An overview of attributes characterization for interoperability assessment from the public administration perspective

José Marcelo Cestari, Eduardo Rocha Loures, Eduardo Santos, Yongxin Liao, Hervé Panetto, Mario Lezoche

An overview of attributes characterization for interoperability assessment from the public administration perspective

José Marcelo A. P. Cestari[1], Eduardo de Freitas R. Loures[1], Eduardo A. P. Santos[1], Yongxin Liao[1], Hervé Panetto[2] and Mario Lezoche[2]

[1] Pontifical Catholic University of Parana, Industrial and Systems Engineering, Curitiba, Brazil
{jose.cestari, eduardo.loures, eduardo.portela, yongxin.liao}@pucpr.br
[2] Université de Lorraine, CRAN, UMR 7039, Vandoeuvre-lès-Nancy, France
{herve.panetto, mario.lezoche}@univ-lorraine.fr

Abstract. The government interoperability frameworks (GIFs) are mainly centered on technical perspective neglecting important actual issues in Public Administration (PA) as the performance of their services and process as well its strategic, legal and politic positioning. This wide spectrum of organizational knowledge demands specific interoperability assessment (IA) methods based on a preliminary step of identifying and organizing adequate attributes that allow a wider view on the PA domain regarding interoperability requirements. This paper proposes a knowledge discovering and extraction method from PA and IA literature, presenting a rationality through a conceptual model in order to identify and organize attributes that will be used as input for an IA model. A preliminary set of attributes is presented, as well as a future approach of how to deal with these attributes using AHP (Analytic Hierarchy Process) in order to evaluate them with experts through a specific structure suited to the IA process itself.

Keywords: Interoperability Assessment, Public Administration, Attributes Extraction, e-Government.

1 Introduction

Organizations are facing a competitive marketplace and they must develop partnerships and work in an integrated way with competitors and stakeholders. As one of the important factor of this collaboration, interoperability [1, 2] is the ability of two or more systems (or components) of: sharing specific information, using them with a specific semantics and creating, with that information, a specific mission in a precise context. Interoperability takes into account dimensions such as concerns, barriers, degrees of maturity and assessment, and consists of more than technical aspects [3]. Interoperability assessment (IA) is the method that defines some formal metrics to quantify and qualify the level of interoperability between heterogeneous systems from one or more organizations [4]. With the increasing use of ICT (Information and

Communication Technologies) in government institutions, the concept of eGovernment emerges in the late 1990s [5]. The term eGovernment, e-gov, eGov and similar are an abbreviation of "electronic government" and refers to the use of ICT to support the government business, providing or enhancing public services or managing internal government operations [6]. Assembling the concepts of interoperability and eGovernment, an eGovernment interoperability domain arises, in a broad sense, as the ability of constituencies (or public agencies) to work together [7].

The literature analysis, presented in [8], reveals that there are many works regarding interoperability models and government initiatives of eGov models and frameworks. Nevertheless, since 2000 it was not possible to identify specific models, frameworks or assessment procedures regarding the evaluation of government entities not related to (almost entirely) technical issues and also it was not found in the literature review papers describing the rules (rationale and procedures) regarding the knowledge discovering steps [9, 10, 11].

The desire to map the concerns and identify the barriers (other than the technical ones) is a recent interest in the literature [12]. This new approach, focused on semantics, opens an important issue regarding the proper collection of attributes related to interoperability of these other dimensions related to public administration (e.g. processes, services, governance, legal and political). Obtaining, identifying and organizing relevant knowledge are important steps to the IA considering also the models and frameworks established in Enterprise Interoperability Assessment (EIA) in the private industrial sector, as well as the work done in eGov.

An exploratory approach into the literature sources will allow the identification of important dimensions (concerns, barriers and attributes) for the evaluation of the ability to interoperate in the governmental entities. There is need of a specific rationale for this broader knowledge extraction of organizational behavior into the public domain. In fact, to characterize the qualification attributes and the IA in their different visions other than the technological, such attributes and guidelines should present the structure, the quality and the completeness of observation in the IA performed by a method that ensures these requirements.

It is important to contextualize that this paper is part of a wider research scope, represented in Figure 1 and more specifically related to the process A2 and A3. The IDEF0 diagram [13] identifies basic steps (which are contained in phases and may contain activities), proposing an attributes' and guidelines' structure collection and composition as well as the assessment method definition. The results of A0 and A1 processes can be summarized as (i) definition of a database of related papers, (ii) analysis regarding the world publication and authoring, helping the justification of the research and mapping the distribution of the subject, (iii) domain definition and (iv) identification of gaps regarding the existence models and related works. Based on the information gathered from A0 and A1, the next step (A2) relates to the identification of preliminary set of attributes and guidelines and in A3, in conjunction with expert assessment, a fact-oriented transformation (FOT) method is applied to evaluate and gather new (if any) attributes from other knowledge sources and types of databases. Those identified objects can be considered as a set of best practices, requirements or desired interoperability characteristics within public entities. The organization, meth-

ods and tools regarding the process A2 is exposed in Figure 2, which also provides an overview of the structure and main goals of this paper.
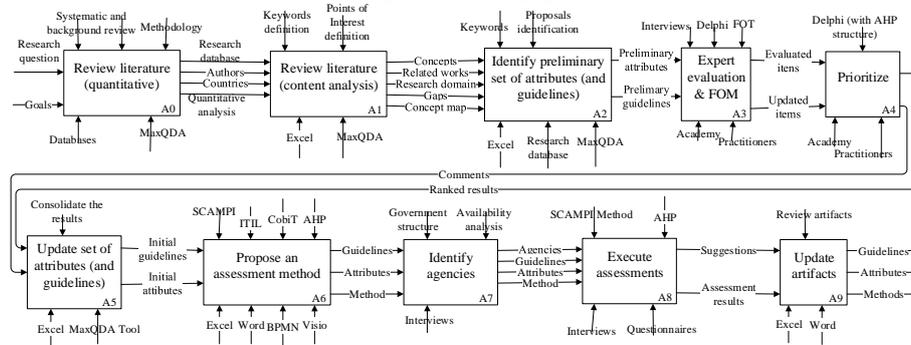

Fig. 1. Full research methodological steps

Considering all the presented aspects, this paper exposes the rationale for the extraction of attributes, as well as a set of preliminary attributes themselves in order to help the IA of public administration entities. Figure 2 shows a general structure of the components adopted as basis for the extraction process, taking into account that the solution space helps to provide information in order to solve issues related to the problem space.
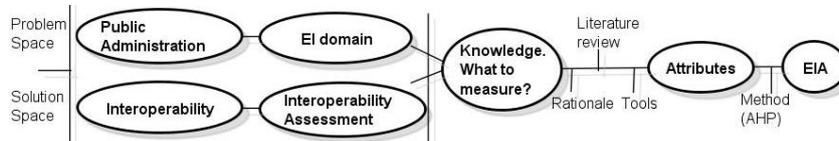

Fig. 2. Structure of the attributes extraction and its application for EIA

After that, a join of the aspects leads to reflections associated to measurement selection, culminating in the generation of a set of attributes and its use into the EIA applications and domain. In addition, a proposal of ranking mechanism (for the attributes) based on AHP [14] is presented as one of the next steps of the research.

2   Methodology

According to [15], knowledge is a justified belief that increases an entity's capacity for taking effective action. In [16] the authors defines knowledge as experience, facts, rules, assertions and concepts about their subject areas that are crucial to the business (customers, markets, processes, regulations). There are two basic dimensions of knowledge: tacit and explicit [15], which can be combined to an ontological dimension. In complement, knowledge management is a process of identifying, capturing and leveraging collective knowledge to help the organization compete [17]. Knowledge can be represented using different formal forms, among others, semantic

nets, rules, ontologies, mind maps and conceptual maps. A general model of the information retrieval process is exposed in Figure 3, based in [18].
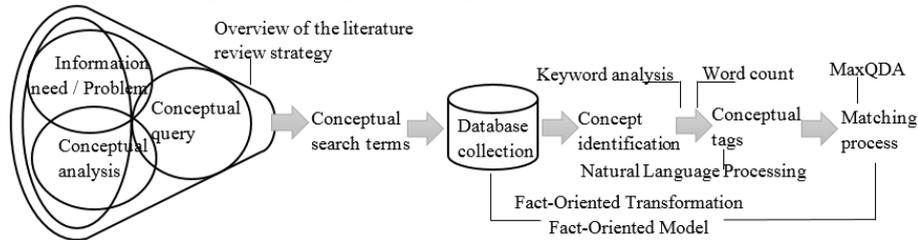

Fig. 3. Overview of the attributes extraction steps

In the initial steps, information regarding the problems of the domain and the concepts involved were processed with conceptual queries in order to filter and select an initial documents database with the execution of conceptual searches. In summary, considering the existence of a database collection, the steps to extract the attributes are: (i) concept identification; (ii) keyword analysis and word count as support mechanisms; (iii) tagging words to identify groups of knowledge and (iv) matching process to search for relations among the words found and root concepts of the research. These steps are described in the subsections bellow.

2.1 Concept identification

As the goal is to extract attributes from the literature (knowledge) database, it is important to define and characterize attributes as in: (i) A quality or characteristic given to a person, group, or some other thing. An abstraction belonging to or characteristic of an entity [19]; (ii) A quality or feature of a person or thing, especially one that is an important part of its nature [20]; (iii) A quality or feature regarded as a characteristic or inherent part of someone or something [21] and (iv) Regarding class diagrams, an attribute represents a data definition for an instance of a classifier. An attribute describes a range of values for that data definition and may describe the structure and value of an instance of a class [22].

Although the grammar structure is a specific body of knowledge (and not the focus of this research), it is important to mention that attributes are nouns and/or adjectives that modify that noun (e.g. collaboration (noun), collaborationist (adjective)), and guidelines are suggestions to show how to behave and usually derives action from its statements [19]. Guidelines can contain (or suggests) attributes and they (guidelines) will be used (along with the attributes themselves) to assess the interoperability of the entities, with a similar approach exposed in [23]. Figure 4 illustrates the relation among important concepts for the research, which is useful in the process of attributes extraction and the idea is to expose the components that are connected (somehow) primarily with an attribute and secondly with a guideline. Considering the database of files regarding the domain (150 documents), a keyword analysis and counting was made as a first attempted to derive some attributes. The rationale adopted was that if a document deals with such important issues of the government interoperability

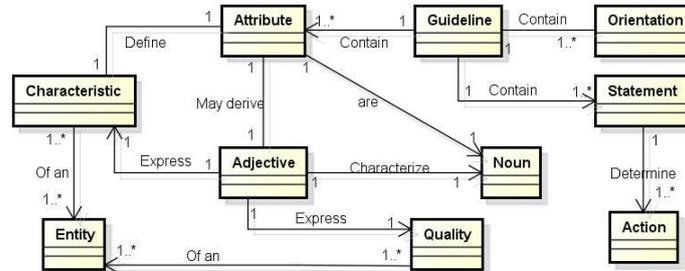aspects, there is a chance that qualification words (attributes) could be cited in the keywords of the paper.

Fig. 4. Concept model of attributes

The keywords are not only "one word", but may be also a group of words separated by commas or other symbols. The MaxQDA software [24] was used to count (and organize) the keywords of the documents and the results are represented in Table 1.

Table 1. Quantitative overview of keywords.

# of appearances	# of words	# of appearances	# of words
1	178	7	2
2	21	8	1
3	6	10	1
4	4	55	1
6	1	53	1

The highest appearance are the words "interoperability" (55 times) and "e-government" (53 times) and this was expected because of the rules adopted during the literature review (these two words were one of the most important composing the search strings) [8]. Nevertheless, as the idea is not have a sort of a "sample" approach, the quantity and distribution of the keywords are not as important as the meaning in terms of creating knowledge [15]. However, after the analysis of all keywords retrieved, it was detected that most of them are related to the paper itself (as expected) but in general ways. That is, the keywords tends to show generic aspects of the paper, containing words of global meanings, in order to introduce to the reader some "tags". Because of this, it was not possible to find words that could be used (directly or indirectly) to compose the preliminary set of attributes.

As the keyword analysis did not bring information about possible attributes, it was defined that the extraction of the attributes would follow a process based on a word counting and Natural Language processing analysis. According to [25], an average English word has from 4.5 to 7 characters and, using the MaxQDA software [24], the whole documents database were scanned searching for words of six characters length. In a first round, 44,111 words were found and, after a second round of execution (with the elimination of authors names, symbols, irrelevant words and words that appeared less than 10 times) a total of 21,644 words were selected. The quantitative analysis by itself does not solve the problem of finding attributes, so it was necessary

investigate some of the semantic meaning of this large group of words. That is, as a next step, try to discover the lexical linguistic category (e.g., verbs, nouns, adverbs) of the words, searching specially for adjectives (considering the relations in Figure 4).

2.2 Conceptual tags, matching process and results

As an attribute may define (or derives to/from) adjectives and also because of the large number of existent words (21,644), a Natural Language Processing Software was used in order to detected the adjectives in the word's database. The software adopted was the Stanford POS Tagger [26], maintained by the Stanford Natural Language Processing Group and built based on the research of a log-linear part-of-speech taggers described in [27]. A Part-Of-Speech Tagger (POS Tagger) is a software that reads text in some language and assigns parts of speech to each word (and other token), such as noun, verb and adjective. The tool was also used in [18] within the natural language processing approach. The list of 21,644 words were processed with the Stanford POS Tagger and the results were 3,739 words tagged as adjectives.

At this point, considering that the set of words identified as adjectives could contain a subset of attributes, there is a need to search for a connection among each of the identified words with the concept words that represents the research subject domain (interoperability within public administrations/government). This process was done using a lexical matching functionality of MaxQDA tool and the results are the following preliminary list of attributes: adaptability, collaboration, commitment, communication (semantic), conflicts, cooperation, economical, efficiency, historical, integration, loyalty, politic, process, responsibility, senior management support, sociological, standardization.

2.3 Conceptualization process (with Fact-Oriented Transformation)

Figure 5 presents an approach related to the process A3 (Figure 1), aimed to evaluate (expert view) and gather new information from other types of knowledge sources and databases (structured or semi-structured data) in order to corroborate and improve/refine the obtained set of attributes.. The first step is a reverse engineering method through which a model from the application or schema level can be derived (Step 1). Then, the resulted model is enriched through a step of Application Knowledge Injection (AKI) where the model is automatically compared and matched with existing domain and application ontology (Step 2). The model is corrected through a Knowledge Extraction and Mapping Validation in a harmonizing step (Step 3). In fact, the model is examined with the help of a domain expert or an end-user, who is a qualified person to describe the context of the peculiar domain and to put in evidence the contextual knowledge. The next step (Step 4) is a FOT through the application of a set of patterns rules for transforming the enriched conceptual model to a Fact-Oriented Model (FOM) with its finest-grained semantic atoms completely displayed [28]. The Step 5 represents a structural optimization through the application of the Formal Concept Analysis methodology. The output of the entire conceptualization

process is a set of formal objects (attributes) inserted in an optimized structure derived from the data analysis and ready to be conceptually analyzed and tagged.
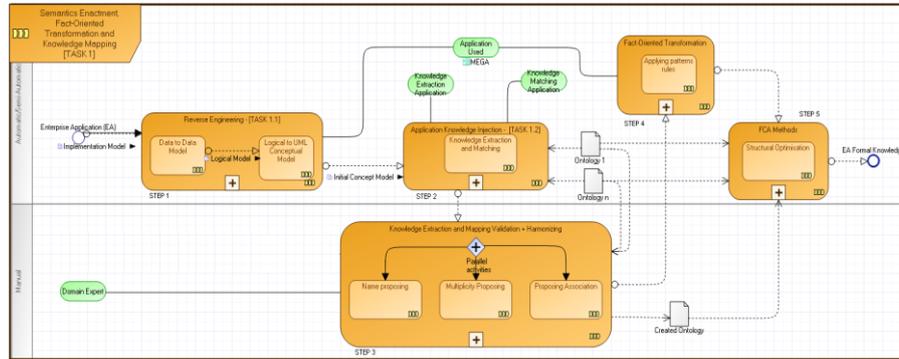

Fig. 5. Entire conceptualization process overview for concept identification

2.4 The use of attributes and guidelines on an AHP structure

The AHP method, proposed in [14], is one of the widely accepted and frequently used mathematic analysis method that supports multi-criteria decision-making. The priority assessment in AHP is based on mathematical analysis of pairwise comparisons defined on hierarchical tree structure, in which, the priority scales between each two related terms are relying on the judgments of domain experts. A significant number of research projects employing AHP method has been proposed in several domains [29], however, as highlighted in [30], little work has been done in specifying and formalizing the important concepts together with their relationships from the AHP method itself. In [30] the authors propose an ontological representation of AHP method to deal with the lack of a deep conceptual understanding on its objects (attributes and reasoning) as well as the lack of flexibility and reusability. This conceptual model (Figure 6) acts in this research as a reference model to facilitate the concept mapping between the domain ontology (obtained in previous knowledge discovering and extraction process) and the IA method.

The importance of having an AHP ontology (Figure 6) merged with a concept model of the attributes (Figure 4) is that, for EIA purposes, each attribute is considered a criterion. Such connection helps the prioritization and evaluation of attributes, made by practitioners and academy, by adopting the AHP technique as a tool to compare the criteria. In addition, during the assessment activities, as each criterion is being compared with others, it is possible to have an adherence degree related to the achievement of the attributes (and guidelines).

The proposed attributes will be firstly evaluated by domain experts and then validated in a selected organization. Figure 7 overviews the evaluation and validation scenario, as a zoom in process A4 (Figure 1). The preliminary attributes (and guidelines) are inputs to expert evaluation process in order to refine those objects. The vali-

dation process is divided in two modes: design and execution mode. In the design mode, a decision hierarchy is firstly created based on refined attributes and the AHP method.
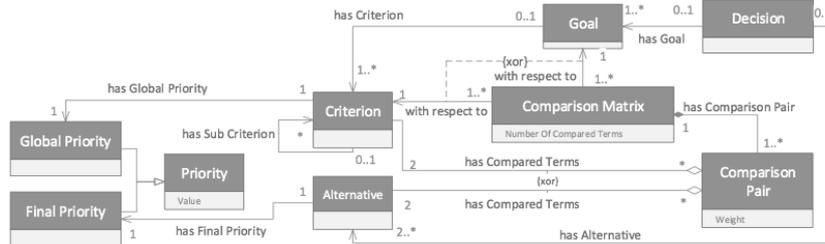

Fig. 6. Ontological representation of AHP

Then, an algorithm will process this hierarchy and produce corresponding questionnaires for criteria assessment and alternative assessment. In the execution mode, practitioners and experts related to the entities will be invited to answer these questionnaires. Finally, based on the assessment of criteria and the alternatives (with respect to each criterion), this validation process will generate the ranked results used as input for the A5 step (Figure 1).
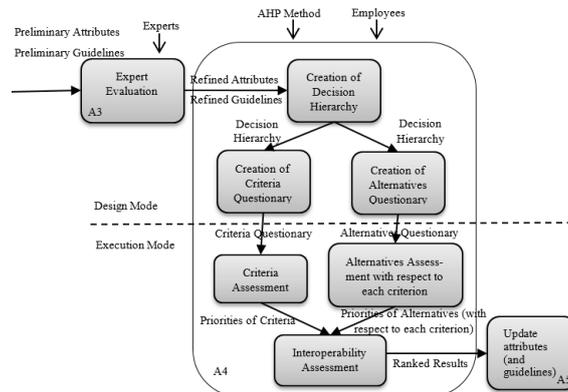

Fig. 7. Overview of the evaluation and validation scenario

Regarding the planning of the assessment and the assessment itself (Figure 1, A6 and A8), the AHP can provide a structure to assess two aspects of EI: potentiality and compatibility [31]. Potentiality can be analyzed as an internal characteristic that reflects the readiness (adherence to attributes) and preparation to interoperate with a possible future partner. Compatibility can also be analyzed regarding internal characteristics, but comparing them between two (or more) known public entities willing to interoperate. In this research, the compatibility approach intends to evaluate which entity, from a group of alternatives (Figure 8, GE B...Z) will better interoperate with a specified entity (Figure 8, GE A). A potential approach implies the ranking of the attributes in order to create ranges that are related to some maturity level. That is,

according to the ranking of the group of attributes of an entity A, the entity is associated to a certain maturity level, with a potential degree for future interoperability actions (Figure 8, ML-1, ML-2 and ML-N).
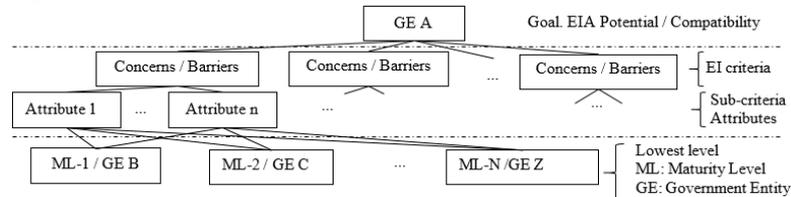

Fig. 8. Potential and compatibility EIA into AHP structure

3   Conclusions

The problem of obtaining, identifying and organizing relevant knowledge is an important issue to the IA in public administration. Considering the models and frameworks established in EIA (in private and public sector), it was detected a gap in the literature regarding the extraction of measurements attributes and criteria. This paper is part of the result of a wider research scope (Figure 1) and presented a rationale of an attributes' extraction from a literature database in order to help the definition and composition of guidelines regarding the IA of public entities. The presented rationale follow steps related to identification of concepts, word count and word analysis, resulting in a set of 17 preliminary attributes and a wide structured and formalized knowledge formal concepts set. The paper also presents an approach of how to adopt a decision-making technique (AHP) in order to organize the attributes (during a validation phase) and assess the interoperability (during an assessment phase).

References

1. IEEE Standard computer dictionary. A Compilation of IEEE Standard computer glossaries, IEEE Std 610, p. 114, DOI 10.1109/IEEESTD.1991.106963 (1991).
2. Carney, D., Fisher, D., Place, P.: Topics in Interoperability: System-of-Systems Evolution. Technical Note, CMU/SEI-2005-TN-002. Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA, http://www.sei.cmu.edu/reports/05tn002.pdf (2005)
3. Pardo, T. A., Burke, G. B.: Improving Government Interoperability: A Capability Assessment Framework for Government Managers. Technical report. Center for Technology in Government, University at Albany, NY, USA. Retrieved from http://goo.gl/20z27B (2008)
4. Yahia, E., Aubry, A., Panetto, H.: Formal measures for semantic interoperability assessment in cooperative enterprise information systems, Computers in Industry, Volume 63, Issue 5, 443-457, DOI : 10.1016/j.compind.2012.01.010 (2012)
5. Camargo, A. R.: Cidade e informática, contatos e interações: explorando manifestações de urbanização virtual, Pesquisa CNPq Proc 301383 /2005-7. Escola de Engenharia de São Carlos, USP, São Carlos, SP (2009)
6. Novakouski, M., Lewis, G.: Interoperability in the e-Government Context. Technical Note, CMU/SEI-2011-TN-014, Carnegie Mellon University, Pittsburgh, PA. Retrieved from http://www.sei.cmu.edu/reports/11tn014.pdf (2012)

7. UNDP (United Nations Development Programme). e-Government Interoperability: Overview. http://goo.gl/bfK0ix (2007)

8. Cestari, J. M. A. P., Loures, E. R., Santos, E. A. P., Lezoche, M.: A Research Strategy for Public Administration Interoperability Assessment, In: Industrial & Systems Engineering Research Conference (ISERC), http://goo.gl/z26gDG, Montreal, Canada (2014)

9. Ray, D., Gulla, U., Dash, S. S., Gupta, M. P.: A critical survey of selected government interoperability frameworks. Transforming Government: People, Process and Policy, 5(2), 114-142 (2011)

10. Solli-Sather, H.: A framework for analyzing interoperability in electronic government. International Journal of Electronic Finance, 5(1), 32-48 (2011)

11. Gottschalk, P.: Maturity levels for interoperability in digital government. Government Information Quarterly, 26(1), 75-81 (2009)

12. Zutshi, A., Grilo, A., Jardim-Goncalves, R.: The business interoperability quotient measurement model. Computers in Industry, 63(5), 389-404 (2012)

13. National Institute of Standards and Technology (NIST). Integration Definition for Function Modeling (IDEF0). http://www.idef.com/pdf/idef0.pdf (1993)

14. Saaty, R. W.: The analytic hierarchy process—what it is and how it is used, Mathematical Modelling, volume 9, issues 3–5, 161–176 (1987)

15. Nonaka, I.: A dynamic theory of organizational knowledge creation. Organization Science, vol. 5, n. 1, 14-37 (1994)

16. Swan, J., Newell, S., Scarbrough, H., Hislop, D.: Knowledge management and innovation: networks and networking. Journal of Knowledge management, 3 (4), 262-275 (1999)

17. Von Krogh, G.: Care in knowledge Creation. California Management Review, Vol. 40, n. 3, 133-153 (1998)

18. Keeling, J.: Development of Systematic Knowledge Management for Public Health: A Public Health Law Ontology. PhD Thesis, Graduate School of Arts and Sciences, Columbia University (2012)

19. Vocabulary Dictionary, http://www.vocabulary.com/

20. Cambridge Dictionaries Online, http://dictionary.cambridge.org/us

21. Oxford Dictionaries, http://www.oxforddictionaries.com/us

22. Object Management Group, http://www.uml.org/

23. Deschamps, F.: Proposal for the systematization of enterprise engineering contributions: guidelines for enterprise engineering initiatives. PhD Thesis, PUCPR, Curitiba (2013)

24. MaxQDA, http://www.maxqda.com/

25. Pierce, J. R.: An introduction to information theory: symbols, signals & noise by John Robinson Pierce, Dover Publications, Subsequent edition, 336 pages (1980)

26. Stanford POS Tagger software, http://nlp.stanford.edu/software/tagger.shtml

27. Toutanova, K., Klein, D., Manning, C., Singer, Y.: Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In Proceedings of HLT-NAACL, pp. 252-259 (2003)

28. Lezoche, M., Panetto, H., Aubry, A.: Conceptualisation approach for cooperative information systems interoperability, in ICEIS 2011 (2011)

29. Vaidya, O. S., Kumar, S.: Analytic hierarchy process: An overview of applications. European Journal of operational research, 169 (1), 1-29 (2006)

30. Liao, Y., Loures, E. R., Panetto, H., Canciglieri, O.: A Novel Approach for Ontological Representation of Analytic Hierarchy Process. Adv. Materials Research (Online) (2014)

31. Guédria W., Naudet Y., Chen D.: A Maturity Model Assessing Interoperability Potential. In Halpin, T., Nurcan, S., Krogstie, J., Soffer, P., Proper, E., Schmidt, R., Bider, I. (eds.), Heidelberg, vol. 81, Springer, 276–283 (2011)