Web X-ray:
developing and adopting Web best practices in Enterprises

Reinaldo FERREIRA
IESF
Valadares, Vila Nova de Gaia, Portugal;
Planeta Virtual
Porto, Portugal

ABSTRACT

The adoption of Semantic Web technologies constitutes a promising approach to data structuring and integration, both for public and private usage. While these technologies have been around for some time, their adoption is behind overall expectations, particularly in the case of Enterprises. Having that in mind, we developed a Semantic Web Implementation Model that measures and facilitates the implementation of the technology. The advantages of using the model proposed are two-fold: the model serves as a guide for driving the implementation of the Semantic Web as well as it helps to evaluate the impact of the introduction of the technology.

The model was adopted by 19 enterprises in an Action Research intervention of one year with promising results: according to the model's scale, in average, all enterprises evolved from a 6% evaluation to 46% during that period. Furthermore, practical implementation recommendations, a typical consulting tool, were developed and adopted during the project by all enterprises, providing important guidelines for the identification of a development path that may be adopted on a larger scale.

Meanwhile, the project also outlined that most enterprises were interested in an even broader scope of the Implementation Model and the ambition of a "All Web Technologies" approach arose. One model that could embrace the observable overlapping of different Web generations, namely the Web of Documents, the Social Web, the Web of Data and, ultimately, the Web of Context. One model that could combine the evaluation and guidance for all enterprises to follow. That's the goal of the undergoing "Project Web X-ray" that aims to involve 200 enterprises in the adoption of best practices that may lead to their business development based on Web technologies.

This paper presents a case of how Action Research promoted the simultaneous advancement of academic research and enterprise development and introduces the framework and opportunities related to Web Science and how the different and overlapping Web generations offer a relevant opportunity for enterprise development. It also discusses the learnings from bringing academic research to the enterprises and how it leads to further development of both the research and the utility of the results. Finally, it introduces the fundamentals for a new and broader Web Technology Implementation Model that may become a reference for Enterprise adoption.

Keywords: Semantic Web, Linked Data, implementation model, Enterprise Web

1. INTRODUCTION

Over the last twenty years, the Web evolved from a laboratory where it consisted in merely an idea to a ubiquitous and universal environment, accessible from basically everywhere, with information from basically everything. Yet, the "Web of Documents" is far from being considered a fully structured version of the Web. The thinking underlying the Semantic Web, providing a common framework that allows data to be shared and reused across application, enterprise and community boundaries [1], puts forward the introduction to data access as well as to data structure and meaning, and enables a more integrated robust and practical environment. It is believed that the next two decades will provide enough time to spread the "Web of Data", being that the expectation of the scientific community.

The introduction of the Semantic Web involves the progressive transformation of the Web based on hyperlinks between documents, the base for its first generation, in the Web based on hyperlinks between data or information, giving place to a Web scale distributed database [2]. This very large database is still in its early days, although a rather significant number of examples that allow to evaluate its potential can be found, since some research efforts and government policies for data publication have already produced satisfactory results [3]. Despite these contributions, Semantic Web technologies usage within the enterprise community is still a rather unexplored theme and in an early adoption phase [4], [5]. The reasons for this reduced level of adoption may point to several difficulties, such as the homogenization and validation of data sources, the definition of knowledge rules and borders that allow to relate data in a uniform way, the analysis of too complex examples, the availability of low cost technological capacity to allow its implementation, the availability of development tools, the recruitment of experienced professionals, the diffusion of success stories and the adoption of a paradigm shift in modelling, design and development [4], [7].

Acknowledging these shortcomings, a model for Semantic Web implementation in Enterprises was developed with two main goals. The first goal is to facilitate the introduction of the technology in organizations with different characteristics and motivations, acting as a guide and providing a roadmap for a quicker and more intensive adoption. The second goal is to evaluate the impact of the introduction of the technology in the applications used in these organizations and in the tasks performed by their users [8], [9].

The economic dimension is one of the most important dimensions of the World Wide Web. It has been so in its first generation, with "long tail" enterprises approaching the global market of customers and products, until the targeting of the smallest segment. It has also been so with the growth of social networks, and these being explored as a competitive advantage to connect people to enterprises, brands and products. And it will eventually be so in its third generation, with enterprises actively participating in the construction of the Semantic Web, the Web of Data. It constitutes certainly a great and broad research opportunity, in many subjects, allowing the space for the creation of the Web Science [10], [11].

Semantic Web technology may transform enterprise software, contribute to the emergence of new business models and reduce costs in areas like data integration, master data management and enterprise information management [7]. Enterprise internal networks based on Linked Data principles constitute a particular subset of Semantic Web technologies that, amongst other benefits, may substantially reduce information integration costs, such as in the integration of information about a product, supplier, materials, legal, market and finance data or other internal and external data sources [12]. However, the growing adoption of Semantic Web technologies and Linked Data principles raise the question of which applications may be developed to take advantage of this potential. The answer may be obtained from identifying the areas where these technologies and principles may constitute a distinctive contribution, when compared with traditional technologies [13].

Meanwhile, the compromise between the issues of computational effort and flexibility tends to favor the latter, in line with Moore's law projection, since additional computing power leads to less concerns about simplification or optimization. Simultaneously, the economy has benefited from technological innovation and became a highly competitive environment where speed and flexibility often play a more important role than robustness and trust. The economic validity of current data normalization models can therefore be questioned in comparison to the flexibility and universality promises of the Semantic Web [14].

The Semantic Web development has been guided by the way different communities envisage its evolution, considering their specific areas of research. One approach is that of semantic annotation, addressing the large volumes of data available on the Web and using different techniques to originate structured data. Another approach is that of data repositories, starting from pre-defined structures that are updated and interlinked with additional structures. Finally, the approach that puts forward the Semantic Web as an agent platform, with applications combining different data sources and, ultimately, executing actions in replacement of individuals [15].

2. THE ENTERPRISE SEMANTIC WEB IMPLEMENTATION MODEL

Acknowledging the different issues to consider in the adoption of Semantic Web technologies in Enterprises, it is envisaged that a model for Semantic Web implementation in Enterprises will constitute a useful approach to guide enterprises in the introduction of the technology as well as in the evaluation of its' impact in the applications used and tasks performed by users in these organizations [8].

The model developed is called "e-swim", an acronym for Enterprise Semantic Web Implementation Model, and is based in four dimensions, where each dimension considers different requirements for the technology implementation and, simultaneously, serves as a guide to identify the desired continuous evolution through subsequent steps [9].

Figure 1 illustrates the e-swim model and its four dimensions: Adoption, Provenance, Accessibility and Activities. The following sub-sections describe the rationale for each dimension.

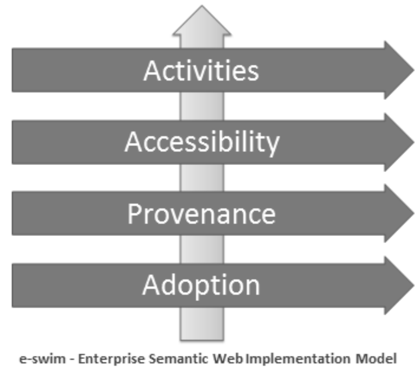

Figure 1 – The e-swim Model

Adoption

The purpose of this dimension is to determine the degree of preparation of the Enterprise to adopt Semantic Web technologies.

The main technological adoption models for enterprise application are the Innovation Diffusion Model and the Technology-Organization-Environment Model [16]. According to the Innovation Diffusion Model [17], organizational innovation is essentially dependent on Leadership, Organizational Structure and Openness and is adopted according to a normal distribution of organizations that includes Innovators, Early Adopters, Early Majority, Late Majority and Laggards. The Technology-Organization-Environment Model proposes three issues of the enterprise context that influence technological innovation, namely Technology, Organization and Environment. By combining both approaches, the e-swim Model considers the following features for the dimension:

1) Technology, including opportunities of technology usage in the organization, in this case, Semantic Web technologies and its applications;
2) Organization, referring to internal organization relations and including:
    a. Leadership, attitude towards change from the top management
    b. Structure, relations between people in the organization
3) Exterior, about the external framing of the organization, including:
    a. Interface, openness of the organization to the outside
    b. Environment, players in the space where the organization is located

These features are summarized in a graphical form in Figure 2.

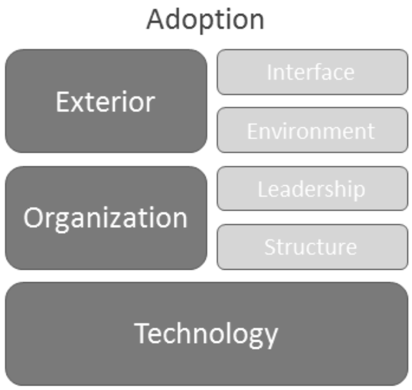

Figure 2 – The technology Adoption dimension

The planned intervention in enterprises takes these features into account and, through observation and interrogation, tries to quantify the position of each Enterprise in the path from a lower to a higher technology adopter.

Provenance

This dimension of the Implementation Model considers data provenance as a determinant factor. Enterprise innovation and competitive advantage depend entirely of its capacity to deal with a constant and always growing information flow. Consequently, information integration efforts must follow that growth. Semantic Web technologies usage in those integration efforts may increase substantially the return, reducing integration costs and increasing subsequent benefits [12]. Opening public data to citizens represents an increasing democratic transparency, possible due to technological availability. In that area, the efforts of the American and British governments, among others, already led to a broad availability of data sources with wide usage possibilities. The need to explore these data sources reveals itself primarily as the possibility to explore wealth sources [18].
Therefore, the purpose of this dimension, as illustrated in Figure 3, is to identify those data sources, according to:

1) applications topology, namely Web Sites, Extranets, Intranets, Web Applications or Web Services
2) location: outside or inside the enterprise

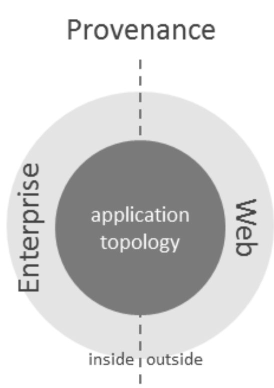

Figure 3 – Enterprise Data Provenance

The planned intervention aims to identify the data sources in the context of the enterprise Information Systems and to quantify their respective usage.

Accessibility

The common user perception about data available in the Web is largely influenced by the format of these data. However, the structure of these data can influence their subsequent usage. The path to the universal availability of data, as shown in Figure 4, was clearly identified by Tim Berners-Lee with a five stars classification or five steps of evolution [19]:

1) available: just having data available in the Web
2) formatted: through proprietary formats
3) open: with open formats
4) semantic: via semantic standards
5) interconnected: with data hyperlinks.

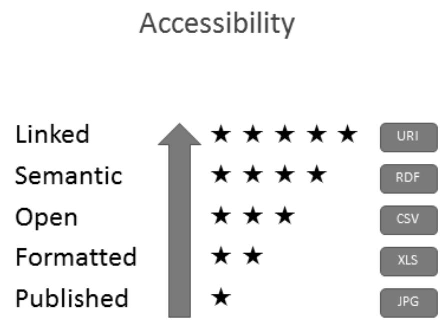

Figure 4 –Data Accessibility, adapted from [19]

The planned intervention will determine the degree of openness of data sources and consequently quantify their suitability towards Semantic Web standards.

Activities

Information Systems have been classified according to different approaches, some with a broader scope, others with more specific purposes [20]. Despite the high number of efforts, the issue is not exhausted and, in this particular case, it is important to find a classification that positions the Web as a base environment, in alternative or in complement to more traditional classifications. With several studies related with Web usage, Tom Heath's work introduces important clarifications and a purpose oriented classification [13]. According to this classification, user activities may be instantiated in the following categories:

1) Locating: look, find
2) Exploring: gather, research
3) Grazing: navigate, browse, follow
4) Monitoring: monitor, check, detect
5) Sharing: distribute, collaborate
6) Notifying: state, inform, communicate
7) Asserting: opinion, suggestion
8) Discussing: comment, respond
9) Evaluating: assess, analyze
10) Arranging: combine, negotiate
11) Transacting: transfer, pay

Figure 5 shows the space for Semantic Web technologies implementation according to the degrees of diversity and conceptualization of the tasks performed.

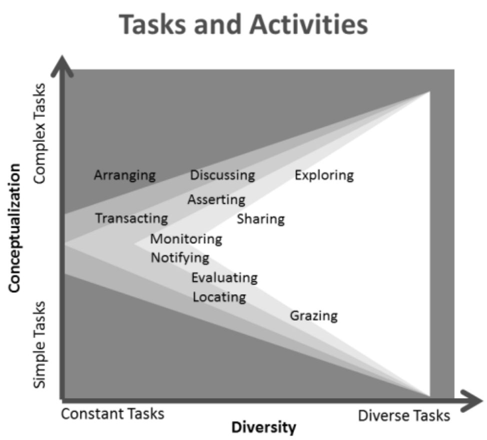

Figure 5 - Activities over Conceptualization and Diversity, a new assessment from [15]

The conceptualization and diversity degrees of the tasks performed influence technology implementation feasibility. Simple and repetitive tasks are those naturally already satisfied by traditional software applications and that will less benefit from Semantic Web technologies. As task conceptualization increases, opportunities arise for the technology, but in the conceptualization threshold, when tasks are hardly typified and demand a person's creative intervention, the effort for their implementation is simply higher than that of just executing the tasks. Low diversity tasks are once again easily supported by traditional software and in the threshold of diversity it will rarely be worth systematizing tasks to be supported by the technology. Hence, the ideal space for Semantic Web technology implementation will be that of some conceptualization or diversity that make tasks too complex or too diverse for traditional software [15].

Field experience was an opportunity to verify this possibility that revealed new insights. The first activities to be observed, associated with early adoption phases, were the most diverse tasks, independently of the conceptualization level, either from being supported by the technology or from giving support to user activities. The Semantic Web implementation space appears tall and narrow at diverse tasks and widens as implementation advances and diversity reduces.

This dimension addresses the wide area of the enterprise Information Systems and introduces a fit to action approach adequate to an Action-Research effort. The planned intervention will determine what are the tasks performed that may better incorporate Semantic Web technologies and quantify their usage through the Enterprise.

3. VALIDATION THROUGH ACTION RESEARCH

Following the development of the model, an Action Research intervention was designed and executed next to 19 enterprises with the primary goal of adopting the proposed model. The intervention was divided in two cycles, the first focused on each enterprise specific environment and the second focused on reapplying learnings across different environments. The average intervention was of 40 hours per enterprise and lasted about one year [21].

In order to better understand the technology adoption, a Measurement Tool was developed. The tool is based in the Implementation Model dimensions and translates observations made in each enterprise into several scales. The dimensions are then averaged into a final overall measurement. The Measurement Tool was used in the beginning and in the end of the intervention, producing two measurements that represent both instant values and a progression between the two moments. In average, all enterprises evolved from a 6% evaluation to 46% during that period.

The evolution observed resulted from practical implementation recommendations, a typical consulting tool. Starting from the observations in the different enterprises, a set of 12 different Recommendations was developed. Each company adopted an average of 3.6, in a total of 68 Recommendations adopted. Individual Recommendations were adopted at different rates, ranging from adoption in 1 to 14 enterprises. The interesting rates of adoption and reapplication validate the Implementation Model and confirm the option for an Action Research intervention.

The major contribution of this research was to conclude that the fast and general adoption of Semantic Web technologies in enterprises can be achieved mainly through Web Sites or Web Services implementing content management, search or integration solutions, used in diversified tasks, mainly in researching and sharing information activities.

Nevertheless, not only the Action Research intervention contributed to the technological development and innovation in the enterprises, but also brought up new Research challenges. Additional research opportunities include:

1) Find and incorporate additional Recommendations in the process (with some of them already appearing as strong candidates)
2) Make training available in a comprehensive set of technologies and applications, under the wide umbrella of e-business
3) Develop a package of services centered in the Webmaster, as a one stop shop for all things Web
4) Determine how Management Systems may evolve with the growing usage of Web technologies
5) Understand the cause and effect of different types of organizations, namely Enterprises and Governments, in the rate of Web technology adoption
6) Search for additional reapplication rates and, consequently, better robustness and return on investment, through proposing the implementation model in larger tests
7) Further determine the correlation between Web technology adoption and enterprise innovation
8) Extension of the Implementation Model to the different Web technologies and generations

This Action Research experience confirmed some warnings issued beforehand:

1) Clear and upfront expectations, outlined in a formal protocol, may be appropriate for scope definition but may not be flexible enough for fast pace environments: solutions implementation, vendor availability and employee training were frequently responsible for most of the feasible adoption
2) Researchers/Consultants may also prefer to be more in control of the process than in pursuit of the optimum or neutral result: some measurements resulted less from the rational than from the degree of optimism of the intervenients
3) Enterprises/Organizations may change priorities up to the point where one intervention may become irrelevant and therefore abandoned: the worst scenario is not when data does not confirm the hypothesis but when there's no data at all

intermediate step towards Web 5.0, whatever that generation will be. Given that Web 3.0 is neither mature enough nor widely spread to be exclusively responsible for the current generation, most likely Web 4.0 is here and is the Web of Context.

4. A NEW WEB TECHNOLOGY IMPLEMENTATION MODEL

Whether Enterprise or Academic Research produce more innovation may be debatable. Enterprises will look for more practical results, focusing on the short term and higher success rates, while academic research will typically look for theoretical results that are more ambitious and focused on the medium term [22].

Though, innovation is certainly a common ground for shared interest. A Web Technology Implementation Model should be able to incorporate high pace innovation while it rescales measurements and stresses enterprise effort. A brief overview of Gartner's Hype Cycle for Emerging Technologies [23] may throw some light on what's to come, featuring:

1) Virtual Personal Assistants, based on Machine Learning, Speech technology and other recent disruptions, providing feedback to user context
2) Internet of Things, or even the Web of Things, depending on the weight of Web technologies in a probably soon to be settled ubiquitous platform that supports billions of almost autonomous devices, like sensors, wearables or appliances
3) Autonomous Vehicles, where an unquantifiable amount of data is needed to support the largest grid of vehicles, with Web technologies (e.g. URI) well positioned to be strongly connected to this revolution

We noticed that Web generations are considerably overlapping each other, coexisting in time, as shown in Figure 6.

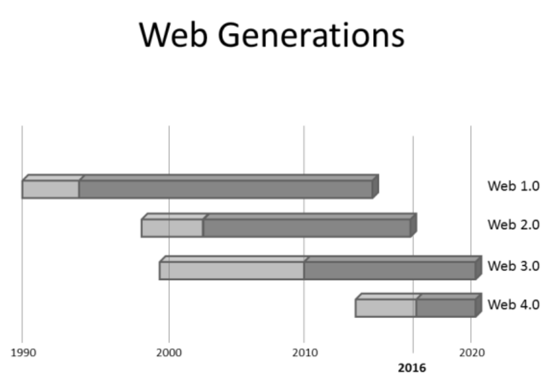

Figure 6 – Web Generations

We also noticed that odd generations, Web 1.0, the original Web or Web of Documents, and Web 3.0, the Semantic Web or Web of Data, are technology driven. So the obvious hypothesis would be checking if even generations are people (or usage or society) driven: that's the case of Web 2.0, the Social Web. Likewise, the Semantic Web Implementation Model is based in two technology and two people oriented dimensions, which also produces important clues for further exploitation.

Considering that Web 4.0 is the result of the intense usage of Web based applications for a growing number of user activities, it is a bit more than a Web OS in the sense its shape is largely undetermined, to be more of an individually adaptable framework, in the sense that activities are executed for and from the context they are involved. It may not be surprising that Web 4.0 is still a bit ambiguous and may be recognized only as an intermediate step towards Web 5.0, whatever that generation will be. Given that Web 3.0 is neither mature enough nor widely spread to be exclusively responsible for the current generation, most likely Web 4.0 is here and is the Web of Context.

The enterprise influence in the research process was determinant in following this research path. It were enterprise expectations and challenges that outlined the need to develop solutions to support the implementation of Recommendations and that introduced new research challenges aimed at further innovation. Being so, it's only natural that the next step in this process includes enterprises in a larger scale.

The undergoing "Project Web X-ray" aims to involve 200 enterprises in the adoption of best practices that may lead to business development and innovation based on Web technologies. With an approach that descends from its' predecessor, a tenfold larger sample may contribute to further understand and develop the Web Technology Implementation Model into one model that can combine the evaluation and guidance for all enterprises to follow. This goal can be achieved if the model can be reused by Consultants on a wider scale. It's then up to Researchers involved to answer the following questions in the upcoming years:

1) Does the Enterprise Semantic Web Implementation Model maintain its coherence when used in and with larger groups?
2) What is a wider set of Recommendations that may assist Enterprise development and innovation through Web technologies?
3) How can all different, existing and upcoming technologies incorporate into an adaptable Web Technologies Implementation Model?

5. CONCLUSIONS

Action Research promotes the simultaneous advancement of academic research and enterprise development and will be the basis for the upcoming Project Web X-ray as it was of its predecessors. Bringing academic research to the enterprises leads to further development of both the research and the utility of the results. It is a process where innovation is triggered and control loses ground to uncertainty. But it is also a process where visible results are within immediate reach of all intervenients.

Project Web X-ray introduces in Enterprises the framework and opportunities related to Web Science and how the different and overlapping Web generations offer a relevant opportunity for development and innovation. Capitalizing on previous Research and the wide acceptance of the Enterprise Semantic Web Implementation Model, it aims to develop a new and broader Web Technology Implementation Model that may become a reference for Enterprise adoption in the emerging Web of Context.

9. REFERENCES

[1] W3C, "W3C Semantic Web Activity Homepage," 2013. [Online]. Available: http://www.w3.org/2001/sw/. [Accessed: 30-Mar-2015].

[2] T. Heath and C. Bizer, "Linked Data: Evolving the Web into a Global Data Space," in *Proceedings of the 17th international conference on World Wide Web*, 2011, vol. 1, no. 1, pp. 1–136.

[3] C. Bizer, T. Heath, K. Idehen, and T. Berners-Lee,

"Linked data on the web (LDOW2008)," in *WWW2008 Workshop on Linked Data on the Web*, 2008, vol. 2008, no. September, pp. 1265–1266.
[4] Z. Ahmed and D. Gerhard, "Role of Ontology in Semantic Web Development," *Knowl. Manag.*, 2010.
[5] W. Kuhn, "Modeling vs encoding for the Semantic Web," *Semant. Web J.*, vol. 1, pp. 11–15, 2010.
[6] S. S. Kang, J. Y. Yang, S. K. Lee, K. H. Gong, J. S. Myung, S. C. Park, and S. G. Lee, "An Enterprise Strategy for Semantic Technology Adoption," in *5th International Conference on Information Technology and Applications*, 2008.
[7] J. Pollock and Oracle, "A Semantic Web Business Case," 2008.
[8] R. Ferreira and I. Seruca, "Modelo de Implementação da Web Semântica nas Empresas," in *Atas da 8ª Conferência Ibérica de Sistemas e Tecnologias de Informação (CISTI'2013), Lisboa, Portugal, 19-22 Junho 2013*, 2013, pp. 615–620.
[9] R. Ferreira and I. Seruca, "e-swim - Enterprise Semantic Web Implementation Model Towards a systematic approach to implement the Semantic Web in Enterprises," in *Proceedings of the 16th International Conference on Enterprise Information Systems (ICEIS 2014), Lisbon, Portugal, 27-30 April 2014*, 2014, pp. 653–658.
[10] R. Ferreira and I. Seruca, "A Ciência da Web : oportunidades de investigação," in *Atas da 13ª Conferência da Associação Portuguesa de Sistemas de Informação, Évora, Portugal, 4-5 Outubro 2013*, 2013, pp. 13–31.
[11] T. Berners-Lee, J. Hendler, W. Hall, N. Shadbolt, and D. J. Weitzner, "Computer science. Creating a science of the Web.," *Science (80-.).*, vol. 313, no. 5788, pp. 769–771, 2006.
[12] K. Janowicz and P. Hitzler, "Creating knowledge out of interlinked data," *Semant. Web*, vol. 1, pp. 97–104, 2010.
[13] T. Heath, "A taskonomy for the Semantic Web," *Semant. Web J.*, vol. 1, pp. 75–81, 2010.
[14] T. Segaran, C. Evans, and J. Taylor, *Programming the Semantic Web*. Sebastopol, CA, USA: O'Reilly, 2009.
[15] J. Domingue, D. Fensel, and J. A. Hendler, Eds., *Handbook of Semantic Web Technologies*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.
[16] T. Oliveira and M. Martins, "Literature Review of Information Technology Adoption Models at Firm Level," *Electron. J. Inf. Syst. Eval.*, vol. 14, no. 1, pp. 110–121, 2011.
[17] E. M. Rogers, *Diffusion of innovations*, vol. 65, no. 5. Free Press, 1995.
[18] C. L. Koumenides, M. Salvadores, H. Alani, and N. R. Shadbolt, "Global Integration of Public Sector Information," in *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line*, 2010.
[19] T. Berners-Lee, "Linked Data Design Principles," 2006. [Online]. Available: http://www.w3.org/DesignIssues/LinkedData.html. [Accessed: 30-Mar-2015].
[20] F. C. Lopes, M. P. Morais, and A. J. Carvalho, *Desenvolvimento de Sistemas de Informação*. Lisboa, Portugal: FCA, 2005.
[21] R. Ferreira, "A Web Semântica nas Empresas: Proposta de um Modelo de Implementação," Universidade Portucalense, 2015.
[22] J. Cardoso, J. Miller, J. Su, and J. Pollock, "Academic and Industrial Research : Do their Approaches Differ in Adding Semantics to Web Services ?," *Lect. Notes Comput. Sci.*, vol. 3387/2005, pp. 14–21, 2008.
[23] H. Levy, "What's New in Gartner's Hype Cycle for Emerging Technologies, 2015 - Smarter With Gartner," 2015. [Online]. Available: http://www.gartner.com/smarterwithgartner/whats-new-in-gartners-hype-cycle-for-emerging-technologies-2015/. [Accessed: 08-Jan-2016].