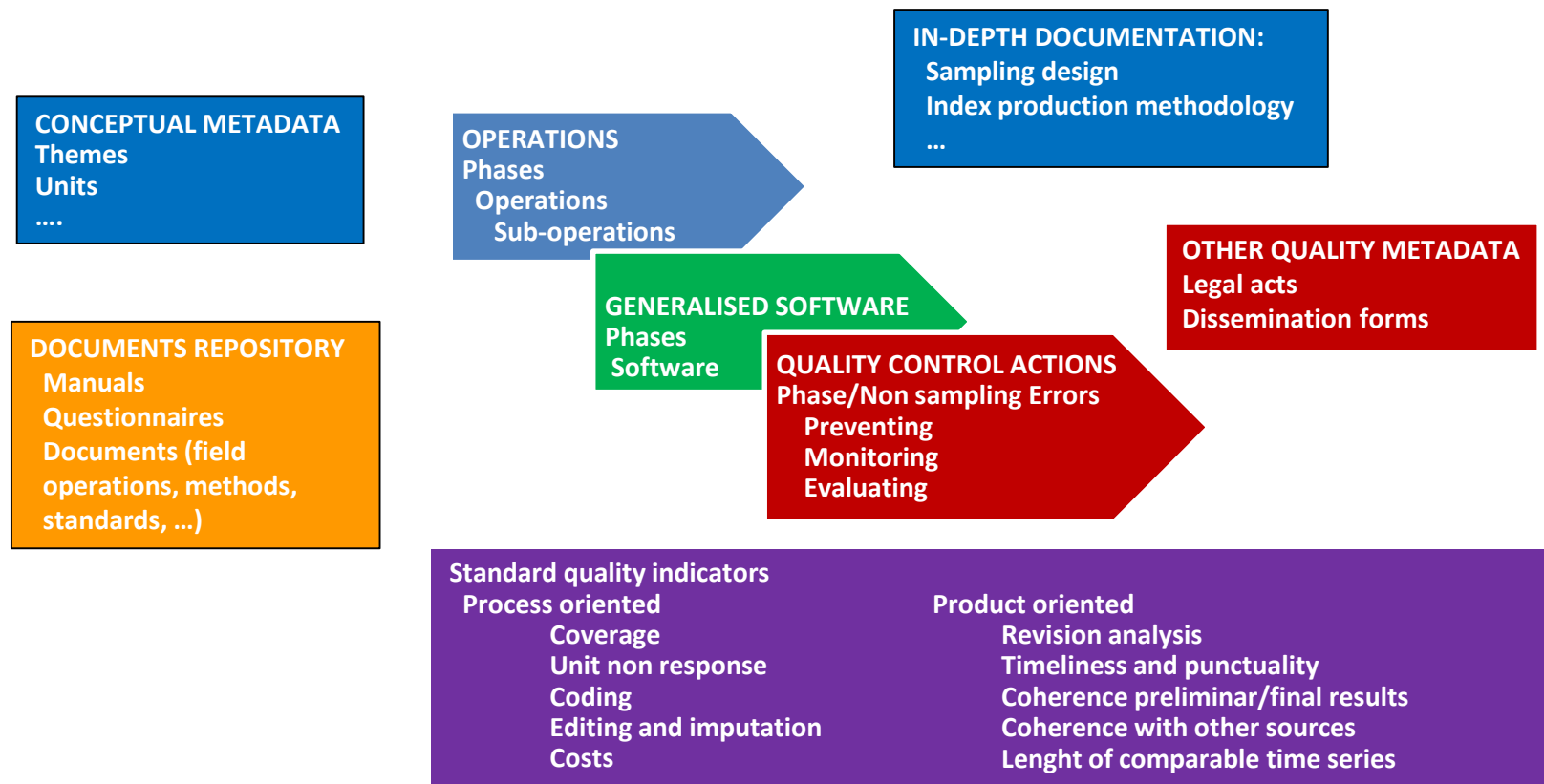
- preserve information already in SIDI/SIQual and SUM-MS
- document, trace (and drive) different kinds of statistical production processes
- manage and store the quality indicators
- ...

State of the art: conceptual modelling

Quality Management / Metadata Management							
Specify Needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Specify needs	1.1 Design outputs	1.1 Build infrastructure	1.1 Collect data	1.1 Process data	1.1 Analyse data	1.1 Disseminate data	1.1 Evaluate data
1.2 Collect & evaluate data	1.2 Design outputs	1.2 Build infrastructure	1.2 Collect data	1.2 Process data	1.2 Analyse data	1.2 Disseminate data	1.2 Evaluate data
1.3 Analyse data	1.3 Design outputs	1.3 Build infrastructure	1.3 Collect data	1.3 Process data	1.3 Analyse data	1.3 Disseminate data	1.3 Evaluate data
1.4 Disseminate data	1.4 Design outputs	1.4 Build infrastructure	1.4 Collect data	1.4 Process data	1.4 Analyse data	1.4 Disseminate data	1.4 Evaluate data
1.5 Evaluate data	1.5 Design outputs	1.5 Build infrastructure	1.5 Collect data	1.5 Process data	1.5 Analyse data	1.5 Disseminate data	1.5 Evaluate data
1.6 Specify needs	1.6 Design outputs	1.6 Build infrastructure	1.6 Collect data	1.6 Process data	1.6 Analyse data	1.6 Disseminate data	1.6 Evaluate data
1.7 Design outputs	1.7 Design outputs	1.7 Build infrastructure	1.7 Collect data	1.7 Process data	1.7 Analyse data	1.7 Disseminate data	1.7 Evaluate data
1.8 Build infrastructure	1.8 Design outputs	1.8 Build infrastructure	1.8 Collect data	1.8 Process data	1.8 Analyse data	1.8 Disseminate data	1.8 Evaluate data
1.9 Collect data	1.9 Design outputs	1.9 Build infrastructure	1.9 Collect data	1.9 Process data	1.9 Analyse data	1.9 Disseminate data	1.9 Evaluate data
1.10 Process data	1.10 Design outputs	1.10 Build infrastructure	1.10 Collect data	1.10 Process data	1.10 Analyse data	1.10 Disseminate data	1.10 Evaluate data
1.11 Analyse data	1.11 Design outputs	1.11 Build infrastructure	1.11 Collect data	1.11 Process data	1.11 Analyse data	1.11 Disseminate data	1.11 Evaluate data
1.12 Disseminate data	1.12 Design outputs	1.12 Build infrastructure	1.12 Collect data	1.12 Process data	1.12 Analyse data	1.12 Disseminate data	1.12 Evaluate data
1.13 Evaluate data	1.13 Design outputs	1.13 Build infrastructure	1.13 Collect data	1.13 Process data	1.13 Analyse data	1.13 Disseminate data	1.13 Evaluate data



# SIDI/SIQual contents



# SIDI/SIQual core contents

**Istat** Information system on quality of statistical production processes

Home Glossary Helped Search Multidimensional search Complete list Report Documents

Versione italiana | Comments | FAQ

## Phases of production process

> home > Complete list > Phases of production process

**Process description**  
**Insights**  
 Summary metadata  
 Legislation  
 Observed phenomena  
 Reporting and observational units  
 Process Design  
 Phases of production process  
 Generalised software  
 Questionnaires  
 Process documents

**Data dissemination**  
 Classifications  
 Publications  
 Online data

**Report**

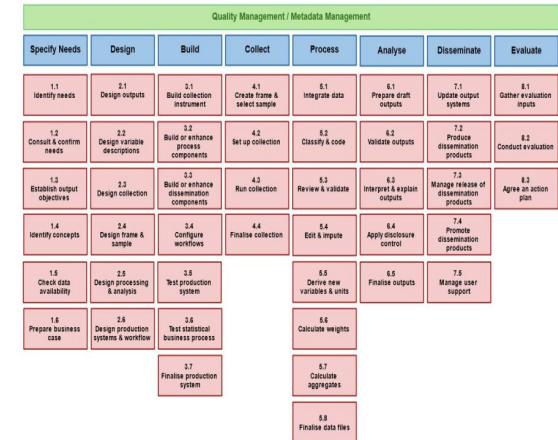
**Phases of the production process** (S) (S)

- Planning
- Frame development
  - on mode
  - processing
  - and imputation
- Design**
  - Data validation
  - Data storage
  - Dissemination
- 2.4 Design frame & sample
- 4.1 Create frame & select sample
- 4.3 Run collection

**Operations, quality control actions and general**  
**Data collection mode**  
 Collapse

- Self-administered data collection via electronic
- Self-administered data collection by e-questionnaire (Computer Assisted Web Interviewing-CAWI) or through upload of datasets on Istat web site INDATA (since 31/12/2015)
  - Control on unit nonresponse
    - Activities for preventing unit nonresponse
      - Survey presentation letter signed by Istat President (since 31/12/2015)
      - Guarantees on statistical confidentiality (since 31/12/2015)
      - Written description of survey objectives (since 31/12/2015)
      - Special care in writing clear instructions to fill-in questionnaires (since 31/12/2015)
      - Establishing a toll free line or telephone number for further explanations (since 31/12/2015)
      - Administrative fines for nonrespondents (since 31/12/2015)
      - Special care in drafting clear instructions for database setting up (since 31/12/2015)
      - Establishing an e-mail address for information on the web procedure (since 31/12/2015)
      - Formal notice for nonrespondents (since 31/12/2015)
      - Publication of the "Survey Information" for respondents on the ISTAT website (since 31/12/2015)
    - Follow-ups of nonrespondent units
      - Telephone follow-ups (since 31/12/2015)
      - Mail follow-ups (since 31/12/2015)
      - Follow-ups by e-mail (since 31/12/2015)
      - Follow-up letter signed by the director (since 31/12/2015)
      - Follow-ups by certified e-mail (since 31/12/2015)
    - Use of GINO++ for web data capturing (since 31/12/2015)

Webinfo Privacy policy Contact us Copyright



# Terminology issues in joint GSBPM-GSIM applications

GSIM IOs (Option 2)	GSIM IOs (Option 1)	GSBPM concept
Statistical Program	Statistical Program ↔	Statistical Business Process
Business process	??	Phase
Process step Process step	Business process Process step	Sub-process
Process control – Metrics – Process step	Process control – Metrics – Process Step	Overarching processes





## GSIM definitions

- **Business process:** *The set of Process Steps to perform one of more Business Functions to deliver a Statistical Program Cycle or Statistical Support Program.*
- **Process step:** A Process Step is a *work package that performs a Business Process.*



## What is missing?

- SUM-MS already modelled successfully according to GSIM Concepts and Structure Groups
- However, the definition and explanatory notes of *Variable* show some contradictions

Definition	Explanatory Text
The use of a <i>Concept</i> as a characteristic of a <i>Population</i> intended to be measured	The Variable combines the meaning of a Concept with a Unit Type, to define the characteristic that is to be measured.  Here are 3 examples - 1. Sex of person 2. Number of employees 3. Value of endowment

«Number» => Concept

«employees» => Unit Type

But what if:

«Number of employees» => Concept

«Enterprises» => Unit Type ?

Not a Population characteristic!

## Proposal

- Change the definition of Variable in GSIM, substituting Population with Unit type.
- Among Unit types, it should be stated the existence of **elementary** and **aggregate** unit types.

*Variable* can be attached to both individual and aggregate unit types:

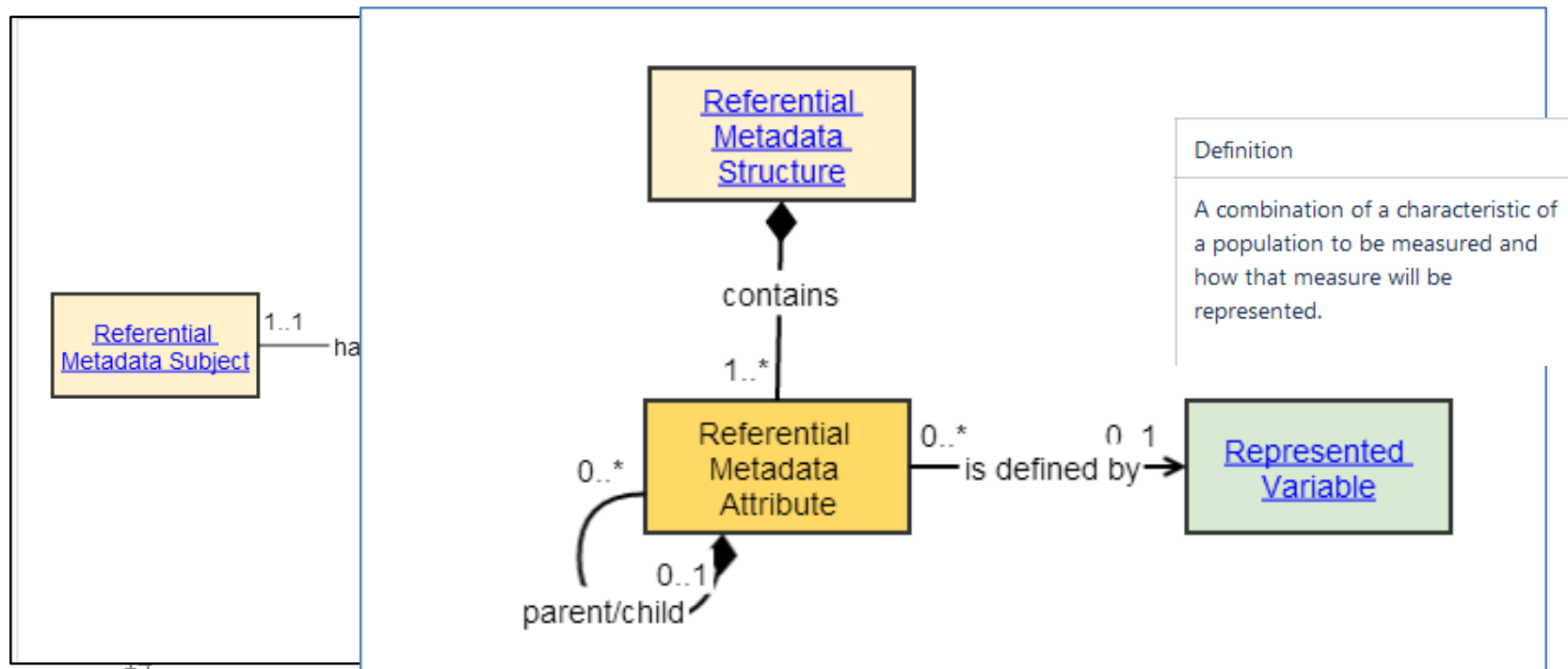
- specifying that *Sex* of person (singular) is restricted to the elementary case,
- referring *Number of employees* to aggregate unit types, possibly grouped according to other variables (observed at the unit level) as the reference area or the NACE code of each enterprise

## What is missing?

- The **reference metadata elements** in SIDI/SIQual are «quality oriented» and highly structured
- Correspondence for most SIDI/SIQual elements has been found in GSIM IOs, e.g. in the Exchange group: Questionnaire, Administrative register...
- However, some IOs could be more detailed, e.g.:
  - *Statistical Need* is a bit too generic to represent Regulations
  - *Product* can be better specialised
  - Relationship between *Information Provider* and different *Exchange channels* can be detailed

## What is missing?

- The **reference metadata model** in GSIM Structure Group resembles the SDMX information model and it is quite simple



# What is missing?

- Reference metadata attribute examples

## 7.1. Confidentiality - policy

[Regulation \(EC\) No 223/2009 on European statistics](#) (recital 24 and Article 20(4)) of 11 March 2009 (OJ L 87, p. 164), stipulates the need to establish common principles and guidelines ensuring the confidentiality of data used for the production of European statistics and the access to those confidential data with due account for technical developments and the requirements of users in a democratic society.

The [European Statistics Code of Practice](#) provides further conditions that have to be respected by statistical offices in regard to statistical confidentiality (Principles 1 and 2). In Italy, according to the article n. 9 of the Legislative Decree n. 136/2008, data collected by statistical offices within the statistical surveys included in the list of confidential data can be disclosed other than in aggregated form such that no reference can be made to individual units. Furthermore, they may be used only for statistical purposes. Data can be disseminated neither to any external subject, public or private, nor to any administration other than in aggregate form and using modalities and procedures approved by the units involved. The Code of Conduct annexed to the Legislative Decree n. 136/2008 (Personal Data Protection Code) provides special rules concerning the use of data for statistical purposes within SISTAN (see section 6.1). In order to ensure the effectiveness of personal data protection, Istat is currently taking appropriate organizational and statistical measures in accordance with internationally established standards. Under the Personal Data Protection Code, respondents are informed of the purpose of the provision of information, and they are assured that the information they provide will be used for statistical purposes only.

Further principles concerning the protection of the data confidentiality, are established by the Code of Professional Ethics and Good Conduct for the Treatment of personal data for statistical purposes and scientific research purposes performed within the framework of the National Statistical System (Legislative Decree no. 196, June 30, 2003).

Institutional relations between Istat and the Bank of Italy are managed by a Coordination Committee, established within the framework of a Memorandum of Understanding which regulates cooperation between the two Institutions in the field of research and exchange of statistical information. The Committee has defined a procedure for the bilateral exchange of micro-data in the cases provided for by Regulation (EC) No 223/2009 of the European Parliament and of the Council, art. 21 and of Council Regulation (EC) No 2533/98, art. 8a. Arrangements for the exchange of data between Istat and the Bank of Italy for research purposes are being developed.

## 6.2.1. Sampling error - indicators

Coefficient of variation (taking into account the sampling design) or estimated sampling error for the number of job vacancies (see guidelines).

Coefficient of variation (%)

Economic activity (Nace Rev. 2)	2016Q1	2016Q2	2016Q3	2016Q4
B	33.9	37.6	4.1	21.7
C	7.9	11.8	6.9	17.5
D	17.5	10.7	16.2	15.3
E	27.5	27.7	15.5	16.5
F	13.7	16.4	16.6	15.7
G	12.3	19.5	17.2	10.9
H	14.7	18.5	21.5	15.4
I	15.2	16.4	13.5	16.3
J	11.4	17.8	16.8	21.0
K	11.4	12.3	10.3	14.0
L	19.0	26.8	28.7	25.4
M	22.3	14.9	13.8	12.7
N	17.9	26.8	21.5	18.7
P	18.4	37.7	23.9	18.2
Q	18.5	28.9	18.3	21.7
R	26.0	35.5	25.6	20.6
S	18.0	23.9	27.5	24.2
Total B-S	4.3	6.8	4.7	7.2

A reference metadata attribute could range from a general text to a structure dataset... the GSIM reference metadata model should be refined!



**THANK YOU  
FOR  
YOUR  
ATTENTION !  
ANY QUESTIONS?**

**ANY... ANSWERS?**