Reinaldo Ferreira

Universidade Portucalense, PhD Candidate; Take a City, CEO; reinaldo@takeacity.com

TITLE

Applying Semantic Web technologies to City Tourism information

ABSTRACT

Tourism is one of the fields where Web technologies have distinctive influence. The proliferation of Web Sites and Mobile Apps providing tourist information exposes the challenge to maintain tourism assets information updated and available for distribution. Semantic Web technologies provide an optimistic 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.

This paper discusses the challenges faced and results obtained in implementing Semantic Web technologies at TakeaCity.com, a global scale web portal developed in Porto and focused on Cities as network hubs of tourism activity. First, it starts with an overview of the technology's adoption potential. Then it addresses an Implementation Model that measures and facilitates that adoption. Finally, it covers the findings gathered so far and the future work to be developed both in the research and economic fields.

KEYWORDS:

Semantic Web, Web Services, Systems Integration

1. Introduction

In about twenty years, the Web left a laboratory where it was just an idea and became a ubiquitous and universal environment, accessible from basically everywhere, with information from basically everything. However, the "Web of Documents" is a less structured version of the Web. With the introduction of data access, structure and meaning, an environment more integrated, robust and practical becomes feasible. The next two decades will provide enough time to spread the "Web of Data", or at least that is 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 (Heath & Bizer, 2011). This very large database is still in its early days, though it's possible to find enough examples to evaluate its potential, since some research efforts and government policies for data publication produced satisfactory results (Bizer, Heath, Idehen, & Berners-Lee, 2008). However, Semantic Web technology usage in Enterprises is a less explored theme (Ahmed & Gerhard, 2010; Kuhn, 2010). Several explanations for a reduced adoption identify several difficulties, as homogenizing and validating data sources, defining knowledge rules and borders that allow relating data uniformly, analysing too complex examples, providing low cost technological capacity to allow its implementation, finding development tools, recruiting experienced professionals, finding success stories and adopting a paradigm shift in modelling, design and development (Ahmed & Gerhard, 2010; Kang et al., 2008; Kuhn, 2010; Pollock, 2008).

Taking into account these difficulties and the experience at Take a City, this article presents an Enterprise Semantic Web Implementation Model with two main goals. The first goal is to facilitate the introduction of the technology in organizations with different characteristics and motivations. The second goal is to evaluate the impact of the introduction of the technology in the applications used in those organizations and the tasks performed by their users.

Tourism is one of the fields where Web technologies have distinctive influence. The proliferation of Web Sites and Mobile Apps providing tourist information exposes the challenge to maintain tourism assets information updated and available for distribution.

Semantic Web technologies are a well fitted candidate to support and promote the exchange of that information over different platforms.

This article is structured as follows: section two reviews the state of the art of the Enterprise Semantic Web; section three presents the Enterprise Semantic Web Implementation Model; section four refers to the work conducted by Take a City; and, finally, section five concludes with a summary and future work overview.

2. The Semantic Web

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, to the smallest segment. It has also been so with the growth of social networks, involving people, enterprises, brands and products. And it will eventually be so in its third generation, with enterprises participating in the construction of the Semantic Web, the Web of Data.

Semantic Web technology may transform enterprise software, originating new business models and reducing costs in areas like data integration, master data management and enterprise information management (Pollock, 2008). Enterprise networks based on Linked Data principles originate a particular subset of Semantic Web technologies that may substantially reduce information integration costs, amongst other benefits, as in the integration of product, supplier, materials, legal, market and finance data or other internal and external data sources (Janowicz & Hitzler, 2010). However, the growing adoption of Semantic Web technologies and Linked Data principles bring the question on which applications may be developed to unleash their potential. The answer may come from identifying the areas where these technologies and principles may contribute distinctively from traditional technologies (Heath, 2010).

Meanwhile, the compromise between computational effort and flexibility tends to favour the latter, in line with Moore's law projection, with additional computing power leading to less concern with simplification or optimization. Simultaneously, the economy benefited with technological innovation and gave way to a highly competitive environment where speed and flexibility take 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 (Segaran, Evans, & Taylor, 2009).

The Semantic Web development has been guided by the view different communities perspective its evolution, considering their specific areas of research. One approach is that of annotation, taking as base large amounts of content available on the Web and using diverse techniques to originate structured data. Another approach is that of data repositories, starting from pre-defined structures and updating and interlinking them with additional structures. Finally, the approach that takes it as an agent platform, with applications combining different data sources and, ultimately, executing actions for individuals (Domingue, Fensel, & Hendler, 2011).

Additionally and taking into account the diverse nature of the organizations, it looks justifiable that the enterprise approach is different from the academic one. Enterprises will look for more practical results, focusing on the short term and higher success rates, while academic research would typically look for theoretical results, more ambitious and focused on the medium term. However, too much divergence may cause a fracture, where enterprises focus on too little and academia in aspects that will never be useful (Cardoso, Miller, Su, & Pollock, 2008), which advises for some equilibrium between both approaches.

The debate over the possible commercial success of the Semantic Web is exhausted and has been replaced by the discussion over what changes in commercial applications may occur with the introduction of the technology. Some authors suggest as possible evolutions the proliferation of highly distributed applications, agile development and dynamic integration of legacy applications, sensor networks and decision support systems (Domingue et al., 2011).

3. Enterprise Semantic Web Implementation Model

Acknowledging the diverse aspects to consider in the adoption of Semantic Web technologies in the Enterprise, an Enterprise Semantic Web Implementation Model would allow to help the introduction of the technology and the evaluation of its' impact in the applications used and tasks performed (Ferreira & Seruca, 2013). The model presented is based in several dimensions, where each dimension satisfies requirements

for technology implementation and, simultaneously, the opportunity to identify the desired continuous evolution through subsequent steps.

Figure 1 presents the model and identifies its four dimensions: Adoption, Provenance, Accessibility and Activities. The following sub-sections justify each of the dimensions.

Adoption

The most important technological adoption models with enterprise level application are the Innovation Diffusion Model and the Technology-Organization-Environment Model (Oliveira & Martins, 2011). According to the Innovation Diffusion Model (Rogers, 1995), 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 characteristics of the enterprise context influencing technological innovation, namely Technology, Organization and Environment. By combining both theories together, the Enterprise Semantic Web Implementation Model considers the following characteristics for the dimension:

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

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 that integration effort may increase substantially their return, reducing integration costs and increasing subsequent benefits (Janowicz & Hitzler, 2010).

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 (Koumenides, Salvadores, Alani, & Shadbolt, 2010).

Therefore, the question posed is to identify those data sources, given:

- applications topology, namely Web Sites, Extranets, Intranets, Web Applications or Web Services
- location, outside or inside the enterprise

Accessibility

The common user perception over data available over the Web is largely influenced by the format of the data. The path to the universal availability of data was clearly identified by Tim Berners-Lee with a five stars classification or five steps of evolution (Berners-Lee, 2006):

- available: just having data available in the Web
- formatted: through proprietary formats
- open: with open formats
- semantic: via semantic standards
- interconnected: with data hyperlinks.

Activities

The ambition to become a global database at the scale of the Web may seem sufficient to enthusiasm towards Semantic Web adoption. In reality, this technology seems to be a higher risk alternative to more conventional technologies, with few development tools, competences hard to recruit for, scarce success references and a true paradigm shift in modelling, design and development (Pollock, 2008).

Simultaneously, Information Systems have been classified according to different approaches, some with wider scope, some with more specific purposes (Lopes, Morais, & Carvalho, 2005). Despite the higher number of efforts, the question is not exhausted and, in this particular case, it is important to find a classification that positions the Web

as base environment, in alternative or complementary to more traditional classifications. With several studies related with Web usage, Tom Heath's work introduces important clarifications and a purpose oriented classification (Heath, 2010). According to this classification, user activities may be divided in:

- Locating: look, find
- Exploring: gather, research
- Grazing: navigate, browse, follow
- Monitoring: monitor, check, detect
- Sharing: distribute, collaborate
- Notifying: state, inform, communicate
- Asserting: opinion, suggestion
- Discussing: comment, respond
- Evaluating: assess, analyse
- Arranging: combine, negotiate
- Transacting: transfer, pay

Figure 2 shows the space for Semantic Web technologies implementation according to the degrees of diversity and conceptualization of the tasks performed.

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 less will 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 rarely will 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 turn tasks too complex or too diverse for traditional software (Domingue et al., 2011). Positioning the previous classified activities in this space, the opportunity for this technology can be visualized.

4. The case of Take a City

Take a City was founded beneath one key concept, the fundamental role of cities in modern economy and, particularly, as network hubs of tourism activity. The company developed a platform, completely Web based, with several applications covering individual and professional users' access, content management, customer relationship, transaction and invoicing, business intelligence and integration services. The platform uses Semantic Web technologies and adopts different ontologies for data structuring. The work done so far allowed testing the possibility to homogenize, validate and regulate data from different data sources and reduce complexity and cost in data integration scenarios. In order to do that, the platform was based in a largely distributed architecture and the modelling, design and development followed agile development methodologies and flexible design features to cope with change and uncertainty. The following sub-sections provide an overview of the main design features considered at Take a City's platform and an overview of the preliminary results obtained.

a. The City model

To build a global scale web portal, the company has developed a city centric model that considers each city a network node. The first challenge the company faced was to identify "all" the cities in the world, knowing that at some extent the weight of a smaller city may be obfuscated by bigger neighbour cities, while simulaneously some limit may have to be drawn to ensure that was a finite effort. The first indicator used to "find" the cities to be considered was the World Tourism Organization Top Cities Destination Ranking. This indicator is compiled from different data sources and shows the number of tourist arrivals in each city for 232 cities. Due to the lack of data from specific cities and the comparability of that data, statistical measurement of city tourism has many shortcomings (Mazanec & Wober, 2010) and may even be a factor of reality distortion (Dias, 2010). Nevertheless, it provides an important measure of each city's importance in the "network" of tourism. Given Take a City's city of origin, the same of this paper's author, and just for reference, Porto, Portugal, was placed in 135th place in this ranking.

Since this Ranking is limited and its' accurateness is not only disputable but certainly decreasing towards lower positions in the ranking, other indicators would have to be used. Another indicator chosen was the availability of internationally published city tourism guides. The assumption taken is that the free economy is a good indicator, in this case, translated by the confidence of publishers to invest in the distribution of a city

guide. The search conducted allowed identifying 498 cities, 333 of which are not on the previous list, raising the total number of cities to 565. For reference, Porto was also included in this list.

It is consensual that the modern society major regional organization is the country, a recognition that the United Nations itself represents as an organization. Even with countries not being considered as much important as cities in Take a City's model, there are many social, economic and political factors that highlight the role of each capital city in every country, which is another factor to consider and led to the inclusion of all capital cities. This inclusion increased the list in 128 additional cities, to a total of 693. Again for reference, in this case very obvious, Porto is not a capital city.

Considering air travel as the major contributor to medium and long distance travel, cities served by one or more close-by international airports take greater importance. Local transport in its many forms completes the interconnectivity of larger and smaller cities, forming a network of transport based accessibility that drives opportunities in tourism. With this additional indicator, 293 additional cities were added, taking the total to 986 cities.

b. The City ranking

The objective of creating such a list of cities to support the development of the city network was not only to identify the cities but to rank them relatively. It's clear that the list can also be further extended but it is not obvious to what extent and with what additional criteria. Therefore, a form of quantifying the importance of cities became necessary.

It was previously mentioned that the number of tourist arrivals was used to identify important cities and this may be one of the most important quantitative factors to consider. Additionally, the availability of international tourism guides, the identification of capital cities and the proximity of an international airport were used only as discrete variables.

Because of the lack of reliable data at city level, another factor that took some role in the model is the country's Gross Domestic Product (GDP), although different authors over different geographies reach different correlations to Tourism economic importance (Assadzadeh, Hojjat, & Nasab, 2012). However, given that GDP per capita is strongly

correlated to cost of living (Spagnoli, 2008; World Bank, 2010), the inclusion of this factor may reveal itself adequate.

At the present time, the City ranking has still a lot of room for improvement but serves well as an indicator of the relative importance of cities. As previously mentioned, when applying the model globally, 986 cities with relevant tourism potential were identified. This is not certainly a finite model, just like the first Web generation demonstrated with the "long tail" opportunity. Future developments of the model will lead to additional identifications and, for instance, will probably increment the relative importance of "shadow cities", cities located near very important cities but that are frequently obfuscated.

At the current level of maturity, there are some important conclusions that can be drawn from the model:

- 3% of the cities take 10% of the global importance
- 10% of the cities take 25% of the global importance
- 20% of the cities take 40% of the global importance
- 29% of the cities take 50% of the global importance

This is a pure "long tail of travel" scenario, as described by Chris Andersson in his reference work (Andersson, 2007) and further essays (Andersson, 2009), drawing opportunities for many smaller destinies to gain relative importance over a few established ones.

For perspective, Porto is currently placed at 147th place, in the top 15% of the cities identified in the model, not included in the top 25% of global importance.

c. Expanding the City model to the City level

With the growing importance of data in the economy, the indisputable role of Information Technology in innovation and the progressively adoption of Web Technologies by enterprises and individuals, Web Technologies in Tourism are thriving of research and economic interest. The application of these technologies in the Tourism space is an important differentiation for Take a City.

The expansion of the City model to the City level is an obvious step, as it approaches the economy and brings up several business model possibilities. The first approach Take

a City took to the market included the objective of characterizing businesses from a technological adoption perspective, guided by some optimism and the belief in the widespread knowledge of the web as a business driver. The samples chosen included businesses from one city, Porto, for convenience, and from one segment, restaurants, for similarity. The approach revealed several barriers to reach robust conclusions, namely:

- low availability of owners or managers to collaborate
- low awareness for technological matters and its relation with business
- low understanding of web technologies, starting on the difference and advantages of a simple web presences versus references in portals, aggregators or guides

The second approach focused on a much smaller sample of bigger businesses, from different market segments, taking a solution implementation perspective. Involvement with these enterprises transformed the desired generalizable and replicable conclusions into specific and personalised solutions that highlighted the following:

- there is a lack of role modelling in the industry, either from an academic or economic background
- there is a lack of signs of collaboration between academia, industry and government
- there is an excessive focus on short term results, including preference for low cost and high risk solutions over high return and medium term investments

Thus, it was difficult reaching conclusions that support the widespreading of the solutions proposed and the technologies associated to it. Even if sales based on individually specified services would still be possible, the deployment of a global scale solution would be an enormous or even impracticable effort and, consequently, robustness and competitiveness of the platform would be considerably lower.

Take a City is currently adopting the Enterprise Semantic Web Implementation Model presented to frame future understanding of the market, develop long term focused solutions and define short term actions that have impact in the industry. With that end in mind, the company is working with established businesses to study their current plans and, using Action Research methodology, conduct two cycles of diagnosis, planning,

implementation, evaluation and recommendation. The first results should be available during this calendar year and help validate the Implementation Model.

Conclusion

Semantic Web presents itself as an innovation opportunity for Enterprises. Current research identified implementation barriers, as organizational structure, and important questions, as what applications may benefit with the technology introduction.

This article presents an Enterprise Semantic Web Implementation Model based in four dimensions:

- Adoption
- Provenance
- Accessibility
- Activities

This Implementation Model should facilitate the analysis of the adoption degree and serve as a guide for Semantic Web technology implementation by pointing possible evolutions. In the near future, Take a City is actively involved in this effort and expects that the model can be validated in real scenarios and can provide insights into how can the Semantic Web generate opportunities for City Tourism information.

Bibliography

Ahmed, Z., & Gerhard, D. (2010). Role of Ontology in Semantic Web Development. *Knowledge Management*.

Andersson, C. (2007). *A Cauda Longa*. Actual Editora.

Andersson, C. (2009). The long tail of Travel. Retrieved March 12, 2012, from http://www.longtail.com/the_long_tail/2009/10/the-long-tail-of-travel.html

Assadzadeh, A., Hojjat, M., & Nasab, N. (2012). Investigating the Relationship between Tourism Industry and GDP in the Islamic Republic of Iran. *International Review of Business Research Papers*, *8*(2), 85–95.

Berners-Lee, T. (2006). Linked Data Desing Principles. Retrieved July 14, 2012, from http://www.w3.org/DesignIssues/LinkedData.html

Bizer, C., Heath, T., Idehen, K., & Berners-Lee, T. (2008). Linked data on the web (LDOW2008). *WWW2008 Workshop on Linked Data on the Web* (Vol. 2008, pp. 1265–1266). ACM New York, NY, USA. doi:10.1145/1367497.1367760

Cardoso, J., Miller, J., Su, J., & Pollock, J. (2008). Academic and Industrial Research : Do their Approaches Differ in Adding Semantics to Web Services ? *Lecture Notes in Computer Science*, *3387/2005*, 14–21. doi:10.1007/978-3-540-30581-1_2

Dias, F. (2010). Estatísticas de Turismo Urbano . O Centro Histórico do Porto e o Turismo. *Actas do Seminário Centros Históricos: Passado e Presente* (pp. 165–197).

Domingue, J., Fensel, D., & Hendler, J. A. (Eds.). (2011). *Handbook of Semantic Web Technologies*. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-540-92913-0

Ferreira, R., & Seruca, I. (2013). Modelo de Implementação da Web Semântica nas Empresas. *Actas da 8ª Conferência Ibérica de Sistemas e Tecnologias de Informação (CISTI'2013), Lisboa, Portugal, 19-22 Junho 2013*.

Heath, T. (2010). A taskonomy for the Semantic Web. *Semantic Web Journal*, *1*, 75–81. doi:10.3233/SW-2010-0003

Heath, T., & Bizer, C. (2011). Linked Data: Evolving the Web into a Global Data Space. In J. Hendler & F. Van Harmelen (Eds.), *Proceedings of the 17th international conference on World Wide Web WWW 08* (Vol. 1, pp. 1–136). Morgan & Claypool. doi:10.1145/1367497.1367760

Janowicz, K., & Hitzler, P. (2010). Creating knowledge out of interlinked data. *Semantic Web*, *1*, 97–104. doi:10.3233/SW-2010-0019

Kang, S. S., Yang, J. Y., Lee, S. K., Gong, K. H., Myung, J. S., Park, S. C., & Lee, S. G. (2008). An Enterprise Strategy for Semantic Technology Adoption. *The 5th International Conference on Information Technology and Applications*.

Koumenides, C. L., Salvadores, M., Alani, H., & Shadbolt, N. R. (2010). Global Integration of Public Sector Information. *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line, April 26-27th, 2010, Raleigh, NC: US*.

Kuhn, W. (2010). Modeling vs encoding for the Semantic Web. *Semantic Web Journal*, *1*, 11–15. doi:10.3233/SW-2010-0012

Lopes, F. C., Morais, M. P., & Carvalho, A. J. (2005). *Desenvolvimento de Sistemas de Informação* (FCA., p. 209).

Mazanec, J. A., & Wober, K. W. (2010). *Analysing International City Tourism*. Springer. doi:10.1007/978-3-211-09416-7

Oliveira, T., & Martins, M. (2011). Literature Review of Information Technology Adoption Models at Firm Level. *The Electronic Journal Information Systems Evaluation*, *14*(1), 110–121.

Pollock, J. (2008). A Semantic Web Business Case. *W3C*. Retrieved July 14, 2012, from http://www.w3.org/2001/sw/sweo/public/BusinessCase/BusinessCase.pdf

Rogers, E. M. (1995). *Diffusion of innovations*. (M. B. Salwen & D. W. Stacks, Eds.)*An integrated approach to communication theory and research* (Vol. 65, p. 519). Free Press. doi:10.1525/aa.1963.65.5.02a00230

Segaran, T., Evans, C., & Taylor, J. (2009). *Programming the Semantic Web*. (M. Treseler, Ed.) (p. 300). O'Reilly.

Spagnoli, F. (2008). Statistics on Gross Domestic Product (GDP) Correlations. Retrieved April 27, 2013, from http://filipspagnoli.wordpress.com/stats-on-human-rights/statistics-on-gross-domestic-product-correlations/#2

World Bank. (2010). PovcalNet: an online poverty analysis tool. Retrieved May 9, 2013, from http://iresearch.worldbank.org/PovcalNet/index.htm

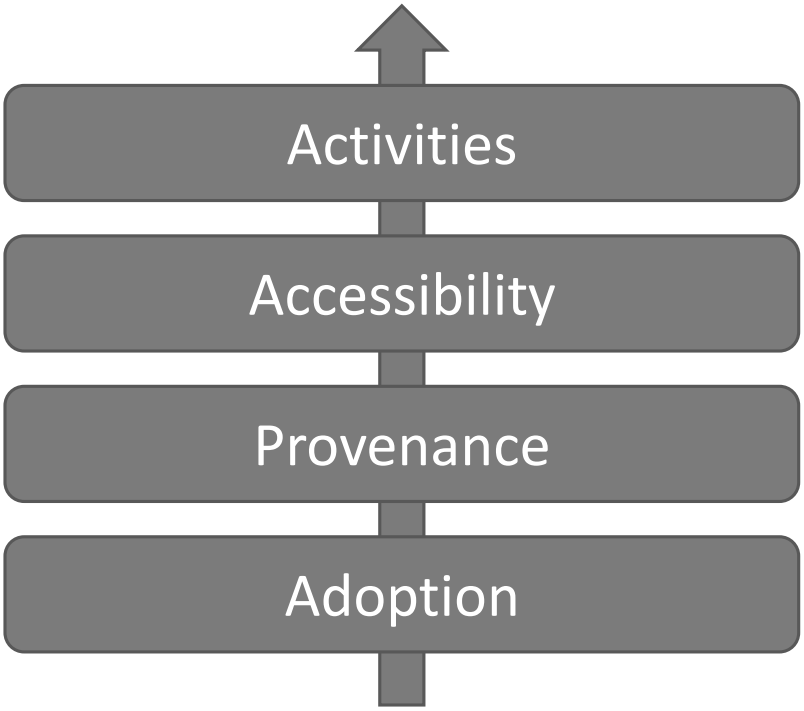

Figure 1 – Enterprise Semantic Web Implementation Model

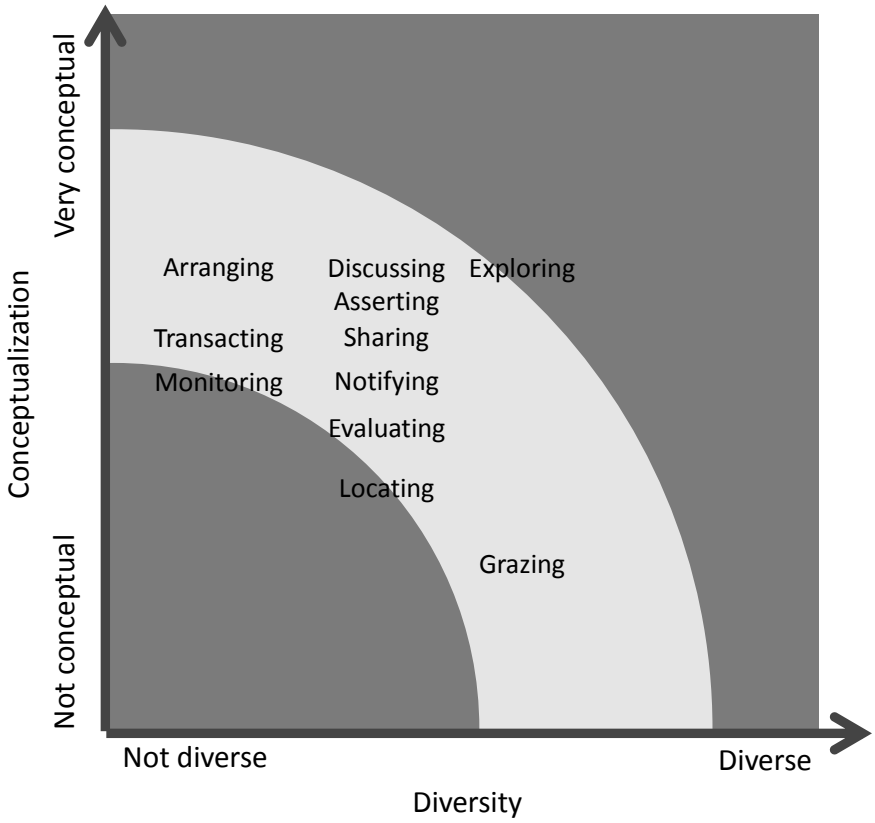

Figure 2 - Activities over Conceptualization and Diversity, adapted from (Domingue et al., 2011)