Business Architecture Principles to Foster Industrialisation and Standardisation at the Italian National Institute of Statistics

Giulio Barcaroli[1], Piero Demetrio Falorsi[1], Alessandra Fasano[1], Nadia Mignolli[1,2]

[1]Istat - Italian National Institute of Statistics, Rome, ITALY

[2]Corresponding author: Nadia Mignolli, e-mail: mignolli@istat.it

Abstract

The Italian National Institute of Statistics has recently adopted the common vision proposed for the European Statistical System, envisaging a different production process of the statistical information, based on its standardisation and industrialisation, thus achieving higher efficiency and quality levels together with lower respondent burden. In order to ensure a proper governance of the transition process, a rigorous definition of the "to be" model is needed: for this purpose, Istat has defined a Business Architecture model, clearly stating the characteristics of the statistical value chain, the way the activities in the different domains interact, the general principles informing the whole process, and the infrastructures required so as to ensure its optimality.

Keywords: modernisation, innovation, business environment, Service Oriented Architecture (SOA).

1. Introduction

The Italian National Institute of Statistics (Istat) has been engaged for years in a series of complex challenges so as to work with a more efficient and optimised approach.
This is due to the need of increasing the production and quality of statistical information and of reducing both respondent burden and production costs.
Istat is presently characterised by multiple organisational models concerning financial, technological, regulatory sectors, etc., that tend to be inconsistent. This lack of homogeneity of language and organisational approaches makes the dynamics of change expensive and complex, and draws the attention to the importance for Istat to adopt a common language, enabling all its components to conceptualise both the given situation ("as is") and the one to be reached at the end of the evolution process ("to be"). The description of these present and future conditions also permits to design a path towards possible changes in a more rational and measurable way, defining specific actions involving different skills that need to interact within a shared view of a tangible progress. For this purpose, the reference to Enterprise and Business Architecture (EA and BA) models greatly help this modernisation process. EA is defined as "*the process of translating business vision and strategy into effective enterprise change by creating, communicating and improving the key requirements, principles and models that describe the enterprise's future state and enable its evolution*" (Gartner 2012). BA can be defined as the conceptual part of the EA, i.e. the part not involving the technological aspects. BA deals with Business, Information and Information Systems (exclusively for what concerns the conceptual aspects) and is called to play a central role in a programme as complex as that of industrialisation and standardisation of the statistical information production. In this context, it is possible to move from the level of a conceptual representation towards more and more operational and technological stages throughout the definition and monitoring of the principles governing BA. A fundamental choice in defining Istat Business Architecture was the adoption of a Service Oriented Architecture (SOA) approach.

SOA is based on structured sets of modules, known as services, that collectively provide the functions required by complex applications. Each service is designed to provide a compactly defined group of functions. This makes it possible to reuse an application in different ways by changing only the way an individual service interoperates with the other services that compose the application, instead of making code changes to the service itself (Bieberstein et al. 2005).
This paper describes Istat BA model, focusing on a fundamental component, that is the set of principles aiming at the modernisation of statistical production processes, well-defined and shared in order to ensure the success of the overall innovation project. In particular, the BA model here illustrated is inspired both by the Dutch Central Bureau of Statistics (CBS) (Bredero et al. 2009) and by the international standard GSBPM (Generic Statistical Business Process Model, Version 4.0) (UNECE/Eurostat/OECD 2009).

2. Description of Istat Business Architecture

2.1 Business domains and activities

Industrial organisations are expected to:

- define their strategic objectives and plan the activities that allow to achieve them (*Strategy*);
- design the processes corresponding to the planned activities (*Design*);
- organise the designed processes taking into account the operational constraints (*Management*);
- implement the processes ensuring efficiency and quality (*Implementation*).

Using a language typical of the organisational approach (Giachetti 2010), these four areas are called *business domains* and are defined in order to guarantee independence from the current organisational structure, so as to ensure their stability with regard to any future processes of reorganisation.
Actually, at the moment Istat is organised in different Departments that tend to replicate inside them the same organisational model, instead of referring to a unique one at the enterprise level. This is a kind of stovepipe model characterised by strong heterogeneity (of procedural, methodological and technological approaches), lack of standards and redundancy of data and applications.
Taking into account these drawbacks, in Istat BA model these four distinct domains have been defined, homogeneous with respect to the purpose of the activities carried out inside each of them, and to the nature of the information processed and/or services that insist on it. A particularity of the model is that each domain produces outputs that are used by the others, i.e. there are tight connections among domains.
Each business domain is characterised by specific activities that are aligned with GSBPM phases and sub-processes, except for the *Strategy*, whose activities are not considered in GSBPM. Each activity realises one aggregate service that transforms the input information into the planned output.
Strategy provides the framework (i.e. all the methods, processes and rules) for the overall statistical process high-level organisation and control. Its products include: budget, regulations, agreements with other bodies/agencies, strategic planning, standards, capacity management, reports from the management. Activities carried out in this domain are:
S1. Maintenance and consolidation of strategic relations;
S2. Budget definition;
S3. Strategic Planning;
S4. Internal and external data source management;
S5. Policy definition for process improvement;
S6. Project portfolio management and Capacity management.
Design produces the meta-information essential for the functional organisation and for the statistical process control. Its products comprise: technical designs, action patterns,

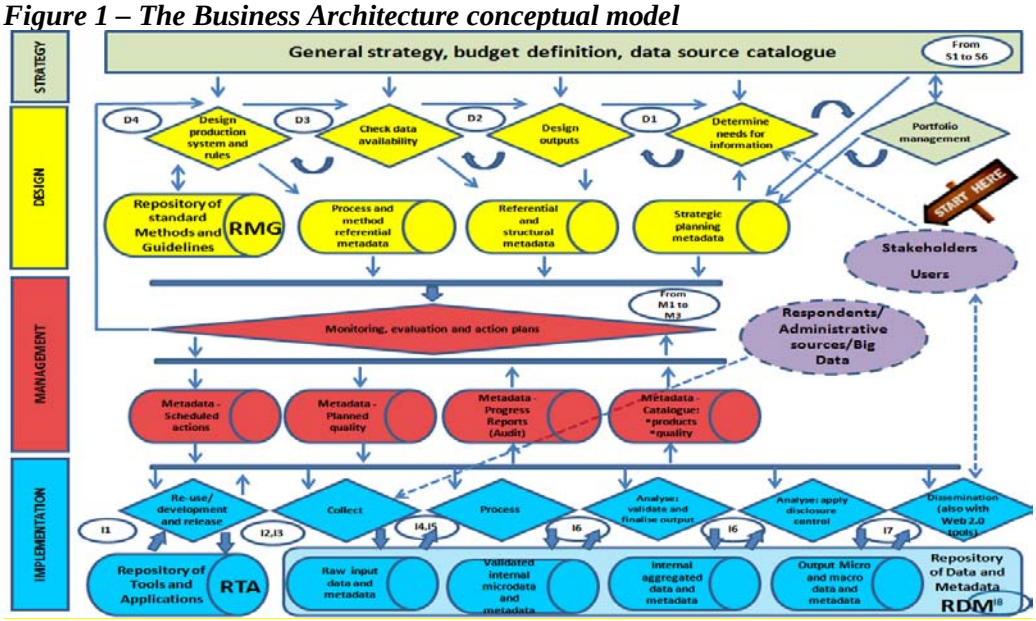
instructions, process indicators and their description. These are the activities:
D1. Determine needs for statistical information (corresponding to GSBPM sub-processes 1.1, 1.2, 1.3, 1.4 and 1.6);
D2. (Re)Design statistical outputs (GSBPM sub-processes 2.1 and 2.2);
D3. Check data availability/(Re)Design data sources (GSBPM sub-processes 1.5);
D4. (Re)Design production system and rules (GSBPM sub-processes 2.6 and from 2.3 to 2.5).
Management utilises the control information in real time. Its products embrace: the scheduling of activities, description of results, state implementation, reports on the achieved quality and plans for the improvement and adjustment of procedures carried out. Management activities are:
M1. Planning (aligned to GSBPM sub-process 3.3);
M2. Monitoring (GSBPM sub-processes 9.1 and 9.2);
M3. Adjustment (GSBPM sub-process 9.3).
Implementation realises the transition from the initial sources to the statistical information. Its products include: data archives and the metadata that describe them, as well as application tools used in the process. Implementation activities are:
I1. Tool and application reuse/development and release for the production (compliant with the whole phase 3 of GSBPM);
I2. Collect: preparatory stage (sample selection; set up collection) (corresponding to GSBPM sub-processes 4.1 and 4.2);
I3. Collect: run and finalise data collection, administrative source acquisition, standardisations (GSBPM sub-processes 4.3 and 4.4);
I4. Process: integration, classification and coding, editing, imputation, new variables and statistical units derivation (GSBPM sub-processes from 5.1 to 5.5);
I5. Process: calculate weights and aggregates (GSBPM sub-processes from 5.6 to 5.8);
I6. Analyse: validate, apply disclosure control and finalise outputs (whole phase 6 of GSBPM);
I7. Disseminate: produce and release dissemination products (whole phase 7 of GSBPM);
I8. Storage in the Repository of Data and Metadata (RDM) (whole phase 8 of GSBPM).

2.2 The value chain

The statistical process represent a logically ordered chain of activities that can also be considered as a value chain, as each step increases the value of the statistical product.
Business domain activities and products are the main elements of statistical process/value chain. If processes are ordered in a logical way, it becomes possible to represent the process chain, which is iterative and starts and concludes with stakeholders and common users.
In Figure 1, the overall scheme is divided in four layers, identified by the vertical rectangles indicating the different business domains. Inside each layer, rhombuses are specific activities or business services (processes); the labels of each activity are indicated inside the little white ovals, and refer to the labels of the activities described previously (paragraph 2.1). Cylinders symbolise information products (data and metadata) or applications; the biggest ones are the common shared repositories. Finally, large ovals indicate actors and external entities.
Although the scheme of BA model is static, the statistical process is dynamic by nature and is realised through the iterative repetition of chain parts, which also depend on the possibility of reusing processes and/or information products. Changes to a node in the step sequence can impact on all subsequent ones, but should have no direct impact on the previous stages. The overall value chain can be divided in two main sub-chains: the first is given by *Strategy* and *Design* activities, the second by *Management* and *Implementation* activities. Once the production process is well underway, in the subsequent iterations of the same process only the set of

Management and *Implementation* activities may be automated and repeated on a regular basis for the release of specific statistical products.

Figure 1 – The Business Architecture conceptual model



Source: Our processing of Statistics Netherlands BA Model and GSBPM, 2012.

3. Infrastructures and principles

In Istat BA model, a central role is played by both infrastructures (elements enabling standardisation and industrialisation) and principles (general rules to be followed to ensure optimality).

3.1 Infrastructures

The most important infrastructures needed for the efficiency and efficacy of the overall process are:

- the *Repository of standard Methods and Guidelines* (RMG), that contains the set of statistical methods, recognised as standards, to be applied to processes;
- the *Repository of Data and Metadata* (RDM), containing input data, intermediate data and output data ready for dissemination, with defined quality standards and metadata;
- the *Repository of Tools and Applications* (RTA), including three distinct categories of software (generalised IT tools, reusable applications and ad hoc applications).

The definition of methods as standard, to be shared in RMG, has to follow a precise procedure that is to be adopted by the statistical organisation, also taking into account the international framework. Following this approach, Istat has already adopted such a procedure that is being used in order to define the first standards in the course of 2013.
The information produced at the various stages of processing should be conceived as a common good, to be stored in the available infrastructure of RDM.
In the same way, the set of production applications of the Institute should be also considered as a common good to preserve and share, maximising the use and reuse, through the infrastructure of RTA.

3.2 Principles

The whole BA model is led by fundamental principles that become practical guidelines for the implementation of each business domain activity and for ensuring the success of the model itself. In particular, ten different governing principles have been suitably defined and addressed specifically to Istat, also taking into account the international and European context. These principles regard the overall governance, the process rules and the specific infrastructures.

In the following they are listed and described, with a synthetic explanation and the indication of BA domains where their impact is expected.

Principle 1: "*The statistical process is a logical and value chain of activities that have to be clearly separated in different business domains and controlled by rules specifically defined*". Within *Strategy* domain, rules and objectives that guide the statistical process are defined. Within *Design* domain, conceptual metadata of process and quality are defined. Within *Management* domain, information (metadata) are defined in order to guide planning, monitoring and adjustment/adaptation of a specific statistical production process. Statistical data are produced within *Implementation* domain following what defined in previous domains.

Principle 2: "*The whole statistical process is output-driven*". The statistical process chain starts from the output desired, that is from required products, and goes backwards, defining the various aspects of the process. Impact on *Strategy* and *Design*.

Principle 3: "*It is necessary to maximise the reuse and adoption of standards contained in the Repository of standard Methods and Guidelines (RMG) in order to standardise all statistical processes: the Service Oriented Architecture (SOA) is the most suitable for achieving this purpose*". Reuse focuses on both what is produced within the Institute and what is issued outside, with particular attention to the standards defined at international and European level. The reuse of existing and available data is generally to be preferred over the decision to conduct a new survey. The "to be" statistical production consists of a series of standardised single processes and of modular services that can be shared and reused in different contexts and statistical areas. The implementation of this principle is based on the development of common IT solutions and a better allocation of tasks and functions within the organisational structure. Impact on *Design* and *Implementation.*

Principle 4: "*It is necessary to ensure the independence between design and implementation, thus promoting the industrialisation of statistical processes*". The assumption of an industrialised process is that it can be realised by agents other than those that have designed it. Design is performed only when needed, while a current statistical process is carried out on a regular basis. Implementation of a new project involving several innovations requires a new design phase. Statistical production has a repetitive nature with a rather rigid organisation style that can be largely automated. Impact on *Design* and *Implementation.*

Principle 5: "*No regular production activities can be conducted without having completed the design of referential metadata of model, process and rules*". Firstly metadata are designed and then data production can start. In principle, metadata have to be generally accessible and, as far as possible, standardised with regard to the types of units, the definition of concepts, classifications, quality characteristics, process. Impact on *Design*, *Management* and *Implementation*.

Principle 6: "*It is necessary to clearly distinguish between the data actually processed and the metadata describing definitions, quality and process activities. The set of data and metadata represent a shared asset to be optimized*". Impact on *Design, Management* and *Implementation*.

Principle 7: "*Quality assessment and documentation are the basis for the continuous improvements of the statistical process*". Quality has to be evaluated and documented at the different stages of the statistical production process: it is defined and planned during design or redesign. It is monitored and assessed in each phase of GSBPM and in correspondence of intermediate and final data releases. Impact on *Design, Management* and *Implementation*.

Principle 8: "*The Repository of Data and Metadata (RDM) and the Repository of Tools and Applications (RTA) represent an Istat shared asset. It is necessary to exactly define its products, so as to ensure that they can be reused, traceable, mutually related to their different versions. For each data release, the output quality represents an input for the next stage*". RDM contains input data, intermediate data and output

data ready for dissemination, with defined quality standards and metadata. At least four release levels are provided: raw micro data; validated microdata; aggregated data for internal use; micro or macro data for external dissemination; data and metadata, including standard indicators of quality, are stored and made accessible in RDM, i.e. the common Repository, so as to ensure their reuse. The set of the Institute production tools and applications has to be stored in the Repository of Tools and Applications (RTA). It contains: generalised IT tools; reusable applications; ad hoc applications, non-reusable. Impact on *Implementation*.

Principle 9: "*It is necessary that the internal statistical information flow goes exclusively through the Repository of Data and Metadata (RDM) infrastructure*". The Institute structures cannot exchange data and metadata through bilateral agreements (taking them directly from other structures) but they have to release them and to take them through RDM. Impact on *Implementation*.

Principle 10: "*It is necessary that the production process be implemented on the basis of generalised tools and the reuse of the applications contained in the Repository of Tools and Applications (RTA): developments from scratch should be limited to what is not already available*". Before proceeding to the development ex novo, it is always necessary to verify the possibility of using generalised software already available, or applications developed for similar types of processes and, therefore, reusable with a limited customisation effort. Impact on *Implementation.*

4. Conclusions

In this paper is outlined a "to be" model of Business Architecture, which offers an integrated view of the statistical production process, that can facilitate the realisation of Istat industrialisation and modernisation. In the medium period, the BA conceptual model can be made more and more operational through a road map properly designed and scheduled, focusing particularly on the implementation of some of the basic infrastructures provided, in terms of both procedures (such as the management of the Portfolio, the compliance assessment of the several statistical production processes with the BA and EA principles, with their subsequent validation, etc.) and shared services (the Repository of standard Methods and Guidelines, the Repository of Data and Metadata, the Repository of Tools and Applications).

In this road map, a particular attention is focused on the realisation of Istat Business Architecture giving priorities to the definition of:

- the governance, to provide the strategic directions, to coordinate, to validate and to monitor all the necessary activities, also involving some Istat reference committees (e.g., methodology, quality, innovation and research etc..);
- the organisational structure, for the different operational phases and to advise the governance arrangements for the basic infrastructure implementation;
- the communication process, to share BA principles within Istat and to disseminate their knowledge.

References

Bredero, R., Dekker, W., Huigen, R. and Renssen, R. (2009). Statistics Netherlands Architecture; Context of change. Discussion paper (09017), CBS Statistics Netherlands, The Hague Heerlen.

Bieberstein, N., Bose, S., Walker, L., Lynch, A. (2005). Impact of service-oriented architecture on enterprise systems, organizational structures, and individuals. IBM Systems Journal, Vol. 44, Number 4.

Gartner (2012). IT Glossary. Defining the It Industry, http://www.gartner.com/it-glossary/enterprise-architecture-ea/.

Giachetti, R.E. (2010), Design of Enterprise Systems, Theory, Architecture, and Methods. Boca Raton, FL (USA): CRC Press.

UNECE/Eurostat/OECD (2009). Generic Statistical Business Process Model (GSBPM Version 4.0), Work Session on Statistical Metadata (METIS, April 2009).