# A Systematic approach to Generate TOGAF artifacts founded on Multiple Data Sources and Ontology

Sérgio Guerreiro<sup>1,2,3</sup>[0000-0002-8627-3338] and Pedro Sousa<sup>1,2,3</sup>[0000-0001-7453-5778]

 <sup>1</sup> Link Consulting SA, Av. Duque de Ávila 23, 1000-138 Lisbon, Portugal
<sup>2</sup> INESC-ID, R. Alves Redol 9, 1000-029 Lisbon, Portugal
<sup>3</sup> Instituto Superior Técnico, University of Lisbon, Av. Rovisco Pais 1, 1049-001 Lisbon, Portugal
{sergio.guerreiro,pedro.sousa}@linkconsulting.pt, {sergio.guerreiro,pedro.manuel.sousa}@tecnico.ulisboa.pt

Abstract. Enterprise Architecture (EA) standards are widely used in industry and recognized as a *lingua franca* between project' stakeholders. However, standards are most of times very extensive, requiring specialized knowledge and consuming a large effort if instantiated manually. Furthermore, the risk of producing inconsistent artifacts increases with project' complexity, *e.g.*, changing the name of a business actor requires propagation to all artifacts using that entity. Moreover, EA generation consumes data from multiple sources, *e.g.*, excel or BPMN files, that need to be normalized, classified, and consistently referenced in the artifacts. This paper proposes a systematic approach where the conceptual understanding of a project is shared using an ontology which in turn supports the entire EA artifacts automatic generation. Results show that there are no similar solution available in the literature. In addition, the usage of our systematic approach in four different EA projects evidences a bounded linear increase in effort as artifacts increase in complexity.

**Keywords:** Artifact · Concept · Document · Enterprise Architecture · Ontology · TOGAF.

# 1 Introduction

Enterprise Architecture (EA) is defined by TOGAF [7] as "the structure of components, their inter-relationships, and the principles and guidelines governing their design and evolution over time". More recently, as noticed by [17], more thorough definitions can be offered, as the one proposed by Greefhorst and Proper [6] that consider three perspectives for architecture: "regulation-oriented, design-oriented, and knowledge-oriented, where the first corresponds to the prescriptive perspective, the second corresponds to the descriptive perspective, and the third corresponding to the high-level design decisions of the system". The EA way of working are proposed by many standards that are widely used in industry,

*e.g.* TOGAF, DODAF, MODAF, *etc.*, and are recognized as a *lingua franca* between clients, suppliers, consultants and/or researchers. However, these standards are most of times very extensive, requiring specialized knowledge and consuming large amounts of time and effort to be instantiated. In practice, large

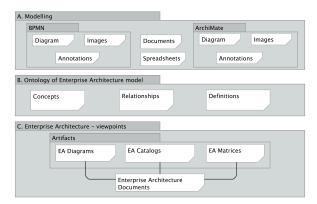


Fig. 1: Conceptual layers to construct EA viewpoints, an example using TOGAF standard.

EA projects are expected to generate, and maintain, complex artifacts (bottom part of Figure 1) that require a large effort if they are produced manually. Furthermore, the risk of inconsistencies between artifacts increases with the dimension of projects, *e.g.*, a change in a single business actor requires propagation to all artifacts that are using that entity to enable traceability. In the top of this problem, EA consume data from multiple sources (top part of Figure 1), *e.g.*, data sources files with modelling elements of the business, application, technology, motivation, or even simple spreadsheets; which need to be normalized, classified, and then, referenced on the artifacts.

From a different perspective, an ontology represents the fundamental **con**cepts, relationships and definitions that are used in a specific application domain [8]. It targets the facilitation and dissemination of its understanding between stakeholders with different interpretations of the same reality. Moreover, an ontology is the result of stakeholder's discussions reaching a consensual agreement between them [11]. We aim at integrating ontology representation techniques with EA artifacts construction, where ontology is used as a scale faithful to guide all the concepts that need to be included during EA generation and maintenance (middle part of Figure 1). In specific, the research question (RQ) addressed by this paper is: "How to generate, and maintain, the consistency of Enterprise Architecture artifacts in a complex organizational model, with a bounded linear increase in effort as TOGAF artifacts increase in complexity?". In short, this paper proposes a systematic approach where the conceptual understanding of a project is shared using an ontology, which is used to classify and to normalize the received data streams, and finally, to automatically generate artifacts that are referenced and synchronized by that understanding. The benefit is to obtain

a bounded linear<sup>4</sup> increase in the project' effort as artifacts increase in complexity. TOGAF [7] is used because is a well known standard in the industry, notwithstanding any other standard can be used in our approach. This document is organized as follows. In Section 2 the concepts used are introduced and the proposed solution is explained. Afterwards, Section 3 discusses the results obtained with the application of this solution in a large EA project. Section 4 identifies, in the literature, the alternative proposals to integrate EA and ontology, and compares them with our proposal. Finally, Section 5 concludes the paper and identifies future work.

## 2 Generation of EA Artifacts founded on ontology

This section details the steps proposed in the systematic approach to generate, and maintain, the EA artifacts. It is founded on the data collected from multiple data sources, and on ontology to normalize, classify and relate that data. As depicted in Figure 1 the collected data relies on previous modelling effort, *e.g.*, business processes using BPMN or any type of available document about the organization' reality.

An Enterprise Model is a direct graph  $G_t = (A_t, R_t)$  being A a set of artifacts, R a set of relationships, and t is a discrete variable representing time [17]. A Viewpoint is "a specification of the conventions for constructing and using a view; a pattern or template from which to develop individual views by establishing the purposes and audience for a view and the techniques for its creation and analysis" [9]. An **ontology** is the representation of the essential understanding of a given domain, consisting is a body of knowledge that is recognized by all the stakeholders involved in a EA project, and also considered as the true source for all the artifacts' instantiations [3]. A **project** holds a set of architectural statements that change artefacts states after its successful completion [17].

We use TOGAF 9.2 [7] to illustrate the approach, however, any other representation language can be considered. TOGAF defines that a document is composed by multiple Artifacts [14], and an Artifact is composed by multiple Concepts (*cf.* depicted in Figure 2). A TOGAF concept is thus the finer grained element requiring common understanding between the stakeholders involved in the EA artifacts generation: suppliers, clients, architects, project managers, *etc.*. The meaning of each TOGAF concept need to be discussed prior to the artifacts' generation in order to align the deliveries expectations and to avoid misunderstandings or biased interpretations.

The solution encompasses six sequential steps, which are described in each one of the following subsections. Next section, the validation, describes how these steps were instantiated in four different EA projects.

<sup>&</sup>lt;sup>4</sup> bounded linear follows the definition from [2]: "normalized systems are information systems that are stable with respect to a defined set of anticipated changes, which requires that a bounded set of those changes results in a bounded amount of impacts to system primitives".

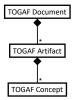


Fig. 2: Decomposition of a TOGAF Document into finer concepts.

#### 2.1 The EA document elicitation

The EA document elicitation step must comply to the organization requirements in order to capture correctly its *as-is* or *to-be* situation. By its turn, related literature refers that organization requirements heavily depend on the fit for purpose [15, 1] of the EA initiative. Therefore, whether the purpose is to create a new architecture, or planning the transition between architectures, or to decommission the current architecture, then different documents can be elicited. Figure 3 depicts an excerpt of the TOGAF documents that are recommended for each one of the TOGAF Architecture Development Method phase [7], where each phase concerns a specific purpose, which can be used as a baseline for any project.

Deliverable	Output from	Input to		
Architecture Building Blocks (see 32.2.1 Architecture Building Blocks)	F, H	A, B, C, D, E		
Architecture Contract (see 32.2.2 Architecture Contract)				
Architecture Definition Document	B. C. D. E. F	C. D. E. F. G. H		

Fig. 3: An excerpt of the prescribed documents by TOGAF organized by their location in the Architecture Development Method [7].

#### 2.2 The EA Artifacts elicitation

Considering TOGAF as the body of knowledge for EA artifacts, they are divided into: *diagrams*, *catalogs* and *matrices*, and then, classified by type of target architeture (Business, Data, Application, Technology, Preliminary, Vision, Requirements Management and Opportunities and Solutions). Knowing that each EA document contains multiple artifacts, and that the previous step 2.1 already elicited the desired document, the decision is straightforward if TOGAF template documents are used. Each TOGAF template document suggests what are the artifacts that should be included. Therefore, this is a stable source to identify which are the artifacts to be elicited for each document. If needed, any other artifact can be added. Furthermore, as long as consistency is guaranteed, an artifact can be placed in more than one document.

Alternatively, the fit for purpose principle can be further extended to elicit what are the exact artifacts that should be produced avoiding eventual biased effects of an EA standard. In this circumstances, each EA artifact should be given a specific and well-defined purpose.

## 2.3 The EA Concepts elicitation

This step is the most relevant in terms of EA generation, encompassing the capability to normalize and classify the data that is going to be extracted using an ontology definition. An ontology comprises a set of concepts, relationships and definitions that are stablished for each EA project and that are only valid within that scope. A first iteration to create the ontology is to identify all the concepts that are required in each one of the selected artifacts from step 2.2, and then, map each concept with a known data source. For instance, if the "Value Stream Catalog" from Business Architecture has been selected for generation, and there is an ArchiMate model available containing the "Value" and "Business Actor" elements, then both ArchiMate elements should be considered as key concepts in the ontology to generate that artifact. This is a straightforward approach, but if done iteratively, with validation milestones, has the possibility to converge for a robust ontology. If the project' stakeholders involved are connoisseurs of the domain, then other option is to bootstrap the ontology from that existing knowledge, requesting them to express all the key concepts and design the ontology from scratch. Either way, a table with concepts and definitions, could be produced with the content similar to that exemplified in Table 1. Relationships are inherited from the data source, e.g., the sequence flows existing between tasks of a BPMN model.

Table 1: An illustration of concepts definition.

The concept	The agreed concept definition	The data source where the	An optional alias
name		concept can be extracted	to other concept
Actor	The intervenient of a business process	Excel File	
Pool	The intervenient of a business process	BPMN file	Actor

#### 2.4 Concepts mapped with Artifacts and Documents

Table 2 exemplifies the mapping between concepts, artifacts and documents. Let us consider a universe of 3 concepts: X, Y and Z; 6 artifacts, namely: 3 diagrams, 1 catalog and 2 matrices; and 3 documents. Diagram 1 and 2 only use the concept X, while diagram 3 requires concepts Z and Y. Catalog 1 requires the concept Z. Matrix 1 requires the concepts Y and Z, while Matrix 2 requires the concepts Xand Y. Then, document 1 requires the following artifacts: diagram 3 and matrix 2. Document 2 requires the following artifacts: catalog 1 and matrix 1. Finally, document 3 includes all the available artifacts.

Table 2: Example of mapping the concepts, artifacts and documents.

		Diagrams			Catalogs					
Arti	facts	Diagram 1	Diagram 2	Dia	agram 3	Catalog 1	Μ	atrix 1	M	atrix 2
Cone	cepts	X	X	$\mathbf{Z}$	Y	Z	Y	Z	х	Y
Documents	Document1			•	•				•	•
	Document2					•	٠	•		
	Document3	•	•	•	•	•	•	•	•	•

#### 2.5 Concepts extraction from multiple data sources

After the previous steps consisting in the ontological construction, a software tool to consume data from the multiple data sources and to compute it, is preferrable than processing manually. Each data source needs to be normalized, we suggest adding annotations, where each piece of data is classified into a single concept, and then stored in a repository. Therefore, the classification part of this step is currently human dependent. AI techniques can be considered as future improvement for this step.

#### 2.6 Automatic documents generation

After provisioning all data annotated in the repository, each document is programmed with a definition of the artifacts required. By its turn, each artifact is constructed with the class of concept and relationships required at each position in a viewpoint. A language for representation need to be chosen here, *e.g.*, BPMN, ArchiMate, or any other; and the tool need to have visualization support for that language. Finally, the documents are automatically generated iteratively selecting the instances of each concept class from repository data, and filling them in the artifacts. This a very repeating task if done manually.

## 3 Validation

The systematic approach is validated using (i) explanations and (ii) results' discussion based on the execution of four EA consultancy projects as listed in Table 3. Each project is characterized by application domain, project' purpose, approximate time employed (for comparison consider that similar man/month allocation was used), and then a set of approximate metrics to understand the complexity involved, namely, the number of concepts found, the data sources, the number of documents produced and the number of artifacts produced.

**The EA document elicitation:** distinct elicitation actions were performed. The exact document list to produce in *Project 1* was defined since the initial specifications. In *Project 2* and 3 the documents were proposed (and accepted by the client) as an initiative of the consultancy team. While in *Project 4* the initial high-level specifications were used to specify a document list baseline that required adjustment during project execution time.

The EA Artifacts elicitation: is a finer definition when compared with documents. In *Project 1, 2* and 3 artifacts were elicited by the consultancy team at project kick-off, while *Project 4* defined them from specifications some time after project start. Again iterative approach is also possible in complex projects where the specification is not closed at project' beginning.

The EA Concepts elicitation: has been designed by the internal development team in all projects, using abstraction of the provided documentation and by organizing meetings to elicit the core concepts. Communication is of uptmost importance at this stage.

Project ID	Application Domain	Purpose of the EA generation	Approxima time employed	atQuantity of Con- cepts	Quantity of Data Sources and format	Quantity of Documents	Quantity of s Artifacts
1	Public administration	IT systems' portfolio to support financial services	6 months	25	5 in excel and BPMN	7	50
2	Public banking	IT services dependencies to increase its resilience	< 6 months	50	1 in workshop	2	9
3	Private banking	The application components portfolio to align business and technology	< 6 months	6	2 in workshop and word	2	5
4	Justice	IT technologies' portfolio to normalize usage in suborganizations	> 1 year	25	20 in excel	3	50

Table 3: Sample of projects where the systematic approach has been applied.

**Concepts mapped with Artifacts and Documents:** for a better visualization, a full example of mapping  $Concept \rightarrow Artifact \rightarrow Document$  that has been used in one of the projects is depicted in Figure 4. It is noticed that concepts are repeated in many artifacts increasing the need to be consistent in their usage and being able to trace the concept' changes in all artifacts. Moreover, the 3 types of artifacts (Diagram, catalogs and matrices) refer to the same concepts. It is also usual that some artifacts are specialized referring to a small set of artifacts, while other refer many of them (or even all of them).

Concepts extraction from multiple data sources and automatic documents generation: concepts from *project 1* were extracted from BPMN sources, and excel files containing BPMN annotations. *Project 2* and 4 used excel files as data sources. In *Project 3*, due to the fact that the organization is in an initial maturity phase, the data source extraction was skipped and the available data was injected directly in the concepts. For all projects, documents were produced using the ATLAS<sup>5</sup> tool. The artifacts (to be included in the documents) were produced using the notation of ArchiMate and BPMN for the diagrams, and excel files for the catalogs and matrices.

Concerning the RQ posed in introduction, we propose to generate, and maintain, the EA using a core ontology of  $Concept \rightarrow Artifact \rightarrow Document$  that requires alignment between the project specifications, the organization' client expectations and the consultancy team. This alignment is produced by a matrix defined with the client' involvement as examplified in Figure 4, and then programmed once in the ATLAS tool. Whenever the data sources change the documentation is recreate again. Therefore, ensuring consistency between all artifacts.

Some questions still remain in this alignment process: Which documents are relevant? Considering some selected documents, what are the artifacts that are

<sup>&</sup>lt;sup>5</sup> Available for consultation at https://atlas.linkconsulting.com.

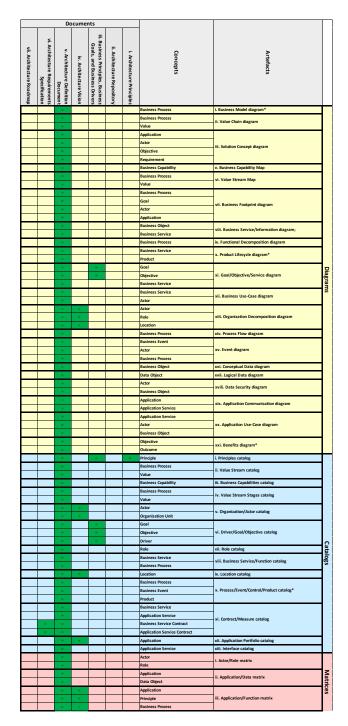


Fig. 4: Example of mapping concepts with artifacts and documents to a given domain.

recommended? Are they enough, or are others needed? Considering the available project' concepts, which one can be used within each artifact? Furthermore, we noticed that keeping the ontology as simple as possible, with many alias, is easier for human understanding and discussion. Visualization has also been identified as a positive asset to understand the ontology.

# 4 Related Work

This section reviews the knowledge available in the literature regarding EA and ontology topics. To that end, a search has been conducted using Web of Science, Google Scholar, IEEE and ACM databases, considering the following topic searches: "enterprise architecture" AND "ontology", until 2022 (included). The returned number of hits for each database is  $N = \{170, 33, 93, 115\}$ , with a total number of 378 papers, where 2 are technical reports, 13 are books, 1 are chapters in books, 288 are conference papers, 72 are journal articles, and 2 are miscelaneous. The yearly distribution reveals a slight decrease in number of publications in recent years when compared with the period: 2007...2018. Yet, many contributions are noticed by this search. The following procedure for the literature review has been used. Firstly, the references are collected from each database, then the duplicates are excluded. Secondly, this selection of papers are screened by title and abstract, where the most referred terms are used to score each paper in a scale [0...1]. The considered inclusion score for title is the top 30%and for abstract is the top 5%. A final collection of 48 papers is considered to a fine-grained analysis. The most relevant and referred concepts, by descending order, in the 48 abstracts are: "enterprise; architecture; models; information; EA; business; management; Ontology; ontologies; process; support; order; framework; modeling; elements; development; role; representation; environment; integration; issues; case; study; knowledge; language; context; concepts; ontological; problem; implementation; holistic; system; tool; maintain; view; engineering; technologies; description; formal; part; communication; components; goal; data; functions; specific: application: benefits: consistent: integrated: domain: sources: comprehensive; evaluate; artifacts; rules; changes; terms; ontology-based; challenges". While the concerns related with architecture exist in this set, also the ontologies, languages, concepts, are present; and the aspects of integrating its concepts along with the consistency. Table 4 organizes the selected papers by interest  $\operatorname{core}^6$ . The

Table 4: Distribution of papers by interest core classification.

Interest core	Number of papers found
Enterprise Architecture	43
Ontology	20
Enterprise Ontology	10
Enterprise Engineering	4
Knowledge Engineering	1

most referred interest core is located in EA, while a smaller amount is located

<sup>&</sup>lt;sup>6</sup> The full citations are not included in this paper due to length limitations

in Ontology. 10 papers uses a specific ontology: the enterprise ontology. Enterprise and Knowledge Engineering have a minority interest. However, Enterprise Engineering is represented in this search with books, which is not directly comparable with papers. Similarly, with our systematic approach, 18 papers are both interest in EA and ontologies, which is a strong indicator about the community interest in this integration. A large diversity of domain are identified in the literature, what corroborates our position that each specific project requires an own ontology: Services computing, Security and/or Privacy, Goal modelling, Repositories, Access management, Concern modelling, Systems' Integration, Government, Semantic Architecture, Libraries, Semantic web, Decision making, and Governance. A different approach than ours, is presented in [11, 12] that propose an integration between an ontology of the business terms with an ontology for EA components and EA relationships, aiming a common understanding between humans and systems to support integrations in enterprises and collaborations between enterprises. However, no practical applications were attempted. [13, 18] propose ontologies to extract knowledge from EA models, what is the opposite that we are proposing. Similarly with our proposal, [5, 14, 16, 10] present approaches to generate EA using an EO ontology: DEMO [3] or OWL. However, prior knowledge to these ontologies is required, and few validation in industry exist. The common understanding between stakeholders, as referred as an hard requirement in our approach is also corroborated by [4] that exemplifies how to construct it using a wiki tool.

# 5 Conclusion and future work

The uncontrolled effort that is required to manually construct artifacts in complex EA projects triggered the need for this research. It is impracticable to start an EA project without knowing the exact effort that is required. Using a metaphor, it is the same as designing a building without knowing the associated cost growth factors. An initial review of the related literature showed us that no full solution is available for this problem. Therefore, we researched an approach and tested it in practice, with four EA complex consultancy projects, to show its usefulness. As reported, our systematic approach evidences a linear, and thus controlled, increase in effort as TOGAF artifacts increase in complexity. The positive impact of our approach applied to a specific domain is given by the following aspects: (i) the network of dependencies between architectural concepts is explicitly presented to all project' stakeholders, (ii) the unique identification of each concept instance allows navigation between them, and thus, the traceability of the dependencies between concepts is easier, and *(iii)* the alignment between architectural layers can generate complementary views of the same model, and thus achieve different documentation purposes.

As future work, we identify the following threads: application of the approach to more case studies or industrial projects to further explore its social and technical implications, and automate the extraction of concepts types from the multiple data sources using AI techniques assisted by human classifications.

# Acknowledgement

This work was supported by a pre-registered project, named as eProcess, which is under national funds with reference C669314338-00003137 (LINK CONSULTING – TECNOLOGIAS DE INFORMAÇÃO S.A.).

## References

- 1. Currie, A.: From models-as-fictions to models-as-tools (2017)
- De Bruyn, P., Mannaert, H., Verelst, J., Huysmans, P.: Enabling normalized systems in practice–exploring a modeling approach. Business & information systems engineering 60, 55–67 (2018)
- 3. Dietz, J.L., Mulder, H.B.: Enterprise ontology: a human-centric approach to understanding the essence of organisation. Springer Nature (2020)
- 4. Fuchs-Kittowski, F., Faust, D.: The semantic architecture tool (semat) for collaborative enterprise architecture development. In: Briggs, R., Antunes, P., DeVreede, G., Read, A. (eds.) GROUPWARE: DESIGN, IMPLEMENTATION, AND USE. Lecture Notes in Computer Science, vol. 5411, pp. 151–163 (2008)
- Gomes, B., Vasconcelos, A., Sousa, P.: An enterprise ontology approach for defining enterprise information architecture. In: PROCEEDINGS OF THE 5TH ECIME. pp. 609–615 (2011)
- Greefhorst, D., Proper, E.: Architecture Principles. Theory and Practice. Springer (2010)
- 7. Group, O.: Togaf(R) standard, version 9.2 (2023), retrieved date: 2023-04-18
- 8. Guerreiro, S.: Conceptualizing on dynamically stable business processes operation: a literature review on existing concepts. BPMJ **27**(1), 24–54 (2021)
- 9. ISO/IEC/IEEE: Systems and software engineering architecture description 42010:2011 (2011), retrieved date: 2023-04-19
- Jabloun, M., Sayeb, Y., Ben Ghezala, H.: Enterprise ontology oriented competence: a support for enterprise architecture. In: 2013 3RD ISKO-MAGHREB. ISKO, IEEE, 345 E 47TH ST, NEW YORK, NY 10017 USA (2013)
- Kang, D., Lee, J., Choi, S., Kim, K.: An ontology-based enterprise architecture. EXPERT SYSTEMS WITH APPLICATIONS 37(2), 1456–1464 (MAR 2010)
- Minaei, B.: Enterprise Architecture Development Based on Enterprise Ontology. academia.edu (2013), query date: 2022-11-03 10:52:59
- Montecchiari, D.: Ontology-based validation of enterprise architecture principles in enterprise models. BIR Workshops (2021), http://ceur-ws.org/Vol-2991/ paper17.pdf, query date: 2022-11-03 10:52:59
- 14. Nogueira Santos, F.J., Santoro, F.M., Cappelli, C.: Addressing crosscutting concerns in enterprise architecture. In: IEEE 18th EDOC. pp. 254–263 (2014)
- Parker, W.S.: Model evaluation: An adequacy-for-purpose view. Philosophy of Science 87(3), 457–477 (2020)
- Silva, N., Mira da Silva, M., Sousa, P.M.M.V.A.d.: Modelling the evolution of enterprise architectures using ontologies. In: 2017 IEEE 19th CBI. vol. 01, pp. 79–88 (2017)
- 17. Sousa, P., Vasconcelos, A.: Enterprise Architecture and Cartography: From Practice to Theory; From Representation to Design. Springer Nature (2022)
- Sunkle, S., Kulkarni, V., Roychoudhury, S.: Analyzing enterprise models using enterprise architecture-based ontology. In: Moreira, A., Schatz, B., Gray, J., Vallecillo, A., Clarke, P. (eds.) MODEL-DRIVEN ENGINEERING LANGUAGES AND SYSTEMS. Lecture Notes in Computer Science, vol. 8107, pp. 622–638 (2013)