An Information Architecture for the Public Administration

Ricardo Castelão

Orientador:
Professor Doutor André Vasconcelos

Co-Orientador:
Professor Doutor José Tríbolet

Acompanhante:
Dra. Maria João Marques

1INESC Inovação, Rua Alves Redol 9, 1000-029 Lisboa, Portugal, {ricardo.castelao, andre.vasconcelos, jose.tribolet}@ist.utl.pt
2AMA, L.P, Rua Abranches Ferrão 10 3ºG, 1600-001 Lisboa, Portugal, maria.marques@ama.pt

Abstract. In today's society, the role played by information is growing in significance as the everyday actions executed by people, and inside the organizations, rest on its manipulation. Aware of this reality, enterprises invest large amounts of their budget in information technology in order to shape up to their needs, which if not properly planned and executed, can result in incompatibilities between different information systems on the same organization. This work proposes a reference information architecture for the Portuguese public administration, and the methodology used to develop it. The development of the architecture was achieved by analyzing multiple contexts and their specific information architectures, following the different steps within the proposed methodology, that allowed to gather the set of information entities and respective attributes transversal to the public administration. As the Portuguese public administration presents as a vast context, this work is focused on the citizen's lifecycle perspective, it's main activities, and different intervenient in the execution of the main processes, specifically the act of being born and dying. The output of the work is a revised version of the reference information architecture, modeled with the CEO framework, and a methodology for gathering a set of transversal information entities from the analysis of specific contexts, that we used for developing this work..

Keywords: Architecture, Citizen, Entity, Information, Methodology, Public Administration
1 Introduction

Nowadays, enterprises are becoming increasingly information based and the improvement in their information activities is a priority among their set of objectives [1]. The activities and processes, executed in the context of an organization, majorly depend on the use, creation, sharing and exchange of information, shifting the costs associated with these practices into high values unless there is some work developed to counteract this. In the light of these occurrences, information technologies are an important tool to achieve organizational goals, and at present constitute a considerable portion of organizational expenditure [2]. Justify the investments in information technologies, however, is not something that all organizations are capable of doing, and therefore, management researchers, focus on understanding the gap existing between the desired and the realized level of objectives in this area.

A key to achieve these objectives and improvements is developing an Information Architecture [1]. A great number of organizations have in place an ad hoc information architecture, which does not comply with what the organization will need in the future because it has an incoherent framework, unnecessary duplications, incompatibilities, missing elements and lack of standards.

Given these problems and scenarios, it is clear that a globalized concern about information should be rapidly emerging in order for the enterprises to shape up with its current needs.

Through the next sections we will address these problems, in the context of the Portuguese public administration, as well as provide the first glance towards a methodology to achieve the desired solution.

2 Problem and Contributions

Considering all the issues identified previously, nowadays, different organisms of the Portuguese public administration have in place different information systems, which creates incompatibilities when a need to transfer data among different locations arises. This situation also creates problems when we need to find where a set of information is stored, or which entities have access and manipulate them. It is by developing an information architecture that we can perceive which are the relevant, and transversal, information entities to the public administration, as well as the relations among them.

After clarifying the context of the problem, we take place by defining the set of objectives and contributions associated with this work, as well as the main question raised. Therefore, this work will aim at answering to the main question Q1 and the sub-question Q2:

**Q1:** What are the common Information Entities through the Portuguese public administration, within the context of the citizen’s lifecycle?

Our contribution in answering this question is the final result of our work. With the work developed, we were able to review the existent version of the architecture, and provide a new one, already incorporating the knowledge gained
during all of our research, therefore containing the set of information entities that we concluded as transversal to the Portuguese public administration.

**Q2: How can we elaborate a global information architecture from specific data models?**

We targeted this question while researching the field of related work. During our investigation, we further researched on different frameworks and enterprise modeling languages that, not only allowed us to conclude our work, but also put us one step closer towards having a solid knowledge base that allows us to answer this question. As the data models may vary between different types of work, and also the specific contexts as well, each work should assess their particular problems and objectives. Concretely, during our work, we found the Spewak's EAP very useful in addressing this problematic, hence having integrate it in our section of related work. The knowledge contained on that section constitutes a working base for each future work to pick on, and then evolve within its context.

### 3 Related Work

After having defined the objectives and contributions of this work, we skim the academic areas of interest that relate to the domain of the thesis. Here we present the main concepts revolving this subject as well as the current work being developed around it.

Although an extended version of the related work can be found in the complete version of our work, here we present the main languages, frameworks and concepts, that constitute the center of it.

#### 3.1 Enterprise Architecture Planning

Developed by S. Spewak [3] the Enterprise Architecture Planning (EAP) uses and complements the Zachman Framework, focusing on the two top levels of it. Spewak considers that these top two levels are directly connected to what the enterprise is, instead of the bottom four which are more concerned with the design, development, and implementation phases of the enterprise architecture planning.

The EAP methodology aims at defining the enterprise architecture for the top two perspectives of the business model and objectives, i.e., define and plan the top two levels of the Zachman framework [4].

As in the Zachman framework, the EAP is strongly data-driven and business-driven, because information is the basis of business operation, therefore being of major importance to it [3]. With this, the EAP defines that the information model and information used by business, in order to run the business, must be defined first, derived from the business context and strategy, which ultimately results in the fact that the enterprise architecture definition should follow a strict sequence of steps, comprehending at first the data architecture, then the application architecture defined from it, and at last the definition of the technology architecture from the application architecture. The purpose of all this actions is to ensure the alignment between the
different architectures and the business, essential factor for obtaining strategic alignment. The Spewak’s EAP presents as a four-step process illustrated on figure 1 [3]:

![Spewak EAP Methodology](image)

**Figure 1:** Spewak EAP Methodology

### 3.2 CEO Framework

The CEO framework was purposed by the Organizational Engineering Center (CEO, in Portuguese) as a standard UML profile for describing, linking and tracing organizational concepts at multiple levels, allowing a common language to be used not only in business, but in software domain as well [5]. These levels are subdivided into strategy and goals, business processes and information systems. By using UML, it is used a common representation language in both business and software domain.

In terms of the relation with this work, the CEO framework comprehends the concept of information architecture, as well as the explicit concept of information entity, under the level of the information system architecture (fig. 2), which are two major advantages considering the objectives of the work [6].

![Simplified Meta-model of the FCEO for Information Systems Architecture](image)

**Figure 2:** Simplified Meta-model of the FCEO for Information Systems Architecture
3.3 Information Architecture

There is no clear distinction between data and information. Information can be perceived as data which is useful to someone at a specific time, being simply data outside that time frame. The utility itself is relative and depends on the context and purpose to which the data is being applied. The role of the information system in this matter is well defined, and resides on processing the data so that its utilization can be maximized.

An information architecture can be viewed as a structured set of multidimensional interrelated elements that support all information processes. It is also an effective framework for acquiring, organizing, and prioritizing a wide range of technological knowledge, facilitating the ability to effectively and appropriately apply it. It provides the framework for planning and implementing a rich, standards-based, digital information infrastructure with well-integrated services and activities [1]. Additionally, it serves the purpose of defining which information entities are necessary and how they relate to each other.

**Information Entity.** An information entity can be perceived as any concept e.g. people, place or object, which has meaning to the business context and about which is relevant for the organization to store information about.

An information entity is characterized for having a name, a unique identifier, a description and its relation to processes, other entities and information systems [7].

**Benefits of Information Architectures.** Enterprises are becoming increasingly information based, creating an urge to improve their information activities so that they can assure continuous competitiveness. Also, organizations suffer from a series of problems that affect the way they conduct business. Among them is the non-existence of an organization or process for creating and sustaining an information architecture, the lack of standards, and the fact that changes in technology occur faster than the ability of an organization, to bring new services and products to their workspace. Due to these facts, developing an enterprise information architecture is a main concern, as it helps to address this concerns.

The benefits that arose from an information architecture relate to strategic advantages, increase effectiveness, and reduce costs, particularly [1][8]:

- Easier information sharing and exchange.
- Improved security and privacy.
- More effective response to customer requirements through easier and faster building of information services.
- Increasingly effective matrix organization structure because of the use of common information services, resources, and tools.
- Promote data sharing.
- Easier incorporation of outside vendors within chains of needed capabilities and better integration with the rest of the departments, academic community, and industry.
- Reduced cost by eliminating data redundancies.
- Reduce time spent on future data collection projects.
4 Proposed Methodology for Developing the Architecture

Facing the problem identified on chapter 2, considering the existing modeling languages, frameworks, methodologies - particularly the Spewak's Enterprise Architecture Planning, and the evaluation methodology to be applied, we devised a coherent sequence of steps/phases that allowed for the obtaining of a solution, and are an adaptation to our problem from Spewak's EAP methodology. A overall viewpoint of the methodology can be viewed on figure 3, and a more detailed activity diagram in Appendix I.

![Figure 3: Proposed Methodology](image)

The development of the solution started with a step pre-dating the one's represented on figure 3, aligned with the first step of the EAP, from which resulted this work. As defined by the EAP, and described through this work, we proceeded to definition of the scope, objectives, contributions, and project plan, in order to achieve the desired solution.

Maintaining the alignment with the EAP methodology, shifting to its second step, in Context we aimed at defining which areas of the citizen's lifecycle we would want to target and proceeded to the analysis of the different set of contexts relevant for our work. This included not only different departments of the public administration, which represent an important role during the life of a citizen and with whom he interacts, but also two projects directly related to the processes of being born and dying. The contexts and projects that served as basis for this work are enumerated as follows:

Administração Fiscal, Identificação Civil, Saúde, Segurança Social, Cartão de Cidadão, Information Architecture developed by AMA, HL7, Project Nascer Cidadão, Project SICO

For each of these contexts/projects we proceeded to its representation using the CEO Framework from a set of gathered data, e.g., specific data models, project documentation, and interviews with enrolled people, and aimed at representing the context-specific information entities while focusing on the interaction with the citizen.

Next, in Analysis, still aligned with the second step of the EAP methodology, we reunited and cross-reference the information gathered from the previous step and, using XML, proceeded to the representation of the entities gathered and their attributes, in a structured manner. Also, after further gathering information about the two projects under analysis, we proceeded to their representation using BPMN. This ended the definition of the As Is, and start the processes of definition of the To Be.
Each of the specific contexts was represented using a XML tree, and for each context-specific information entity, a new node was included in that tree. The same procedure was followed to deal with the information entities' attributes, this time by creating elements for each corresponding node of the XML tree.

In Solution, aligned with the third step of the EAP, we elaborated the information architecture based on the work developed here and until here. In order to do so, we took all the knowledge gathered from the previous phases and analyzed the relations between the different topics investigated.

While the results are presented on the next section, we started by picking up each context's XSD and match it against the one correspondent to the actual revision of the information architecture. To do so, we introduced the use of the mapping tools available in BizTalk, and mapped as much attributes as possible between the two sides, using the application and the XML provided by AMA, which originate on the Portuguese Interoperability Platform.

Taken that we were able to model two processes that represented precisely the beginning and ending of a citizen's lifecycle, and with limited information and resources, we tried to identify the relation between the activities in the processes and the information entities in the context.

Following this work, another analysis phase took place and at the end of it we were able to identify what were the most recurrent attributes to not have a mapping with the current version of the architecture. In order to do so, we picked on the results of the mappings established, and identified which attributes did not have a correspondence with the architecture proposed by AMA. We devised a set of statistics, were we identified how many times a certain attribute appeared as unmapped and on how many contexts, the global number of attributes and entities analyzed, and the ratios between the attributes analyzed and the ones categorized as unmapped. By considering these two key factors, we chose a set of attributes viable for inclusion on our version of the architecture. After having defined the set of attributes to include in the new version of the architecture, we applied traceability on the attributes, used the XML to define which information entities were affected by their inclusion, and developed a new representation of the architecture using the framework CEO.

Finally, in Conclusion, we will evaluate the work done and make considerations about the future work to be developed. As the definition of the migration and implementation plans towards the solution is out of the scope of our work, the fourth step of the EAP will not be concluded in this work.

5 Proposed Information Architecture

Through this chapter we provide the results obtained with the concretization and execution of the different phases of our proposed methodology.

As comprehended within our methodology, we started by representing each of the context-specific information entities using the CEO framework. As we progressed through the different steps, we developed an equivalent representation using XML, where each context was represented under a XML tree, each information entity as a
node, and each entity's attribute as an element. Having the representation of the As-Is, we proceeded to the mappings stage where we mapped each attribute from the specific contexts against the ones existent on the AMA's proposed version of the architecture. From here, we were able to gather a set of data that allowed us to identify the unmapped attributes that occurred more often and with greater dispersion over the different contexts (fig.4), as well as a set of overall statistics (fig.5).

<table>
<thead>
<tr>
<th>Unmapped Attributes</th>
<th># Occurrences</th>
<th>Contexts</th>
<th># Contexts Present</th>
</tr>
</thead>
<tbody>
<tr>
<td>NumeroProcesso</td>
<td>30</td>
<td>X</td>
<td>4</td>
</tr>
<tr>
<td>Prioridade</td>
<td>21</td>
<td>X</td>
<td>4</td>
</tr>
<tr>
<td>EstadoProcesso</td>
<td>12</td>
<td>X</td>
<td>4</td>
</tr>
<tr>
<td>PretendeInscrição</td>
<td>8</td>
<td>X</td>
<td>3</td>
</tr>
<tr>
<td>GUID</td>
<td>8</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>FotoCidadao</td>
<td>7</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>CodigoRepartiacao</td>
<td>6</td>
<td>X</td>
<td>2</td>
</tr>
<tr>
<td>DataValidade</td>
<td>6</td>
<td>X</td>
<td>2</td>
</tr>
<tr>
<td>AnotacoesJustica</td>
<td>6</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>Inscrito</td>
<td>5</td>
<td>X</td>
<td>3</td>
</tr>
</tbody>
</table>

Figure 4: Results from attribute's mappings

| # Overall Unmapped Attributes | 256                   |
| # Overall Attributes          | 546                   |
| ≈ Overall % Unmapped Attributes | 46,886,446,89 |
| ≈ Overall % Mapped Attributes  | 53,113,553,11 |

Figure 5: Overall statistics

Having identified the top unmapped attributes, and considering the data gathered, we were in position of suggesting which attributes should be added to the architecture (fig.6).
<table>
<thead>
<tr>
<th>Attribute</th>
<th>Type of Modification</th>
<th>Destination Information Entity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numero Processo</td>
<td>Add attribute</td>
<td>Evento Contacto</td>
</tr>
<tr>
<td>Prioridade</td>
<td>Add attribute</td>
<td>Evento Contacto</td>
</tr>
<tr>
<td>Estado Processo</td>
<td>Add attribute</td>
<td>Evento Contacto</td>
</tr>
<tr>
<td>Pretende Inscricao</td>
<td>Add attribute</td>
<td>Cidadao</td>
</tr>
<tr>
<td>Inscrito</td>
<td>Add attribute</td>
<td>Cidadao</td>
</tr>
<tr>
<td>Código Deficiencia</td>
<td>Add attribute</td>
<td>Pessoa</td>
</tr>
<tr>
<td>Indicador Obito</td>
<td>Add attribute</td>
<td>Pessoa</td>
</tr>
<tr>
<td>Data Falecimento</td>
<td>Add attribute</td>
<td>Pessoa</td>
</tr>
<tr>
<td>Indicador Ordem Gravidez</td>
<td>Add attribute</td>
<td>Pessoa</td>
</tr>
<tr>
<td>Multipla</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 6: Final set of attributes to be added

After completing the analysis of the entire set of contexts, the practical case (which is not related in this article), and the processing of the emergent data from the instantiation of an entire cycle of our proposed methodology, we developed a new version of the information architecture, incorporating the modifications resultant from our entire work. The modifications applied to the AMA's proposed version of the architecture are at the level of the information entities' attributes, targeting the ones that most of all relate to the citizen's lifecycle, and can be viewed in the representation present in Appendix II, highlighted in green.

6 Conclusions

The results obtained reflect the changes that, based on our research and analysis, improve the alignment of the reference architecture with each of the contexts analyzed. Given the type of work performed, the granularity of the changes and improvements proposed is aligned with our objectives.

By using the knowledge gathered under the initial phases, and by applying it into the further development of the work, we were able to analyze and process the different data gathered and ultimately provide a revised version of the existent architecture, with improvements based on the entire work we developed.

If we focus our main contributions on the proposed architecture and methodology, the representation chosen for the architecture, the work developed during the context and analysis stages, the mappings, and the practical case and CRUD matrixes, qualify as accessory contributions to the field targeted by this work.

Using as base for our analysis the statistics produced during our step of context analysis, and comparing them to a set of new ones produced using as basis our proposed information architecture, we see that our proposal demonstrates an alignment improvement of 13.6%, while still having margin for improvement (fig.7).
### Statistics from AMA's Proposed Architecture

<table>
<thead>
<tr>
<th>Statistics from AMA's Proposed Architecture</th>
<th>Statistics from our Proposed Architecture</th>
<th>Alignment Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td># Overall Unmapped Attributes</td>
<td>256</td>
<td>182</td>
</tr>
<tr>
<td># Overall Attributes</td>
<td>546</td>
<td>546</td>
</tr>
<tr>
<td>Overall % Unmapped Attributes</td>
<td>46.9</td>
<td>33.3</td>
</tr>
<tr>
<td>Overall % Mapped Attributes</td>
<td>53.1</td>
<td>66.7</td>
</tr>
</tbody>
</table>

**Figure 7: Impact of proposed changes to the architecture**

From a global perspective it is our opinion that considering the objectives of the work, the means at our disposal, and the time frame available, we positively contributed to improve the existent version of the information architecture for the public administration, while also performing work that will enrich the academic universe, and can serve as basis for future research on similar subjects.

### References

Appendix I: Activity Diagram for the Proposed Methodology
Appendix II: Representation of the Proposed Architecture