Towards an Enterprise Architecture for an NSI

Boris Lorenc[1], Ylva Fossan[2], Eva Holm[2], Ingegerd Jansson[1]
[1]Statistics Sweden, Stockholm, e-mail: firstname.lastname@scb.se
[2]Statistics Sweden, Örebro, e-mail: firstname.lastname@scb.se

Abstract

This paper summarizes ongoing work on development and implementation of enterprise architecture for Statistics Sweden. For structuring the different levels of the analysis, our work borrows from the Zachman framework, but as a general methodology for developing and implementing the architecture, we consider The Open Group Architecture Framework (TOGAF). TOGAF is a structured and widely deployed framework into which Statistic Sweden's systems development model and project management model may be fairly easily integrated. The framework supports creation of repositories that, in the course of practicing TOGAF, are populated with architecture building blocks and solution building blocks, which in turn enables standardized and reusable solutions that are economically viable.

Development and implementation of enterprise architecture (EA) at Statistics Sweden goes on prominently in areas corresponding to the Generic Statistical Business Process Model (GSBPM). As a component of EA, business architecture (BA) is developed for the processes in GSBPM.

Main modelling activities include modelling of information (objects, attributes and relations) and of processes that the business needs to carry out in order to fulfil its goals. This includes creating conceptual models and adopting architecture-wide definitions. BA supports development of IT-solutions that in turn support a process-oriented approach to the production of statistics.

Keywords: Enterprise architecture, business architecture, TOGAF, information modelling, process modelling.

1. Introduction

The mission of a national statistical institute (NSI) and other producers of official statistics is to provide – in an economically effective way – statistics of sufficiently high quality that meet users' needs. As quality itself is a multifaceted concept (Eurostat 2003) encompassing relevance, accuracy, timeliness, and other important aspects, running and developing modern statistical production is a complex task that involves heeding several driving forces: changing needs of users, methodological achievements, improvement of process quality, an ever-changing IT-landscape, and budget requirements. To manage a modern statistical production system and to manage the transition of an NSI towards such a production system are therefore rightly seen as being exceptionally complex activities.

In managing the production of official statistics, several NSIs have experienced a strategic gap between high-level goals of the organisation (its vision and strategic aims) on the one hand and the capability of goal implementation and change management at the lower levels (applications and IT infrastructure, middleware, etc) on the other. Consequently, official statistics production has been tailor-made within subject matter areas and individual products within them, resulting in a wide variety of technical solutions and ways of implementing them, often weakly documented. High-level goals like "better use of process data" or "more effective contact strategies" are difficult to implement as any change requires hands-on activities, often in turn being dependent on close, personal knowledge of each of the systems to be changed.

Demands for increased economic efficiency and swifter reaction to changes in users' needs have led many NSIs to work towards improved production management (see also European Commission's COM 404, EC 2009). The directions of change include – among others – standardisation of the production process, increased reuse of tools, and increased transparency and manageability through greater use of process data (both for process quality and cost follow-up) and metadata. Change management process at NSIs is thus in itself a complex endeavour.

However, production of statistics is just one of the many endeavours that are getting more complex in modern society. When the complexity is growing, there is an increased need to reduce it and simplify the solutions through use of modern know-how. The latter includes EA, with the aim of connecting the high-level goals of an NSI with lower-level processes and tools for statistics production. By building reusable components, quality is improved and costs reduced.

The emergence of EA was in fact the result of dealing with two related issues (Sessions 2007):

- increase in system complexity: the number and extent of IT systems of an enterprise,
- poor business alignment: an increase in the pace of change in external business environment that requires faster business change, and consequently faster IT change, which the old ways of managing and developing IT tools were not able to keep the pace with. Thus, enterprises were "finding it more and more difficult to keep those increasingly expensive IT systems aligned with business needs" (Sessions 2007).

This paper argues that NSIs should embrace enterprise architecture much more than what has been the case thus far, and that – in doing so – they should not "reinvent the wheel" but instead borrow from experiences of the business activities in general (amongst them, the knowledge-based service sector: NSIs are not unique). In the course of the preceding 25 years or so, EA has become the way to structure and manage complexity of running and developing modern companies, and nothing precludes NSIs from being able to use this knowledge in their field.

The paper has the following structure: in Section 2 we introduce EA with respect to its history and its currently existing main varieties, including a discussion of aspects that has guided Statistic Sweden's reasoning about EA choice. In Section 3 we present the approach and main components of the work carried out at Statistics Sweden with respect to EA and BA, especially process modelling and information modelling. In Section 4 we give two concrete examples of work with some aspects of EA, namely details about process and information modelling for some parts of the data collection process of the GSBPM, and a preparatory TOGAF work on an envisioned data base of all Statistics Sweden products. We conclude with some reflections and recommendations in Section 5.

2. Some components of enterprise architecture

EA as a concept emerged from the need to manage complex information systems. The reference often quoted as the cornerstone of EA is Zachman (1987), which however in fact does not use the term "enterprise architecture" but rather "information systems architecture". The realisation that information systems cannot be treated by themselves, separated from the business they are carrying out and the whole of the enterprise that they are supporting, came gradually in the course of the past two decades or so.

There are a large number of definitions of EA. Here we give a fairly recent one:

> Enterprise architecture: a coherent whole of principles, methods and models that are used in the design and realisation of an enterprise's organisational structure, business processes, information systems, and infrastructure [Lankhorst 2009].

One of the major characteristics of an EA is its hierarchical structure: in order to handle complexity, a system is simplified by specifying different levels. In the above definition, these are business, information, and infrastructure.

The Zachman architecture framework, considered by many to be the leading one due to its completeness in specifying the ontology of EA (e.g. Sessions 2007), identifies these commonly used levels:

- context
- business
- logic (alternatively, information)
- technology
- components
- operation instantiation

But, to define the ontology of an enterprise, yet another classification is necessary, corresponding to answering the interrogatives *who*, *where*, *when*, *why*, *how*, and *what*. This gives the other main dimension of Zachman's framework, which is concerned with organisational, spatial, temporal, motivational, process and content decompositions of an enterprise. By crossing the levels with responses to these interrogatives, one gets the rich structure of documents that, when in place, defines an EA.

However, while abundant in structure, Zachman's framework says virtually nothing about how the work to define the architecture should be carried out. In looking for methodologies for its EA work, Statistics Sweden is considering The Open Group Architecture Framework (TOGAF). This is not a proprietary framework and it incorporates knowledge from a vast number of EA professionals and experiences from their organisations (The Open Group 2009). The main reasons for using TOGAF that Statistics Sweden identified are that it:

- is a structured and widely deployed framework,
- supports creation and population of repositories for architecture building blocks and solution building blocks,
- contributes to standardized and reusable solutions that are economically viable.

Statistic Sweden's systems development model, project management model and modelling as means of building transparent and flexible systems may be fairly easily integrated with TOGAF.

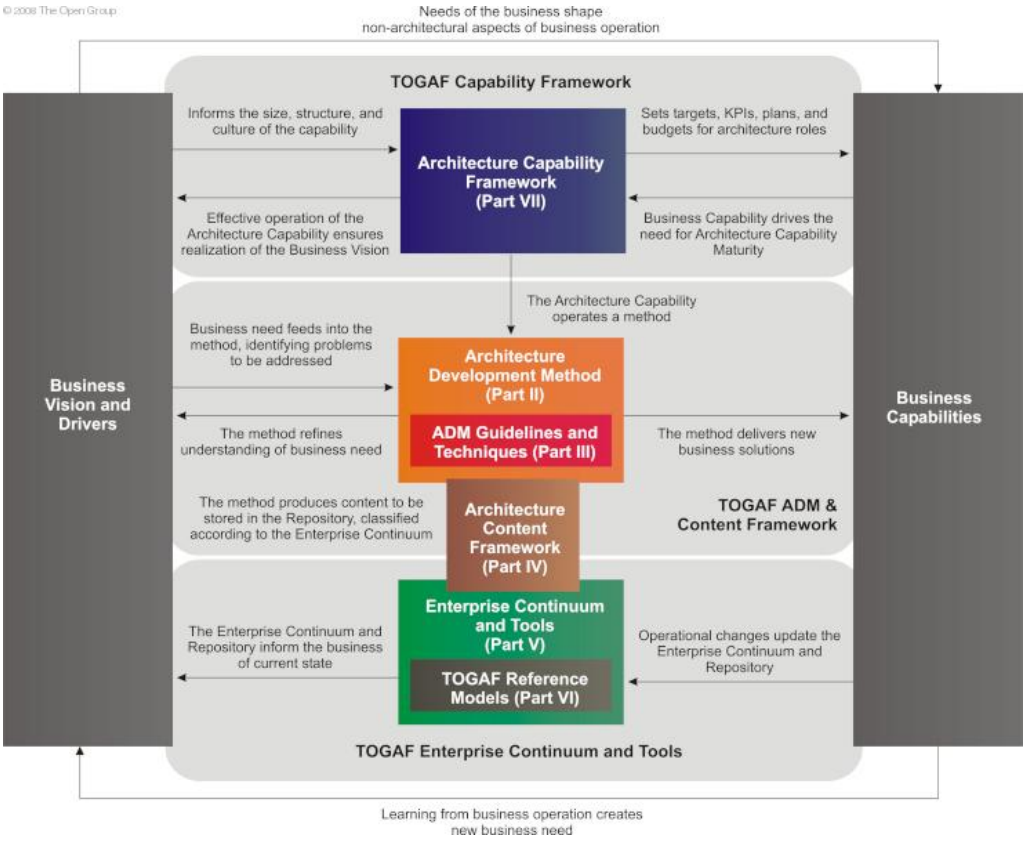

Figure 1: TOGAF content overview (The Open Group 2009)

The TOGAF model, a depiction of which is given in Figure 1, can be – as presented in Josey (2009) – seen as centring the EA on the Architecture Development Method (ADM). EA further requires an architecture capability to run the method. The method is supported by a number of guidelines and techniques. The EA work produces content – a component of the enterprise continuum – which is stored in a repository, initially populated with the TOGAF Reference Models. Thus, the basic components of an EA according to TOGAF are:

I. the method: ADM, to be presented in some more detail below,
II. repositories containing reusable building blocks: architecture building blocks and solution building blocks,
III. know-how for running the EA: an architecture capability as well as guidelines and techniques that support it.

A characteristic of the ADM is that it can be performed on an enterprise as a whole, on a component of an enterprise, or on an arbitrary subcomponent – the main steps of the method will be the same in all cases. However, a reasonable partitioning, i. e. definition of a business case, simplifies the modelling and results in a better manageable EA (c.f. Sessions 2008). The main steps are depicted in Figure 2.

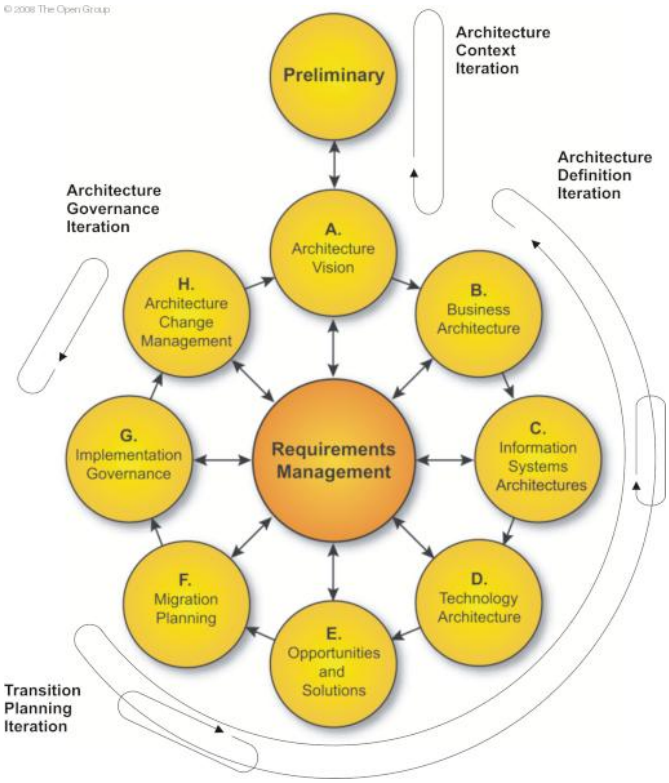

Figure 2: The Architecture Development Method of TOGAF (The Open Group 2009)

ADM consists roughly of the following phases:

- preliminary, for defining "where, when, what, why, who, and how we do architecture", including determining a rough scope of the project, its drivers, stakeholders, high-level requirements, architecture principles and the framework; relations with management framework(s) are determined, and the level of ambition of the architecture-to-be,
- an architecture vision,
- three definition phases, determining the business architecture, the information architecture, and the technology architecture,
- a number of phases regarding solution, migration and implementation planning of the architecture-to-be, and
- an architecture change management phase.

It cannot be overstated that, except for the preliminary phase, the work within a phase is not completed once and for all (as in the waterfall model). Previous phases are by the framework required to be iterated or revisited (c.f. Figure 2), while at the same time brought in mutual concordance with the architecture requirements. It can thus occur, for instance, that issues in the solutions phase lead to changes in information or business architecture phases and changes in requirements. However, all such changes are well documented and their impact assessed through iterations of the phases until a satisfactory EA is reached.

It aids application of the method that TOGAF offers, for each of the phases, clearly stated objectives, inputs, steps to be performed, and outputs. The framework provides templates for these documents.

In testing TOGAF at Statistics Sweden, we have seen that many of the documents in the ADM have close similarities to those that are already in use in regular project management at Statistics Sweden. Request for architecture work, statement of architecture work, organisational model, communication plan, and stakeholder analysis are examples of this. However, the documents differ somewhat in the way they are intended to be used and, in particular, in that the TOGAF artefacts clearly fit in a well-defined structure for supporting development. Among the new artefacts that we especially appreciate are the architecture repositories (consisting of architecture building blocks and solution building blocks), which contribute to an enterprise continuum and enable reusability of the blocks.

It is important to realize that the output of the ADM is a proposal for an architecture, while practical development of systems is not carried out during the phases of the ADM. Once the ADM is completed, it is the role of the project management or architecture board of the enterprise to decide if the projects proposed by the ADM should be realised. Thus ADM will give the enterprise a complete picture of the requirements, capabilities, and prerequisites of the organization, as well as the connections to other activities and projects within the organization and the consequences of realising (or not) of the proposed architecture.

3. Implementation of EA at Statistics Sweden

In this section, we present some principles for work with EA and BA at Statistics Sweden, comprising parts of the EA competence, as well as some results and an overview of the most relevant components.

Statistics Sweden's architects use four basic layers in structuring their work on EA and BA: business goals, processes, information and applications. Business goals span a space in which both business processes and information for carrying them out are realised. A relation of mutual compatibility needs to exist between processes and information. An exact specification of the goals, processes and information provides the context for implementation and infrastructure solutions. A consequence is that business rules, process flows, and information structures become represented in the models in a systematic, consistent, and unitary way.

BA and EA are carried out at Statistics Sweden by modelling on the aforementioned four levels. We do these models in their as-is and to-be versions. When it is done, an analysis of differences between the versions in the pair will indicate gaps in our current architecture and the way ahead. Results of the analyses will also provide a clear overview of the state of BA and EA at the NSI.

3.1. Process modelling

Formal description of the business process model at Statistics Sweden largely coincides with that contained in the generic statistical business process model (GSBPM, c.f. METIS 2009 for its version 4). However, a complete EA will provide corresponding models also for processes supporting the business, for instance human resources management or legal framework management.

Process modelling is carried out on a level that enables identification of common business subprocesses. The goal of the modelling is to produce a representation of the process flows. Currently, we have models for the following subprocesses in place:

2. Design
   i. Choose data source(s) and data collection method(s)
   ii. Choose contact strategy and identify relevant population groups
   iii. Decide on level of editing within data collection and choose data input method
   iv. Design the production flow
   v. Verify administrative routines
   vi. Plan and book resources

4. Collect
   i. Update sample
   ii. Prepare questionnaire distribution
   iii. Prepare scanning
   iv. Prepare web data collection
   v. Provide support to data provider
   vi. Manage double submissions
   vii. Manage reminders
   viii. Scan and check

5.3 Edit
   i. Error alerting
   ii. Automatic editing
   iii. Manual check

3.2. Information modelling

Information modelling helps formalise the description of the information used in the enterprise's business, with the aim of achieving compatibility with the IT infrastructure that is going to process the information. The models are conceptual, formal, and independent of, and able to be unequivocally interpreted by, the lower layers of application and infrastructure. We distinguish between two levels of models: object group models and (detailed) object models.

Object group models may be used for assigning ownership of information. Object models are the basis for a common concept repository in the enterprise. The models also are part of – and help specify – the general requirements space. For instance, no business goals are allowed to be implemented in physical data models without first being put into object models. By doing so, interpretation of the business goals is formalised and recorded, rather than left to arbitrary (and undocumented) interpretation of the staff doing application implementation.

Among architecture principle there is one stating that specified types of information shall not be created and/or stored in more than one place, such as *statistical survey*, *edit rule*, *status code*, etc. This, however, needs to be weighed against other architectural concerns, like creating too strong system dependencies (and thereby an unstable system).

Statistics Sweden has thus far developed the following information models:

a. Object group (for the models below)
b. Statistical survey
c. Sample object
d. Reporting object
e. Contact person for data collection
f. Sending material
g. Questionnaire
h. Sending occasion
i. Arrival occasion
j. Observational object
k. Data item
l. Edit round occasion
m. Edit round result
n. Auxiliary information
o. Contact occasion
p. Comment
q. Status code
r. Edit rule
s. Collection variable

3.3. Concept modelling

In order for a system to be shared by many, the concepts it uses need to be well defined. Concepts are implicitly present in process models and information models; however, crucial concepts may need to be formalised into well-defined concept models. We did such models for 5.3 Editing and 6.4 Disclosure control.

3.4. Architecture capability

Regarding the need for continued work on EA and structuring the different architectural functions involved, Statistics Sweden is implementing the following architecture-related roles: i) business architect, ii) solution architect, and iii) software architect. Additional competencies may be needed, for instance for infrastructure or specific technical platforms.

4. Examples of EA work

A so-called *architecture function* was established in 2008 within the R&D department of Statistics Sweden. The function consists of about ten senior methodologists and senior IT experts (who, however, all devote their time also to other duties than architecture). In the course of the group's existence, the architecture work consisted of benchmarking similar developments at other NSIs and of the EA field in general, in identifying and establishing working relations with the stakeholders at the NSI of relevance for the EA work, and in carrying out some actual EA work. Here we have space to briefly mention some of the latter instances.

4.1. General data collection platform

Statistics Sweden is in the process of developing a general platform for data collection and editing (Engdahl 2010). It will support the Build and Collect processes of the GSBPM, as well as subprocesses 2.3 and 2.6 of the Design process and 5.1-5.4 of the Process process. In contrast to traditional solutions, which focus on data storage, the platform is built to support business processes with needed data and metadata. An event-driven approach is used, where data and metadata are transferred between processes using Business Objects (as parts of a Business Information Model). In practice, a value chain is created for each Business Object type, taking

care of where the Business Object is created and what processes/services should be used to add value to the object. There is a communication platform that supports the information flow.

4.2. Product data base

Associating precise costs to products and to processes and subprocesses in the statistical value chain is difficult. However, a solution to this challenge would result in a much improved potential for cost-effective management of statistics production, as well as for empirically-based decision making of choices for process improvement (and in increased transparency of the production process as a by-product).

In cooperation with researchers from Royal Institute of Technology in Stockholm, we have been carrying out a pilot project that brought TOGAF and the quality management framework EFQM (e.g. Hakes 2007) to bear on this task.

4.3. Further development

We plan to continue applying TOGAF, next being the development of a coordinated system of registers within a data warehouse solution. Further, Statistics Sweden is currently evaluating the process of managing and coordinating its improvement and development processes. Integration with the basic thoughts from EA and application of TOGAF on relevant areas of development are possible ways towards effective production and high quality solutions.

5. Concluding remarks

While having a strong tradition of developing IT solutions for statistical production process, the experiences of Statistics Sweden with EA are fairly new. However, we see EA as a reasonable solution – in fact, the only solution – to the problems that have vexed the statistical production process and its management during a considerable period of time. By providing a unified and transparent representation of the NSI, EA helps bridge the gap between high-level goals and possibilities of implementation of these into the practices of the NSI; aids addition of support processes on top of the production processes, like process data collection or quality control; improves cost management and production management; supports effective development project portfolio management; and so on.

We have chosen to consider TOGAF as the framework from which to use methods and solutions for developing EA. TOGAF is an open framework which embodies experiences from a large number of enterprises with good track record in connecting business goals and IT. However, our basic belief is that the specific choice of an EA framework is not essential, and thus we encourage NSIs to simply start by applying any EA framework that they may find appropriate for their needs: Federal Enterprise Architecture Framework (or another government EA), Gartner's, TOGAF, or something else. Our impression is, judging from our experiences and those of the others, that the way to work on EA is not as a paper product but rather by being directly involved in projects and relevant areas of development.

A point possibly in need of making is that, in our opinion, NSIs should use existing frameworks and methodologies rather than invent their own. The latter is comparable to reinventing the

wheel, since existing frameworks have in general been developed with a far larger EA competence than what an NSI usually has at its disposal. EA frameworks are nowadays used not only in production enterprises but also in services, and there is no principal difference between providing statistical services and other kinds of services. We propose, however, that open frameworks (e.g. TOGAF) be preferred to proprietary ones, and accepted standards for statistics production be used (e.g. GSBPM), as this will enable easier integration of production systems across European NSIs. Together with more cost-effective management and development, this would speed up realisation of the vision of statistics production for the next decade (EC 2009).

Acknowledgements

The authors are grateful to Jakob Engdahl and Hans Irebäck, both at Statistics Sweden, for constructive comments on an earlier draft of the paper. Opinions expressed here are those of the authors, and do not necessarily reflect the views of Statistics Sweden.

References

EC (2009). *Communication from the Commission to the European parliament and the Council on the production method of EU statistics: a vision for the next decade*. COM(2009) 404. Available online: http://eur-lex.europa.eu/LexUriServ/LexUriServ. do?uri=COM:2009:0404: FIN:EN:PDF (accessed 17 January, 2011).

Engdahl, J. (2010). *An Event-Driven Architecture for Data Collection*. Presented at MSIS 2010 meeting, Daejeon, Republic of Korea, 26-29 April 2010.

Eurostat (2003). Methodological Documents – Definition of Quality in Statistics. Available online: http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/ess%20quality %20definition.pdf (accessed 17 January, 2011).

Hakes, C. (2007). The EFQM Excellence Model for Assessing Organizational Performance. Zaltbommel, NL: Van Haren Publishing.

Josey, A. (2009). *TOGAF Version 9 Enterprise Edition: An Introduction*. Document No. W094. The Open Group: San Francisco, CA.

Lankhorst, M., et al. (2009). *Enterprise Architecture at Work: Modelling, Communication and Analysis*. Berlin, DE: Springer-Verlag.

METIS (2009). *Generic Statistical Business Process Model*: Version 4.0 – April 2009. UNECE Secretariat. Available online: http://www1.unece.org/stat/platform/download /attachments/8683538/GSBPM+Final.pdf?version=1 (accessed 17 January, 2011).

Sessions, R. (2007). *A Comparison of the Top Four Enterprise Architecture Methodologies*. ObjectWatch: Austin, TX. Available online: http://www.objectwatch.com/whitepapers/ 4EAComparison.pdf (accessed 17 January, 2011).

Sessions, R. (2008). *Simple Architectures for Complex Enterprises*. Englewood Cliffs, NJ: Prentice-Hall.

The Open Group (2009). *The Open Group Architecture Framework (TOGAF), Version 9*. The Open Group: San Francisco, CA.

Zachman, J.A. (1987). A Framework for Information Systems Architecture. *IBM Systems Journal*, Volume 26, Number 3.