Workshop on Methodologies in Official Statistics

Session 4: Standardization of methods and processes

Chair

Maurizio Lenzerini

Sapienza University of Rome, Italy

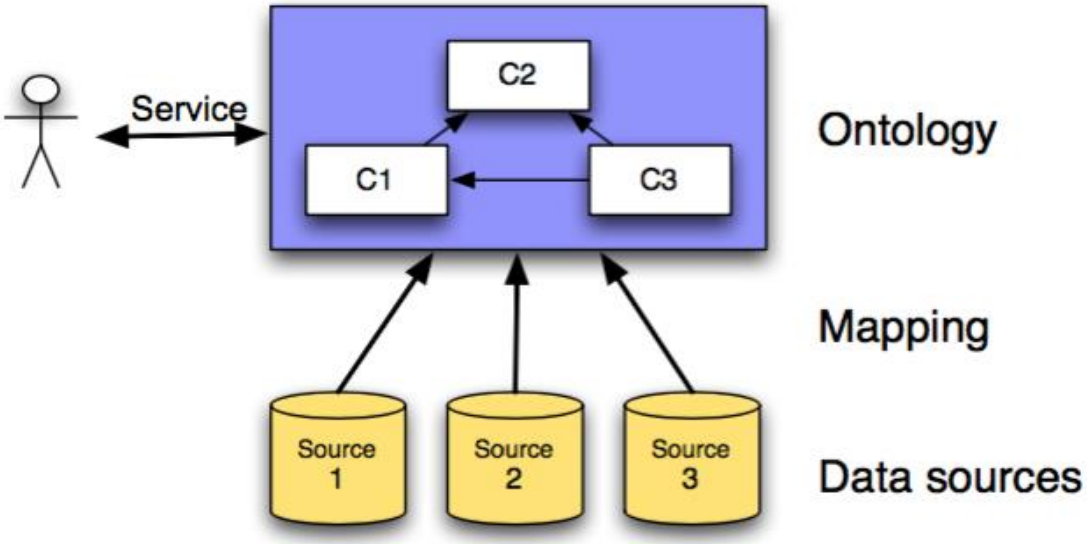
The context

- Since 2014 ISTAT has been carried out a modernization program with the aim of improving efficiency in statistical production
- Standardization is one of the key elements of the program
- Models such as GSIM (Generic Statistical Information Models) and GSBPM (Generic Statistical Business Process Model) play a central role
- Main aspects to standardise:
  - Metadata and domain modeling
  - Workflow in the statistical process
  - Methods and tools for the various phases of the process

Metadata and domain modeling

- Ontology-based data integration, whose architecture consists in
  - ontology
  - data sources
  - mapping between data sources and ontology

The proposed approach: Ontology-based data integation



Based on three main components:

- Ontology, a declarative, logic-based specification of the domain of interest, used as a unified, conceptual view for clients
- Data sources, representing external, independent, heterogeneous, storage (or, more generally, computational) structures
- Mappings, used to semantically link data at the sources to the ontology

The proposed approach

- Services:
  - analysis of formal specification of the domain
  - data governance based on formal specification
  - ontology-based data quality
  - ontology-based query answering
  - semantic open data publishing

Advantages of the proposed approach

- Domain (metadata) documentation
- Data source documentation (what's in a source?)
- Abstraction: data access based on concepts/relationships
- Access to integrated data
- Reasoning implicit in all usage of the system
- Data quality checking
- Open data publishing

Risks of the proposed approach

- Designing the ontology/mapping can be costly
- Getting consensus on the ontology can be hard
- Getting good source profiling may be difficult
- Maintaining the ontology/mapping can be hard
- Choosing the right implementation can be problematic
- Methodologies for ontology-based open data publishing not yet consolidated
- Designing the right services to users is also an issue

Papers discussed in meeting with the Advisory Committee on Statistical Methods

2017
On the design and implementation of a Generalized Process for Business Statistics,
by *M. Bruno, D. Infante, G. Ruocco, M. Scannapieco*

The Italian Integrated System of Statistical Registers: design and implementation of an ontology-based data integration architecture,
by *R. Radini, M. Scannapieco, L. Tosco*

2018
A new framework for quality assessment of processes based on Integrated Administrative Data,
by *F. Rocci, R. Varriale, G. Brancato, O. Luzi*

2021
Longitudinal and cross-sectional analyses of data in the Integrated System of Statistical Register
by *F. Altarocca, M.R. Aracri, R., Benedetti, R. Radini, G. Vaste*

2022
On designing aggregated data as Statistical Data Cubes,
by *M. Scannapieco, M. Scanu, L. Tosco, A. Bianco, M. Riccio*

This session

1. Overview of the ISTAT activities and open problems,
   Carlo Vaccari (ISTAT)

2. Metadata for statistical processes on registers,
   Mauro Scanu (ISTAT)

3. Overview of new approaches on the topic,
   Fabio Ricciato (Eurostat)

4. The point of view of the Statistical Production Department,
   Alessandro Faramondi (ISTAT)