Semantic Repository in Internet Infrastructure knowledge domain: Methodology Position Paper

Julião Braga[1], **Nizam Omar**[1]

[1]Department of Electrical Engineering – Mackenzie Prebisterian University
01302-090 Rua da Consolação, 930 - Consolação - São Paulo – SP – Brazil

juliao@braga.eti.br, nizam.omar@mackenzie.br

Abstract. *This is a position paper that proposes the development of a methodology to build Semantic Repositories in the Internet Infrastructure domain.*

1. Introduction

This work aims to establish an appropriate methodology to build a *Semantic Data Repository on Internet Infrastructure*, *SDRII*, which is defined as an information base, geographically distributed and ubiquitous, whose contents are data, with their equivalent semantic representations, procedures and vocabulary. The *data* are those which become available as outputs of historical or active processes from the *IETF*, *Internet Engineering Task Force*, either aggregated or operationally associated institutions. The *semantic representations* are all techniques, resources and facilities available in the context of Semantic Web that allows *SDRII* to transform into a knowledge base or ontology. *Procedures* are components that relate to and / or support themselves, proactively and mutually, making *SDRII* a smart knowledge base, adaptive and autonomous. *Vocabulary* or *terms* in this context has the same meaning as given by W3C[1] which broadly define the concepts and relationships used to represent, describe and disclose the contents of one or more *SDRII*, exchanging information with each other or with other repository terms, specialized or not, as the *Dublin Core Metadata Initiative*[2]. This exchange of information will be accomplished through *URIs* (*Universal Resource Identifiers*), described in RFC3986, [Berners-Lee et al. 2005]. This entire structure together enables a *SDRII* be understood by machine, without human intervention.

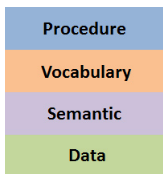

Figure 1: O *SDRII* in layers

Figure 1 presents a graphical view, in layers of a *SDRII* and Figure 2 situates *SDRII* in a wider context, such as the Internet which provides the basic input: the *connectivity*.

[1]http://www.w3.org/standards/semanticweb/ontology
[2]http://dublincore.org/

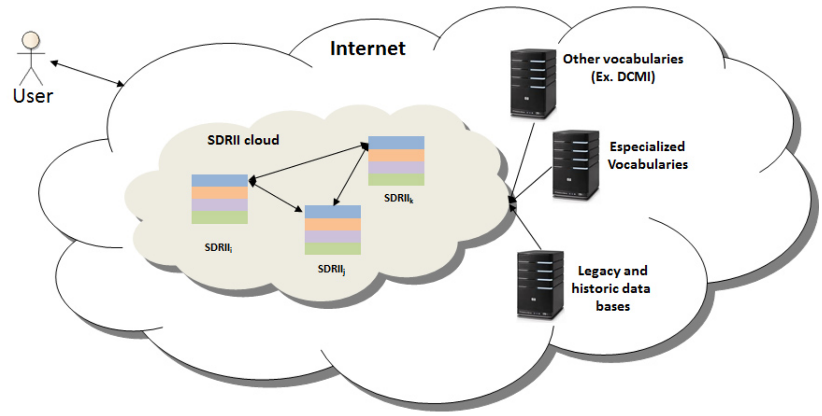

Figure 2: SDRII connectivity model in the Internet.

In knowledge where Semantic Web has not yet been applied extensively (as in the case of the Internet Infrastructure), and particularly under the influence of legacy databases, the work will point to characterize and analyze the complexity associated to the resources available for construction of *SDRIIs* and proposes a methodology intended to mitigate the difficulties related to the tools and techniques applicable, currently available, as mentioned, among others, by [IBM 2013], regarding the treatment of *Big Data*.

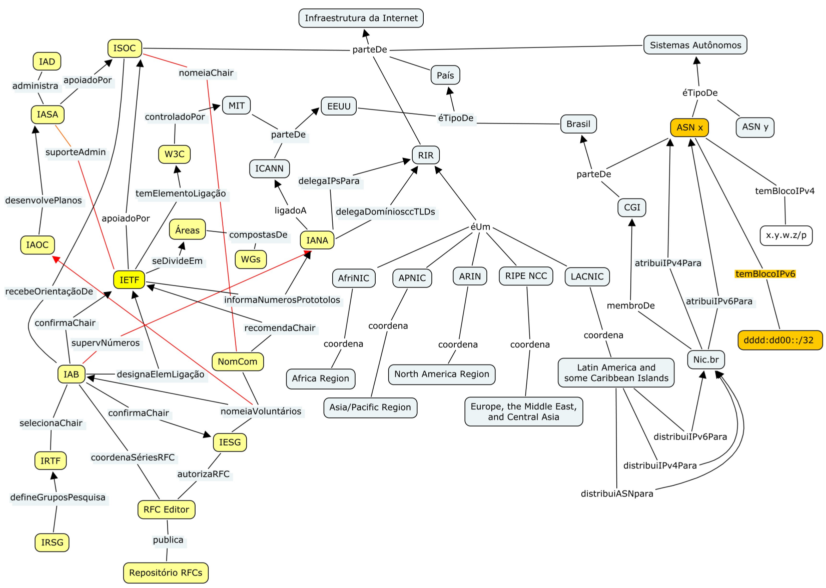

Figure 3: Partial view of the data repositories in the Internet Infrastructuture.

The databases currently available - legacy, heterogeneous, reformulated, integrated or merged - have data built on traditional storage structures, i.e., under the management of the database with several orientations – relational, graph, columnar, key-value, documents, hybrids – described as *NoSQL*, [Redmond and Wilson 2012], and even in ASCII mostly associated with metadata. Many of them, for some time, add the XML information, [W3C 2014], which allow an program-to-program (or computer-to-computer), the understanding of creating facilities on the effort required to capture and manipulate of data. This is the case of most of repositories maintained by the IETF RFC Editor[3], but

[3] `http://www.rfc-editor.org/`

not enough to establish a *SDRII*.

A partial view on the Internet Infrastructure is shown in Figure 3, illustrating a set of institutions which produce data and their relationships. Following the path [IETF–>IANA–>RIR–>Nic.br–>ASN x], is possible to observe how the information is distributed across multiple repositories not providing any effective integration.

2. Project Structure

The project will be divided into six steps, not necessarily interdependent but described below.

Step 1: This step is to present the theoretical basis for the application of Semantic Web in the domain of Internet Infrastructure identifying restrictions and current difficulties that prevent the effective and efficient use of existing information by system administrators, in addition to characterize the various existing repositories of unstructured, dispersed and not integrated data. A typical and interesting to observe example is the *whois*, [O'Reilly 2011], a classic not integrated and accessible syntactically dispersed repository. Other repositories should require special attention, with the application of data mining techniques, [Tan et al. 2009], to ensure reliability of the data, as is the case of simple lists attendance (*attendance lists*), for the IETF meetings[4], simply displayed on a Web page created without the additional use of more recent recommendation resources, such as RDFa, [Herman et al. 2013] which would allow people to relate with I- Ds, RFCs, RIRs, location and participations, among other attributes .

Step 2: This step focus on interoperability issues, syntactic, structural and semantic repositories, bothering to formally characterize the *SDRII*.

Step 3: This step focus on the *Upper Level Ontology*, where the reuse of vocabularies stimulates cooperation in the development of *Semantic Repositories*, with the implementation of *Suggested Upper Merged Ontology* (*SUMO*), [Pease et al. 2002] and [Peace 2011b], justifying the reason for this choice, the opportunities and facilities arising mainly the feasibility of its integration with WordNet ([Pease et al. 2012], [Borra et al. 2010], and [Niles and Pease 2003]); and use of a language of high expressive semantics, *Standard Upper Ontology Knowledge Interchange Format* (SUO-KIF), [Pease 2009]. A hierarchy of expressiveness of several languages available or more observed, can be seen in Figure 4. The SUMO will be used as a *training model*.

Step 4: This step research will focus on the possibilities of *cooperation* between individuals and groups and, initiate the use of tools such as Apache Hadoop, [Apache Foundation 2014] and [Zikopoulos et al. 2012].

[4]`https://www.ietf.org/registration/ietf89/attendance.py`

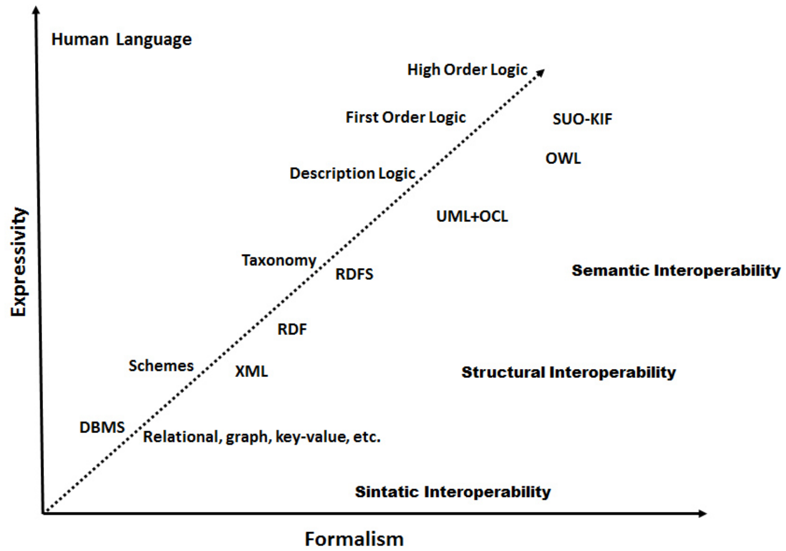

Figure 4: Languages: expressivity and formalism. Adapted from: [Peace 2011a] e [Peace 2011b], page 33.

Step 5: This step will focus on the set of available alternatives by comparing them with the *SUMO*. It is true that the *SUMO* has a language, the *SUO - KIF*, with high expressiveness. However there is a disadvantage when it requires the implementation to use it effectively. Other languages such as OWL, [Hitzler et al. 2009] and [Patel-Schneider et al. 2004], are traditional and generally sufficient for expressing ontologies without running the risk of exceeding the allowable limits of logical reasoning. One of the alternatives studied is the *JOINT* in [Holanda et al. 2013], which admits the possibility of adjustments and changes, appropriate to the *SDRII*, as proposed in this project, and to facilitate the use of traditional research tools on semantic repositories in general, as the *SPARQL*, recommended by W3C in [Harris and Seaborne 2013]. Additionally, *JOINT* is a rich experience, since it uses an array of resources, including OWLIM, described in [Bishop 2011], and in its generalization allows prospecting ontologies constructed by editors as the *Protégé*[5].

Step 6: This step will address the methodology itself, oriented to satisfy the demand of establishment of *SDRIIs* anticipating up facilities to creation, adaptation and clustering of traditional data repositories and its maintenance of its traditional distributed and functional features, but integrated, where it applies. In summary, the *Step 6* is reserved for consolidation and basis of the final results of the research conducted in the previous steps.

This step will also be responsible for the identification of future contributions of this research project, including the resulting methodology. Among other contributions, it is worth emphasizing those related to the consolidation of ideas around the infrastructure and common requirements for implementation of *autonomic network* exposed in recent references, which can be seen in the history of NMRG (IRTF) mailing list[6], as the observations of [Trammel and Happe 2013], although such references are oriented to protocols.

[5] `http://protege.stanford.edu`

[6] `http://www.ietf.org/mail-archive/web/nmrg/current/maillist.html`

The notions of *autonomic networks* is derived from a manifesto drawn up by IBM in 2001, which sets a challenge to create self-managed computing systems, featuring elements of the so called *autonomic computing system*, described in [Kephart and Chess 2003]. The *autonomic elements* that encapsulate the *RSDIIs* to ensure the interconnections shown in Figure 2 features the definition of *RDSII* given in Section 1 should follow in principle, the proposal contained in [Parashar and Hariri 2005], when the first experiments, and away to the most recent recommendations from the literature and results for any others experiences. Concerns the implementation of *autonomic computing system* will be restricted to the behavior records of *RDSIIs*, or *log*, structured in ontologies, which allow inference to the purposes of self-management / self-regulation.

3. Additional Information

There is a Web site with updated information about the project: `http://ws.org.br`.

References

Apache Foundation (2014). Hadoop. Disponível em `http://hadoop.apache.org/`. Acessado em 02/02/2014.

Berners-Lee, T., Fielding, R., and Masinter, L. (January 2005). Uniform Resource Identifier (URI): Generic Syntax. RFC3986. `https://tools.ietf.org/rfc/rfc3986.txt`. (Obsoletes RFC2732, RFC2396, RFC1808) (Updates RFC1738) (Updated-By RFC6874) (Also STD0066) (Status: INTERNET STANDARD) (Stream: IETF, WG: NON WORKING GROUP). Acessado em 27/02/2014.

Bishop, B. (2011). OWLIN Primer. `http://owlim.ontotext.com/display/OWLIMv43/OWLIM+Primer`. Acessado em 25/02/2014.

Borra, A., Pease, A., Roxas, R., and Dita, S. (2010). Introducing Filipino WordNet. In Bhattacharyya, P., Fellbaum, C., and Vossen, P., editors, *Principles of Construction and Application of Multilingual Wordnets: Proceedings of the 5th Global WordNet Conference*, pages 306–310. Narosa pub, Mumbai, India.

Harris, S. and Seaborne, A. (2013). SPARQL 1.1 Query Language. `http://www.w3.org/TR/sparql11-query/`. Acessado em 31/10/2013.

Herman, I., Adida, B., Sporny, M., and Birbeck, M. (2013). RDFa 1.1 Primer - Second Edition: Rich Structured Data Markup for Web Documents. `http://www.w3.org/TR/xhtml-rdfa-primer/`. Acessado em 18/01/2014.

Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P. F., and Rudolph, S. (2009). OWL 2 Web Ontology Language Primer. `http://www.w3.org/TR/2009/REC-owl2-primer-20091027/`. Acessado em 25/02/2014.

Holanda, O., Bittencourt, I. I., Isotami, S., Elias, E., and Bandeira, J. (2013). A Tool for Efficient Development of Ontology-based Applications. `http://ceur-ws.org/Vol-938/ontobras-most2012_paper10.pdf`. Acessado em 25/02/2014.

IBM (2013). Using semantic enrichment to enhance big data solutions. `http://www-01.ibm.com/software/ebusiness/jstart/semantic/`. Acessado em 04/01/2014.

Kephart, J. and Chess, D. (2003). The vision of autonomic computing. *Computer*, 36(1):41–50.

Niles, I. and Pease, A. (2003). Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. In *Proceedings of the IEEE International Conference on Information and Knowledge Engineering*, pages 412–416.

O'Reilly (2011). How to use and understand 'whois' in its many forms. http://answers.oreilly.com/topic/408-how-to-use-and-understand-whois-in-its-many-forms/. Acessado em 24/01/2014.

Parashar, M. and Hariri, S. (2005). Autonomic computing: An overview. In Banâtre, J. P., Fradet, P., Giavitto, J. L., and Michel, O., editors, *Unconventional Programming Paradigms*, volume 3566, pages 257–269. Springer.

Patel-Schneider, P. F., Hayes, P., and Horrocks, I. (2004). OWL Web Ontology Language Semantics and Abstract Syntax . http://www.w3.org/TR/2004/REC-owl-semantics-20040210/. Acessado em 25/02/2014.

Peace, A. (2011a). Formal Ontology and the Suggested Upper Merged Ontology. http://www.youtube.com/watch?v=EFQRvyyv7Fs&noredirect=1. Acessado em 05/01/2014.

Peace, A. (2011b). *Ontology: A Practical Guide*. Articulate Software Press, Angwin, CA, first edition.

Pease, A. (2009). Standard Upper Ontology Knowledge Interchange Format. Disponível em http://sigmakee.cvs.sourceforge.net/viewvc/sigmakee/sigma/suo-kif.pdf. Acessado em 13/01/2014.

Pease, A., Li, J., and Nomorosa, K. (2012). WordNet and SUMO for Sentiment Analysis. In *Proceedings of the 6th International Global Wordnet Conference (GWC2012)*, Matsue, Japan.

Pease, A., Niles, I., and Li, J. (2002). *The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications*. Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web.

Redmond, E. and Wilson, J. R. (2012). *Seven Databases in Seven Weeks: A Guide to Modern Databases and the NoSQL Movement*. Pragmatic Programmers, LLC, USA, 1 edition.

Tan, P., Steinbach, M., and Kumar, V. (2009). *Introdução ao Data Mining: Mineração de Dados*. Editora Ciência Moderna, Rio de Janeiro.

Trammel, A. and Happe, M. (2013). Autonomic network architecture - position paper. http://people.ee.ethz.ch/~arkeller/nmrg_ietf_88.pdf. Acessado em: 06/03/2014.

W3C (2014). Extensible markup language (xml). http://www.w3.org/XML/. Acessado em 24/02/2014.

Zikopoulos, P. C., Eaton, C., deRoos, D., Deutsch, T., and Lapis, G. (2012). *Understanding Big Data: Analyticas for Enterprise Class Hadoop and Streaming Data*. McGrawHill, New York, USA, 1 edition.