Information Reference Architecture for the Portuguese Health Sector

Tiago José Gonçalves Brás
tiago.g.bras@ist.utl.pt
Técnico Lisboa – October 2015

Abstract — The need to improve the information activities has been proved to be fundamental for the competitivity of all organizations. One of the identified ways to reach that goal is the creation of Information Reference Architecture (IRA) that enable the guidance and restriction of the instantiations of a set of individual architectures and solutions.

In this work, we suggest a method of develop reference IA able to ensure an easy maintenance and the semantic interoperability. Such method follows a bottom-up approach and involves both the mapping of the main information systems in a certain sector of activity, and the use of reverse engineering, models enhancements and integration models techniques. The main contribution of the method resides in the model integration step, where a process enabling the identification of correspondent informational entities in different models, created following a set of similarity rules, is stressed.

This methodology is instanced in the specific Portuguese Public Health Sector, culminating in the development of a reference IA for the sector.

We used the Design Science Research Methodology to conduct our research. The method suggested in this work and the corresponding instantiation to the Portuguese Public Health Sector are assessed with the help of metric systems of IA and models.

Keywords — *Easy Maintenance, Informational Entities, Information Reference Architecture, Interoperability, Model Integration*

1. Introduction

In an ever more global and competitive world, organizations have become very dependent on information to implement their activities. Therefore, improving the informational activities is fundamental to ensure a continued competitive edge and to increase their global output [1].

The achievement of these improvements is based greatly on the development of Information Architectures (IA), that facilitate the sharing and exchanging of information, or on the easy and fast development of information services [1].

Applying this concern to the different sectors of activity, the difficulty in assuring continuous competitiveness is even higher, as it is necessary to take into account the huge amount of organizations and Information Systems (IS) that integrate a given sector. Assuring interoperability among all the organizations and IS is, quite often, dependent on the existence of an Information Reference Architecture (IRA) guiding and restricting the instantiations of a group of architectures and individual solutions and easing the maintenance of the different IS [2].

This work deals with the theme of IRA and suggests a method that, applying a bottom-up approach and starting with a group of IS, allows us to get to an IRA that can assure the maintenance efficiency and the semantic interoperability.

As far as the public Portuguese Administration is concerned (PA), there is only an IRA [3] for the whole PA. Different sectors of activity integrating the Portuguese PA, as it is the case of the public health sector, have no IRA. That is the reason for the many semantic incompatibilities existing among the IS in the Portuguese Public Health Sector that can, not only complicate or even make communication among them impracticable, but also affect their connection with the IRA of the PA in general. Furthermore, these incompatibilities are responsible for increasing the difficulty in keeping the IS involved.

After having developed the method suggested in this investigation, we proceed with its instantiation and validation in the concrete case of the Portuguese Public Health Sector.

The structure of this investigation follows the principles and guidelines that the Design Science Research Methodology (DSRM) [4] indicates.

2. Problem

2.1 Problem Identification and Motivation

Following the frame presented in the introduction, there appears the evidence and motivation for the work:

- There is no IRA for the Portuguese Public Health Sector.

Generalizing this motivation, and having in mind that the present work can be applied to another sector of activity, the problem to which this investigation intends to find an answer is the following:

- **There is not a well-defined method for the creation of an IRA that follows a bottom- up approach and that assures an easy maintenance and semantic interoperability.**

2.2 Research Questions

The main research question for the investigation that guides this document is:

- Which method to follow for the development of an IRA that can assure an easy maintenance and the semantic interoperability?

In order to be able to answer this question, there appear three other questions without which it is impossible to correctly orient this work:

- How are we going to compare the informational entities of different models?
- Which IRA will be adequate for the Portuguese Public Health Sector?
- Which informational entities ought to be present in the IRA for the Portuguese Public Health Sector?

In possession of the answers to the above research questions, we can, then, correctly address the problem identified for this work and develop a method that allows us to create an IRA easy to maintain.

3. Related Work

In this chapter the most relevant concepts of the work related to the topic under consideration that can contribute to the resolution of the previously identified problem are analyzed.

3.1 Enterprise Architecture

Enterprise Architecture (EA) is a set of principles, methods and models that are used in the design and implementation of the organizational structure of a company, its business processes, IS and infrastructure [5]. An enterprise architecture consists of several layers [6]. As far as the development of this work is concerned, we will focus on the informational layer of an enterprise architecture, since this layer is responsible for creating and managing IA.

3.2 Information Architecture

IA describes the principles and guidelines that allow a consistent implementation of information technology solutions [6]. Furthermore, it aims to identify and define the main types of data that support the organization business development [7].

IA brings many advantages, such as [1]: to facilitate the sharing and exchanging of information; to improve security and privacy; to respond more rapidly and effectively to customer demands; to facilitate the integration of systems, processes, data and information; to increase the business understanding.

An IA is composed by Informational Entities (IE), characterized by their attributes and relationships that the different IE establish among themselves [8].

3.3 International Reference Architecture and Document produced in the Context of the Portuguese Public Health Sector

Although the survey is not restricted to a specific sector of activity, an instantiation to the Portuguese Public Health Sector will be done in order to tackle the motivation identified in section 2.1. That's why it is useful to have in mind the work carried out in a similar context.

3.3.1 HL7 RIM

The Reference Information Model (RIM) [9] is the fundamental support for the development of version 3 of the HL7. This model aims to represent all clinical information from HL7, and can be seen as a reference model for the Health Sector.

3.3.2 Document Health Operational Programme 'Saúde XXI'

In the context of the Operational Programme for Health 'Saúde XXI', the Organizational Engineering Centre developed a document [10] where the most relevant concepts for the context of health are identified.

3.4 Interoperability, Reverse Engineering, Model Enhancement and Model Integration

3.4.1 Interoperability

The concept of interoperability refers to the ability of different heterogeneous applications to share processes and data in distinct systems or platforms [11]. To make this communication possible, it is necessary to take into account the existence of different types of interoperability - organizational, semantic, technical [12].

Regarding the specificity of this work, we highlight the semantic interoperability, since it is the one that occurs in the informational layer of an enterprise architecture and is related to the way different IA are represented.

Among the various ways identified to improve that interoperability, the reverse engineering, the model enhancements and the model integration acquire great importance [13] [14].

3.4.2 Reverse Engineering

According to [15], reverse engineering is the process of analyzing a particular system in order to identify its components and the relationships between these components, in order to create representations of the system in another form or at a higher level of abstraction.

In [16], a set of transformations used in reverse engineering at the database levels is presented.

3.4.3 *Model Enhancements*

According to [17], model enhancements help models to become more correct, simple and understandable. Besides using good modeling practices, in [18] and [19] some of the most common transformations in model enhancements are defined.

3.4.4 *Model Integration*

According to [20], model integration consists of building a global view, taking as its starting point a set of models developed independently. Since the development of such models occurs in an independent way, the structures and terminologies presented are often very different amongst themselves.

Therefore, it is necessary to understand how we can combine these differences in order to create a single and coherent model.

In [17] the following steps in model integration are identified: pre integration, models comparison, conforming models and models unification.

3.4.4.1 *Pre Integration*

At this step it is necessary to analyze the models we wish to compare, and choose the best strategy to follow regarding the integration. It is the time to take decisions, namely the ones related to the quantity of models we want to compare. We can compare two or various models simultaneously.

3.4.4.2 *Model Comparison*

At this step the models are compared in order to determine correspondences between concepts and identify possible conflicts. Thus, it is necessary to identify inter-model relationships. This activity is known as correspondence between models.

3.4.4.2.1 CORRESPONDENCE BETWEEN MODELS

It is possible to define the correspondence between models as follows [21]: Given two models M1 and M2, for each concept identified in M1 we have to find another one in M2 that is semantically similar to the one identified in M1.

3.4.4.2.2 SIMILARITY RULES

According to [22], there is a set of rules that can be used to verify the similarity between IE of different models. The rules that we consider relevant to this work are:

- **Rule 1** – If two IE have the same Uniform Resource Identifier (URI), they are equal;
- **Rule 2** – If two IE have the same instances, they are equal;
- **Rule 3** – If the description of two IE is similar, it is likely that the two IE are also similar;
- **Rule 4** – If the attributes of two IE are equal, it is likely that the two IE are also equal;
- **Rule 5** – If the name of two IE is the same or similar, it is likely that the two IE are equal or similar.
- **Rule 6** – If the hierarchical path to the IE is equal, the compared IE are similar;
- **Rule 7** – If the super entities are the same, the compared IE are similar;
- **Rule 8** – If the sub entities are the same, the compared IE are similar;
- **Rule 9** – If two IE have equal “sisters IE”, the compared IE are similar;
- **Rule 10** – If two instances have the same “mother IE”, they are similar.
- **Rule 11** – IE that have equal instance quantity are similar.
- **Rule 12** – If two instances are connected to another instance through the same property, they are similar amongst themselves.

In order to get to a correct correspondence we must follow a process which uses several of the rules above mentioned. In [22] it is suggested a correspondence process between models, where the above identified similarity rules are used.

3.4.4.3 *Conforming Models*

At this step the identified conflicts are solved, allowing the unification of the models. In order to achieve this, we can use several techniques such as type transformation, restructuring, renaming, etc. [18]. In situations where it is not possible to solve the conflicts arising from basic inconsistencies, someone responsible for the models in question should guide the designer in the resolution of conflicts [17].

3.4.4.4 *Model Unification*

It is the phase where the unification of the models compared previously is made, resulting in a new model. In [18] and [19] the most common transformations are described.

4. OBJETIVES AND SOLUTION PROPOSAL

After having identified the problem and the research questions, as well as having analyzed the related work, we are now able to define the solution goals. The main objective of this work is:

- Method creation to develop an IRA using a bottom-up approach that can assure an easy maintenance and a semantic interoperability.

In addition, we also want to achieve the following objectives:

- To identify correspondent IE in different models;
- To develop a IRA for the Portuguese Public Health Sector;
- To identify the most important IEs for the Portuguese Public Health Sector.

According to the above defined objectives, we now present the artifact that this research proposes to achieve those goals and to address the problem identified in chapter 2.

In this work, the artifact consists of a method that uses a bottom-up approach to develop an IRA capable of ensuring an easy maintenance and semantic interoperability.

Then, we describe the different steps involved in the proposed method. It should be noted that Model Integration is the main proposal in this work and adds real value to the context where it is inserted.

4.1 Mapping the main IS in a Sector of Activity

In a given sector of activity there can exist tens or hundreds of IS. In order to design an IRA, it's necessary to understand, with the people in charge for the IS of that particular sector of activity, which are the most important IS that can form the basis for the creation of such IA.

With the identification of such IS, it becomes possible to accede to the most important IE, attributes and relationships for the sector, and thus initiate the development of the IRA.

4.2 Reverse Engineering

After the main IS of a particular sector of activity have been identified and analyzed their models and original database tables, it is necessary to carry out some changes and proceed with their modeling. In this step there are two distinct activities:

1. Application of Reverse Engineering transformations;
2. Model representation with UML and XML.

In the first activity some reverse engineering transformations are applied (section 3.4.2) in order to make an abstraction of the analyzed models and tables and reach individual models for each of the analyzed IS. This abstraction allows us to get models where there is no reference to characteristic database information.

In the second activity the representation of models with UML and XML is carried out. The use of UML is related to the fact that this is the language chosen to represent the IA to develop. On the other hand, the use of XML allows us to represent the attributes of each IE separately, besides facilitating the model integration to be applied in section 4.4 of this work.

The final result of this phase is an individual model for each analyzed IS without characteristic database information.

4.3 Model Enhancements

After the abstraction of models and database tables has been concluded, we proceed with their improvement. To achieve this, some model enhancement transformations (section 3.4.3) are used, as well as UML good practices, such as IE generalization, IE specialization, inheritance, concept separation [23].

Here, we need to remember that there can be an IRA for a more general sector of activity. In that case it is necessary and useful to consider the wider IRA, in order to improve interoperability between the sector under study and the broader sector.

The final result in this phase is an individual model for each of the analyzed IS, with enhancements that increase their understanding and correctness.

4.4 Model Integration

After enhancements in the models of the analyzed IS have been made, we proceed with their integration in order to finish the method proposed in this work and so get an IRA for a specific sector of activity, where there is no duplication of corresponding entities.

Section 3.4.4 provides a more comprehensive and theoretical description regarding model integration. For this reason, it is necessary to frame the theme in the specific context of this work. To reach this purpose, the four steps were adjusted to the solution of model integration with special attention to the second phase – model comparison.

4.4.1 Pre Integration of the Solution Proposal

In the context of this work, and in order to achieve the objective of identifying IE corresponding to different models, only binary comparisons will be made, that is, only two models are compared at a time.

4.4.2 Model Comparison of the Solution Proposal

We will base ourselves in the process described in [22] to make the IE comparison of different models.

First, it is necessary to quantify the concept of correspondence in this work, set a minimum threshold for concluding that two IE are correspondent and assign weights for each of the similarity rules described above. After that, the correspondence process to be used in this investigation is defined.

4.4.2.1 Name and Attribute Correspondence

For the scope of this study, we consider that two names are correspondent if they have the same or synonymous labels, or if the name correspondence techniques allow us to establish which are the same or similar. Regarding the attributes correspondence, we consider that two IE are correspondent if at least 75% of the attributes of an IE have correspondence in the attributes of another IE. These parameters can be configured according to the needs and the degree of certainty required by the situation.

4.4.2.2 Cut-off

As far as the minimum threshold to conclude that two IE are correspondents (cut-off) is concerned, we choose the value of 0.75 because we think that, following the process defined in 4.4.2.5, this value is high enough to avoid false positives. This parameter can be configured according to the needs and the degree of certainty required by the situation.

4.4.2.3 Grouping Similarity Rules

Table 1 brings together the similarity rules described in 3.4.4.2.2 graduated by their level of importance. In Group 1 the most important rules to verify the similarity between IE are present and in Group 4 the less important ones. When drawing up the group, we considered that the model integration to carry out this work is based on a bottom-up approach, where the final model is achieved by applying bottom-up primitives, starting from basic concepts and building more complex concepts [24]. For this reason, we consider that the attributes or the description of an IE (rules 3 and 4) are more important than the name given to an IE (Rule 5).

Group 1	Group 3
Rules 1 and 2	Rule 5
Group 2	**Group 4**
Rules 3 and 4	Rules 6 to 12

Table 1. Grouping Similarity Rules

4.4.2.4 Assigned Weights by Similarity Rules

The assigned weight to each rule is given according to the equation:

$$\boldsymbol{Weight}_{\boldsymbol{SimilarityRule}} = \frac{1}{Group_{Number}}, \text{where} \quad (1)$$

$Group_{Number}$ refers to the group number that contains a certain similarity rule.

Thus, Table 2 shows the weights of each of the similarity rules described in 3.4.4.2.2.

Rule	Weight	Rule	Weight
1	1	7	0.25
2	1	8	0.25
3	0.5	9	0.25
4	0.5	10	0.25
5	0.33(3)	11	0.25
6	0.25	12	0.25

Table 2. Assigned weight by Similarity Rules

4.4.2.5 Correspondence Process between Models

In this work, we will use the following process of model correspondence:

1. Given two models, we intend to calculate the similarities between any pair of EI;
2. Choose a pair of IE to be compared;
3. Iterate, in order, for all the similarity rule groups defined in 4.4.2.3 and apply, the maximum, one rule of each group;
4. Sum weights of the rules applied;
5. When sum of weights is higher than the cut-off (0.75), IE are considered correspondent and the process ends;
6. If at the end of the process, the value obtained is lower than the cut-off (0.75), IE are not considered correspondent.

In case any evident correspondence is validated through this process, we should proceed to a contextual interpretation, using, if necessary, someone responsible for the IS involved.

At the end of this step it will be possible to present the correspondent IE between different models.

4.4.3 Conforming Models of the Solution Proposal

By using the similarity rules and the correspondence process between models described in 4.4.2, conflicts are minimized. For example, we do not run the risk of two homonymous IE (same spelling and/or pronunciation but different meanings) be considered correspondent just because the descriptions and attributes of the two IE are taken into account.

4.4.4 Model Unification of the Solution Proposal

In this phase we proceed with the unification of the models of each analyzed IS. To accomplish this unification we use the transformations presented in section 3.4.4.4.

The final result of this step is a model that is the result of the unification of the analyzed IS, where there is no duplication of corresponding entities. This model is also the final output of the method suggested in this work to define an IRA for a particular sector of activity.

5. DEMONSTRATION

After the proposed solution has been dealt with, it is now time to carry out a demonstration to clarify how that solution allows us to solve the problem defined above, answer the research questions and achieve the objectives defined.

In this demonstration, an instantiation of the proposed solution for the Portuguese Public Health Sector is implemented, culminating in the creation of an IRA for that sector.

However, and because the documentation of the IS of that particular sector is confidential and highly complex, in this document we will only fully describe an instantiation by using an academic example.

Still, some important considerations and the result of the instantiation for the Portuguese Public Health Sector will be presented.

5.1 Demonstration with an Academic Example

In this demonstration we used an academic example where the steps of the solution proposal of this work are followed (chapter 4).

For a better understanding, this demonstration follows the structure:

1. Demonstration of the Reverse Engineering and Model Enhancements steps;
2. Demonstration of the Integration Model step.

5.1.1 Demonstration of the Reverse Engineering and Model Enhancement steps

In this section we intend to make the demonstration of the Reverse Engineering and Model Enhancement steps of the proposed solution of this work (section 4.2 and 4.3). The following academic example is considered:

Given five database tables (Figure 1), we intend to obtain a model that conceptually represents these same tables and the relationships established between them.

On the table, *FN* indicates the number of the field of the table, *PK* indicates if a certain field is or is not Primary Key of the table and *FK* indicates if a certain field is or is not Foreign Key of the table. The Physician table and the Auxiliary table represent concepts that are specifications of the Person concept.

Database table "Person"

FN	Field Name	PK	FK
1	personID	X	
2	name		
3	age		
4	genre		
5	addressID		X
6	dbCreatedBy		
7	dbSincDmlID		

Database table "Address"

FN	Field Name	PK	FK
1	addressID	X	
2	district		
3	city		
4	localitaly		
5	postalCode		
6	street		
7	apartment		
8	foreignAddress		
9	foreignLocality		
10	foreignPostalCode		

Database table "Physician"

FN	Field Name	PK	FK
1	physicianID	X	
2	name		
3	age		
4	orderNumber		
5	speciality		
6	workplace		
7	contactID		X
8	dbCreatedBy		

Database table "Contact"

FN	Field Name	PK	FK
1	contactID	X	
2	contactType		
3	number		
4	dbCreatedBy		

Database table "Auxiliary"

FN	Field Name	PK	FK
1	auxiliaryID	X	
2	workSchedule		
3	contactID		X
4	dbCreatedBy		

Figure 1. Academic Example's database tables

Starting from the tables in Figure 1, it is necessary to understand how these tables are related to each other, which fields represent primary or foreign keys, which fields represent attributes that must be removed at a higher level of abstraction, etc.

To do that, we followed the strategy described in section 4.2 of this document. Specifically, IE are developed (in the example, each table happens to be represented as IE) attributes that act as foreign keys are removed (e.g., attribute addressId of the Person table), attributes which only have meaning at the level of databases are also removed (for example, the attribute dbCreatedBy of the Person table).

After all those modifications are identified and performed, the UML is used to carry out the modeling, having as a result the model presented in the Figure 2 (from now on referred as model A).

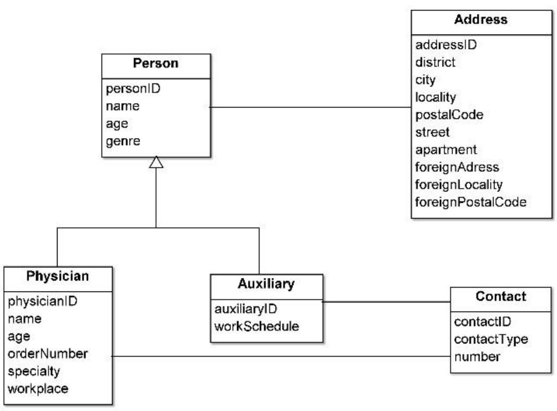

Figure 2. Academic Example's Model A after Reverse Engineering

Once we have reached to a model where no characteristic information of database are present, it is necessary to enhance that model.

The strategy followed is described in the section 4.3.

Summarizing it, the concepts separation is made (eg, IE Address can be split into two IE, as clearly represents the concept of Portuguese address and foreign address); redundant attributes are removed (attributes Name and Age in the Physician IE are redundant because they are also present in the Person IE); the relationship generalization is performed (the relations between the IE Contact and Physician and between the IE Contact and Auxiliary should be generalized, by removing these two relationships and creating a new relationship between Contact IE and Person IE).

After these changes are performed, we get the model presented in Figure 3.

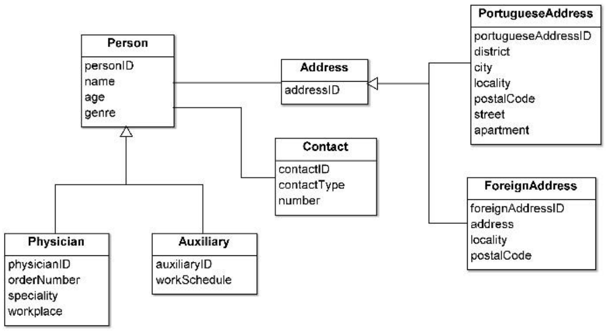

Figure 3. Academic Example's Model A after Model Enhancement

5.1.2 Demonstration of the Integration Model step

In this section we intend to make the demonstration of the Integration Model step of the proposed solution of this work (section 4.4), where the phases of comparison and unification of models are specially treated. The following academic example is considered:

Given two models A and B, where model A refers to the outcome of the previous demonstration (Figure 3), and model B is shown in Figure 4, we initially intend to realize which IE are the correspondent IE in the two models. In this case, A.Person means: "IE of model A with the name Person".

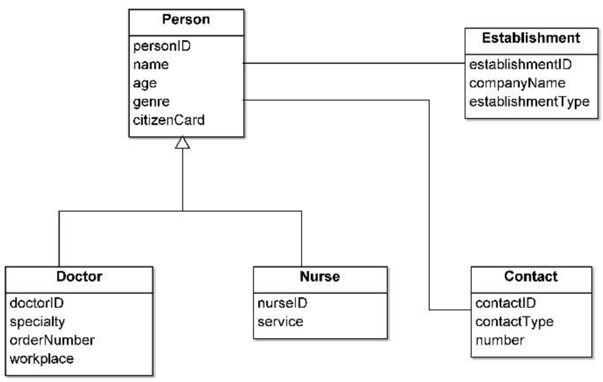

Figure 4. Academic Example's Model B

By simply analyzing the models in Figures 3 and 4, it is possible to estimate that the IE pairs *A.Person/B.Person*, *A.Physician/B.Doctor* and *A.Contact/B.Contact* are correspondent. However, we must demonstrate that this idealization is true. For demonstration purposes, we will proceed by comparing only the second identified pair.

Iterating through the groups of the similarity rules defined in 4.4.2.3, the first rule that can be applied is Rule 4 - *If the attributes of two IE are equal, it is likely that the two IE are also equal.* To apply this rule, there must be a parity at the attribute level of 75% or more. We count the weight of this rule (0.5) in a variable starting at zero. Since the value accumulated in this variable does not exceed the cut-off established of 0.75, we must continue the iteration. So, the second rule to apply is the Rule 5 - *If the name of two IE is the same or similar, it is likely that the two IE are equal or similar* - because IE names are synonymous. By adding the weight of this rule (0.33) to the variable initialized earlier, we come to the value 0.88. Since this value already exceeds the cut-off of 0.75, we consider that the two IE are correspondent and the correspondence process comes to an end.

Following a similar reasoning, we conclude that the other two IE pairs previously mentioned also contain correspondent IE, and that there is no other pair of corresponding IE.

At the end of model comparison, we conclude that *A.Person* is correspondent to *B.Person*, *A.Physician* is correspondent to *B.Doctor* and *A.Contact* is correspondent to *B.Contact*.

Based on these conclusions, we shall now proceed to solving the existing conflicts. In this example, there is a name conflict, since *A.Physician* and *B.Doctor* have synonym names. Therefore, we will rename one of these IE so that both keep the same name.

After solving all the conflicts, it is then possible to proceed with model unification. Exploring the unification techniques referred in 3.4.4.4, we will use the "*Merge IE*" transformation to collapse all the corresponding IE pairs.

The final outcome of this step is a model resulting from the unification of the analyzed IS, where there is no duplication of corresponding entities. Thus, we get the final unified model, which ensures that there is no duplication of corresponding IE. That model is presented on Figure 5.

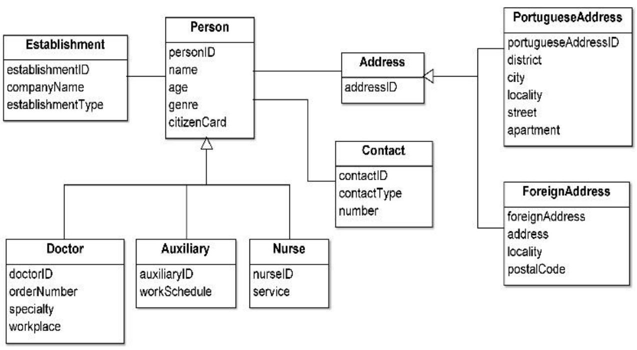

Figure 5. Academic Example's Model A and B Unification

A remark to point out that the resulting IE name of the union of the IEs A.Physician and B.Doctor is Doctor as it is the most usual term.

5.2 *Information Reference Architecture for the Portuguese Public Health Sector*

In this section we present the instantiation to the Portuguese Public Health Sector.

As the reasoning was the same used in the demonstration of the previous section and because the documentation on the base of this instantiation has a strong nature of confidentiality and a high level of complexity, we decided to present just the final result of the instantiation accompanied by some important considerations.

Together with the people in charge of the IS in the Portuguese Public Health Sector we came to the conclusion that the development of an IRA was of utmost importance. Concomitantly, it had to have as support the three IS considered the pillars of the Portuguese Public Health Service in the future. The three IS identified are the *Registo Nacional de Utente*s (RNU), *Registo Nacional de Profissionais* (RNP) and the *Sistema de Gestão de Entidades Informacionais* (SGES).

Since there is a common IRA for all the Portuguese PA (broader than the Portuguese Public Health Sector), on this instantiation we will use concepts representations that are present in the Portuguese PA IRA. That is, in order to improve the alignment between the PA IRA and the analyzed IS, and because there is not a unique representation for some concepts (for example, the concepts Address, Contact, Birth), we have decided to represent these concepts just as they are represented in the Portuguese IRA.

The resulting product of this phase is a final individual model resulting from the unification of the analyzed IS - **the proposal of this work - that can serve as IRA for the Portuguese Public Health Sector**.

Because of its dimension, this model can be found on the last page of this document.

6. Evaluation

After having described the solution proposal and its instantiation, it is now time to refer the way the evaluation of the work will be carried out.

According to [4], evaluation is considered one of the most important components in the DSRM since it is in this activity that we validate the contribution of the artifact developed to respond to the problem identified above (chapter 2) as well as its utility, quality and efficacy.

For better understanding the scope of the instantiation carried out to the Portuguese Public Health Sector, Figure 6 conceptually shows the identified current situation on the motivation and the situation we want to achieve with this solution. The representation on the left illustrates the identified motivation: there is no IRA for the Portuguese Public Health Sector to facilitate interoperability between the Portuguese Public Health Sector IS and the PA IRA; the oval shape represents the lack of an IRA that takes into account the Portuguese Public Health Sector IS. The representation on the right illustrates the solution proposed in this paper: the creation of an IRA for the Portuguese Public health sector, which is a specification of the PA IRA and takes into account the Portuguese Health Sector IS and uses some of the best practices and PA IRA definitions and representations.

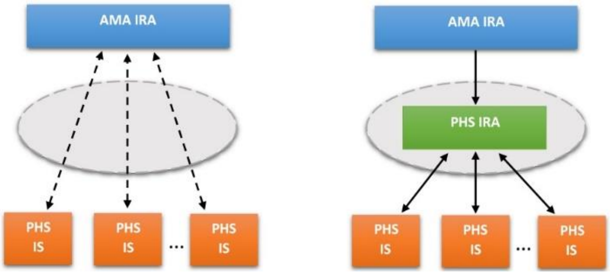

Figure 6. Current and Expected Situation

As stated in [2], the fact of using an IRA improves the interoperability amongst models. **Therefore, this solution leads, by itself, to an increase on the interoperability level referred previously.**

The evaluation carried out in this work is done with the help of metrics adequate to evaluate, in a practical way, the proposed artefact - method for bottom-up development of an IRA that can ensure an easy maintenance and semantic interoperability. To achieve this, we use the proposed model as an IRA for the Portuguese Public Health Sector, and compare it with the IS which form the basis of its development, and with other works developed in a similar context.

In these comparisons, the following metrics are used [25] [26]:

- Number of IE;
- Number of relationships between IE;
- Number of attributes

These metrics are used to assess the dimension of an IA or of a model, and are directly related to the efficacy of their maintenance [27]. That is, the lower the value of these metrics, the higher efficacy of maintenance of an IA or of a model.

6.1 Evaluation Conclusions

As mentioned at the beginning of this chapter, the fact that a specific sector follows an IRA leads to an improvement in the interoperability between models of that same sector. Thus, we conclude that the instantiation of the work performed in the Portuguese Public Health Sector leads, in itself, to an increase in the level of the interoperability referred above.

Furthermore, by implementing the metrics identified in this chapter it is possible to draw conclusions regarding the proposed method of bottom-up development of IRA that ensures easiness of maintenance, and its instantiation to the Portuguese Public Health Sector.

As for comparisons with IS that underpinned the development of the model proposed in this work, it is possible to verify that the IRA for the Portuguese Public Health Sector contains less IE, fewer relationships between IE and fewer attributes than the sum of models that are the basis of the same IRA. Thus, **it is possible to conclude that the proposed method in this assignment allows to create IRA that are easier to maintain than the IS set that are at its basis.**

It is also possible to specialize this evaluation, by recurring to the conclusions drawn from the comparison of the suggested model as IRA for the Portuguese Public Health Sector with other similar works. The following sections describe the main conclusions from these comparisons.

6.1.1 Comparison with HL7 RIM

The fact that the proposed model in this work as IRA for the Portuguese Public Health Sector results from the integration of three different models, inevitably leads to the consequence of containing a larger number of IE, attributes and relationships than the HL7 RIM. Therefore, **it is possible to state that the IRA proposed has lower maintenance efficacy than the HL7 RIM**.

On the other hand, the fact that HL7 is international, and doesn't take into account the specific context of the Portuguese Public Health Sector or the PA IRA, leads us to the conclusion that **the proposed model ensures higher interoperability between the Portuguese Public Health Sector IS and PA IRA**.

6.1.2 Comparison with the Document Operational Programme for Health 'Saúde XXI'

The document Operational Programme for Health 'Saúde XXI' aims to identify the most relevant IE of the Portuguese health sector, so, we can say that it fills in a context similar to the one of this work.

This document identifies a larger number of IE and relationships between them than the proposed model in this work for the Portuguese health sector IRA.

Therefore, **any implemented information architecture based on the document 'Saúde XXI' would always be more difficult to maintain than the model proposed in this paper**.

The fact that the document 'Saúde XXI' does not take into account the AP IRA, also allows us to conclude that **the proposed model ensures higher interoperability between the Portuguese Public Health Sector IS and the PA IRA**.

7. Conclusions

One of the identified ways to improve continuous competitiveness of organizations is through the creation of IA that can help in the sharing and exchanging of information or in the fast and easy development of IS.

As far as sectors of activity are concerned, the lack of an IRA to guide and restrict the instantiations of a set of architectures and individual solutions leads to interoperability problems, translating, among other things, great difficulties in maintaining the existing IS.

The solution that this assignment proposes is defined as a method that uses a bottom-up approach and comes from a set of IS to get to an IRA that will be easy to maintain.

The method suggested in this study and its instantiation to the Portuguese Public Health Sector were assessed with the help of evaluation metrics that allow us to reach to conclusions concerning the maintenance of an IA or model.

7.1 Main Contributions

We believe that the proposed solution described in this work adds some value to the context in which it is inserted: the creation of IRA to ensure easiness of maintenance and semantic interoperability.

The objectives defined for this investigation were successfully achieved, and each one of them represents an important contribution of this work.

As general contributions we underline the creation of a method for the bottom-up development of an IRA that can ensure both easiness of maintenance and semantic interoperability, and the identification of IE correspondence between different models.

As far as the specific contributions for the Portuguese Public Health Sector are concerned, we emphasize the identification of the most relevant IE, as well as the development of an IRA.

7.2 Communication

In the context of this investigation, a scientific articled was published in the 17th IEEE Conference on Business Informatics. The step Model Integration of the solution proposal suggested in this work deserved special attention (section 4.4).

Besides, this document also aims to contribute to the communication of the undertaken investigation, by offering, to the academic community, a new study component for future research.

7.3 Limitations and Future Work

In spite of the contributions coming from this solution proposal, there are some limitations that must be considered. These limitations can be directly associated to improvements to be worked in the future.

Thus, we suggest a deeper investigation on the Reverse Engineering and Model Enhancement steps of the suggested model; the instantiation of the suggested model to other sectors; the development of other evaluation metrics to assess the international level of an IA; and the use of HL7 RIM to improve the suggested model of this work for an IRA in the Portuguese Public Health Sector.

REFERENCES

1. **Watson, R. W.** *An Enterprise Information Architecture: A Case Study for Decentralized Organizations*. IEEE Computer Society (2000)
2. **CIO, Office of the Dod.** Reference Architecture Description, 2010
3. **Agência para a Modernização Administrativa.** *Arquitetura Informacional*. [Online] https://m6.ama.pt/docs/ArquiteturaInformacional.pdf. [Acedido em 4-10-2015]
4. **Henver, A.R., et al.** *Design Science in Information Systems Research*. MIS quarterly **28**(1) (2004) pp. 75-105
5. **Lankhorst, M., et al.** *Enterprise Architecture at Work*. Springer (2005)
6. **Godinez, M. et al.** *The Art of Enterprise Information Architecture: A Systems-Bases Approach for Unlocking Business Insight*. IBM Press (2010)
7. **Spewak, S. H. and Hill, S.** *Enterprise architecture planning: developing a blueprint for data, applications and technology*. QED Technical Publishing Group (1993)
8. **Vasconcelos, A.** *Arquitecturas de Sistemas de Informação no Contexto do Negócio. Dissertação para a Obtenção do Grau de Mestre em Engenharia Informátíca e de Computadores (*2001)
9. **Health Level Seven International.** [Online] http://www.hl7.org/implement/standards/rim.cfm [Acedido em 16-09-2014]
10. **Centro de Engenharia Organizacional.** *Documento de Análise: Arquitectura Informacional e Análise aos Processos de Saúde*. (2003)
11. **Bernstein, A.** *Middleware: a model for distributed system services*. ACM (1996)
12. **Interoperability Solutions for European Public Administrations.** *European Interoperability Framework (EIF) for European public services*. (2010)
13. **Tran, H., et al.** *View-based reverse engineering approach for enhancing model interoperability and reusability in process-driven SOAs*. High Confidence Software Reuse in Large Systems. Springer Berlin Heidelberg (2008) pp. 233-244
14. **Chiticariu, L.** *Enabling data interoperability with provenance and schema integration*. University of California (2008)
15. **Chikofsky, E. J. and Cross, J. H.** *Reverse Engineering and Design Recovery: A Taxonomy*. IEEE **7**(1) (1990) pp. 13-17
16. **Bagui, S. and Earp, R.** *Database Design Using Entity-Relationship Diagrams*. Auerbach Publications (2003)

17. **Batini, C., et al.** *A Comparative Analysis of Methodologies for Database Schema Integration*. ACM Computing Surveys **18**(4) (1986)

18. **Rizopoulos, N. and McBrien, P.** *Schema Merging Based on Semantic Mappings*. Springer-Verlag (2009)

19. **Rizopoulos, N.** *Schema Matching and Schema Merging based on Uncertain Semantic Mappings.* University of London (2009)

20. **Rahm, E. and Bernstein, P.** *A survey of approaches to automatic schema matching*. The VLDB Journal 10(4) (2001)

21. **Su, X.** *A Text Categorization Perspective for Ontology Mapping*. Dept. of Computer and Information Science Norwegian University of Science and Technology

22. **Ehrig, M. and Sure, Y.** *Ontology Mapping – An Integrated Approach*. LNCS (2004)

23. **Fowler, M.** *UML Distilled: A Brief Guide to the Standard Object Modeling Language*. Pearson Education (2003)

24. **Batini, C., et al.** *Conceptual Database Design: An Entity-Relationship Approach*. The Benjamim/Cummings Publishing Company, Inc (1992)

25. **Vasconcelos, A.** *Arquitecturas dos Sistemas de Informação: Representação e Avaliação. Dissertação para a Obtenção do Grau de Doutor em Engenharia Informática e de Computadores*, Instituto Superior Técnico, Universidade Técnica de Lisboa (2007)

26. **Vasconcelos A., et al.** *Information System Architecture Metrics: An Enterprise Engineering Evaluation Approach* (2006)

27. **Genero, M., et al.** *Defining and Validating Metrics for Assessing the Maintainability of Entitity-Relationship Diagrams* (2003)

Information Reference Architecture for the Portuguese Public Health Sector