Semantic Business Intelligence - a New Generation of Business Intelligence

Dinu AIRINEI, Dora-Anca BERTA
Alexandru Ioan Cuza University of Iaşi
Faculty of Economics and Business Administration
adinu@uaic.ro, anca.berta@feaa.uaic.ro

Business Intelligence Solutions represents applications used by companies to manage process and analyze data to provide substantiated decision. In the context of Semantic Web development trend is to integrate semantic unstructured data, making business intelligence solutions to be redesigned in such a manner that can analyze, process and synthesize, in addition to traditional data and data integrated with semantic another form and structure. This invariably leads appearance of new BI solution, called Semantic Business Intelligence.
Keywords: *Business Intelligence, Semantic Business Intelligence, Semantic Web, Semantic Integration, Unstructured Data*

1 Introduction

The transition from social web (web 2.0) to the semantic web (web 3.0) opens new doors in terms of semantic data integration, one of the consequences being the appearance of a new type of BI solutions called **Semantic Business Intelligence** (SBI).

This research aims to respond to the question „*What is the difference between Semantic Business Intelligence and traditional solutions of Business Intelligence*?"; moreover, undertakes to demonstrate that SBI is not a fancy term or a marketing strategy but a new generation of Business Intelligence solutions.

The importance of this research approach is justified by the fact that all the Internet users begin to be involved, consciously or not, in phenomenon of semantic integration, a phenomenon that will be felt in all applications.

2 Research Methodology

This study starts from observing the impact it has the Semantic Web to data and application that produce, manage, process and analyze this data, such as Business Intelligence. Its goal is to define and analyze the emergence of a new generation of business intelligence in the context of web transition from version 2.0 to version 3.0.

The hypothesis we propose to research is: "*Semantic Business Intelligence is a new generation of Business Intelligence*".

Hypothesis' construction was based on use of the difference method. This method explains a phenomenon depending on differences that occur when certain factors are varied [30] in the case differences that arise between SBI and BI due to the impact of the Semantic Web. Also, in construction of hypothesis was used inductive inferring. Thus, the research will have an inductive approach in that it will be analyzed on two applications on the assumption that will be generalized.

Hypothesis testing will be done using qualitative analysis, comparative functionalist type [31]. Whereas deterministic analyzes the influence of a situational context, in this case the introduction of new technologies in certain applications. The appearance of a new generation of Business Intelligence generically called Semantic Business Intelligence based on the Semantic Web specific technologies will confirm the proposed hypothesis.

3 Related works

Even if it occurred in 1958 [5], Business Intelligence begin to become known since 1989, when Howard Dresner defines it: "an umbrella concept that describes a set of concepts and methods to improve business decisions by assist systems use decisions based on facts " [24].

In our vision, Business Intelligence is a set of economic applications used for analyze data from companies in order to transform them into information that will substantiate the decisions taken by managers.

Along the time, attempting to keep up with technology, Business Intelligence passed through several phases, becoming from operational BI real-time BI [26], adding socialization modules [6], and expanding its functionality to be used by mobile phones, turning to Mobile BI [1] and now by adding new semantic technologies becomes Semantic Business Intelligence.

Given that the subject addressed is a relatively new, literature is rather poor in this area of Semantic Business Intelligence. Internationally, the field began to attract the interest of researchers since 2006 due to research projects to find viable solutions for the integration of semantic technologies in Business Intelligence solutions.

MUSING, a European research and development project funded by FP6 and passed off in 2006-2010, had as mission developing a new generation of Business Intelligence. The goal of this project was to support research that aimed to develop tools and modules for BI which integrate elements of semantic web and human language processing technologies to improve knowledge acquisition of BI solutions [16].

This project has managed work up an interest to some researchers, who treated Semantic BI, but their number was quite limited compared to the area to be exploited, and their interest after the project decreased significantly.

Fortunately, this mega-project finds its continuity in projects funded by FP7 Cubist (Combining and Uniting Business Intelligence with Semantic Technologies).

CUBIST is a project funded by the European Commission on the priority axis "Intelligent In-Formation Management" and will take place between 2010 and 2013. The major pillars are represented by semantic technologies, business intelligence and Visual Analytics, and as areas of interest are covered: market and competitive intelligence, computational biology / biomedical informatics and control centers operations [36].

What makes this project truly extraordinary is combination between theory and practice: researchers receive support and cooperation from the developers of applications or technologies. The most important partners in research work are: SAP as a conceded producer of BI and Ontotext Lab, a developer of semantic technologies all too well known. On the other hand, researchers are selected from prestigious universities such as Sheffield Hallam University (England), CENTRAL Recherché in Paris and Heriot-Watt University (England).

Expected results of project implementation aim to incorporate unstructured data into existing solutions. For this to be possible, researchers propose that the level of technology to design new architecture that supports integration of unstructured data, hence the need for alloying with specialist manufacturers of applications.

By consulting the literature, we found that in Romania there is no post to be treated this subject. One reason could be related to the fact that on the market has not been promoted semantic solutions yet; most manufacturers are currently working on them. Looking at things from another angle, the Romanian companies do not seem interested in semantic integration of data which they hold.

4 Semantic Web– premise of SBI appearance

The Semantic Web "was born" as a desideratum [29]: to have data in a Web defined and interconnected (linked) in a manner that computers can be used not only to display but especially for automation, integration and reuse of data from various applications. Berners-Lee's vision about the Semantic Web takes the form of a highly interconnected network of data that could be easily accessed, understood and used by any desktop or portable computer [4].

According to the dictionary [32], the term "semantic" considered by the light of informatics is seen as "The theory of a formalized particular system interpretation by other formalized system." On the literature [27], [19], [10] were identified several types of semantics applied in informatics [27]: real-world semantics, axiomatic semantic and model-theoretic semantics, But any of them would

be used the purpose will be the same: to provide a well defined sense to that word / phrase to which it applies.

The appearance of the Semantic Web, also called Web data [9] promises significant improvements to the current web. Large amount of information online, returning irrelevant results sometimes without accuracy in web searches and the lack of applications to process natural language in idea of being understood by computers are just some of the shortcomings that experts hope to solve the Semantic Web [34].

- *Excessive amount of information* - people create an unimaginably large amount of digital information, which grows exponentially in relation with time, reaching to the size of zettabyte [35] (a zettabyte represent 10^{21})
- *Inefficient search keywords in search engines* – currently, any search information on the web will not only return the requested information on a domain, but also information with the same shape but different meaning (intended only as a word or phrase, not meaning, namely semantic)
- *Distrust in the veracity* – there is no control mechanism to manage web content so as may be posted only real information
- *Lack of information systems for natural language processing* – currently there is no common language to be perceived in the same way as man and computer.

Semantic Web is based on XML that develops new standards coding and description of data from and about the Web. Semantic Web technologies were developed as a prerequisite for process automation and improvement of service enabling data integration and interoperability. The most known and used are used technologies semantic RDF Schema, OWL, SKOS, SPARQL, GRDDL, etc.

In terms of the link between Web and BI, we can say that Web 1.0 has no influence on him. Things change, however, with the emergence of web 2.0 and social networks when implementing CRM applications BI modules that include social messages and posts on social networks to analyze their influence on companies [6]. We are currently witnessing the appearance of a new generation of BI inspired by the Semantic Web. In Figure 2 we tried to plot the relationship between the two concepts.

Fig. 1. Web impact's on Business Intelligence

5 SBI – a theoretical approach

The literature by the economic data and found the web organizations are divided into three categories. [28]:

- structured data: data companies from SGBD: ERP, CRM, SCM, BI;
- semistructured data: data from RSS and XML documents;
- unstructured data: email-uri, blogs, social networks, mobiles, etc.

Semantic Business Intelligence has emerged in response to the question „*How can be analyzed unstructured data*?" given the fact that BI solutions currently used applies only to data that are significantly less structured than those unstructured - according to Paquet unstructured data represents about 80% of all existing data worldwide [18].

Researchers from CUBIST project believe that to be able to integrate unstructured data

in its analysis, business intelligence solutions should be created based on new architectures. They suggest a possible SBI structure, shown in figure below.

Hereinafter, we will try to highlight with reference to this figure we compared differences appear between BI classic and Semantic BI.

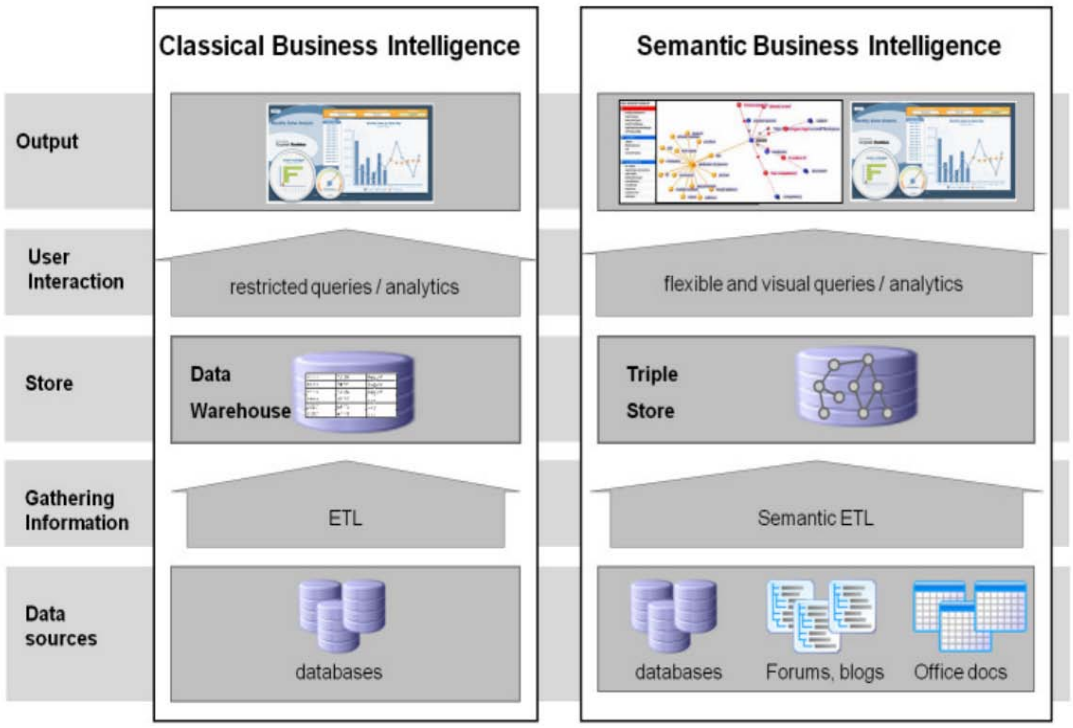

Fig. 2. BI versus SBI [37]

a) **Data sources:** apart from data from company departments, SBI intends to analyzing the data found on the web that relate to the organization, too. For example may be relevant for a company to know its
customers Reviews relating to products tested, or are their preferences for a product. Also, analysis of unstructured data gains importance, given that it can reveal hidden information at first glance, being useful to BI applications in providing information about the competition.

b) **The storage:** in traditional BI, data warehouses' role is to organize, integrate and store data from all departments, of an organization. Inside SBI their place is taken by triplestores.
The use of triplets is a way that makes statements about web resources by RFD consisting of subject, predicate and objects, where [7]:

- The subject is the resource (identified by a URI)
- Object is the resource's values
- Predicate – specify the relationship nature between subject and object

For example in Fig. 2 is rendered a triplet.

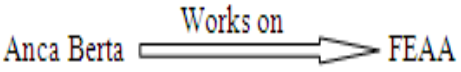

Fig. 3. Example of a triplet

c) **Data analysis:** on traditional BI solutions the data are analyzed through data mining tools. Into SBI, data mining tools will be replaced by FCA (Formal Concept Analysis). FCA is a method used for data analysis, knowledge representation and information management [25] where the data is structured in units in the guise of formal abstractions of concepts of human thought, allowing a comprehensive and easily understandable interpretation [8].

As in the literature will appear design specifications, models of architecture and structural schemes of SBI certainly that this list will grow, but for now it is enough to validate the proposed hypothesis. As you can see SBI comes with two new technologies, and Formal Concept Analysis triplestores, unprecedented in traditional BI.

6 SBI market

Business Intelligence Semantic Model, produced by Microsoft, is the first BI solution on the market that tries to integrate semantic data. According to market's data [22] published by the prestigious company Gartner, in 2011 Microsoft ranked the five most popular buyers of BI, with a percentage of 9% which means 1.05 million dollars from total sales of BI, in these circumstances the appearance on the market with a semantic solution is inspired and strategically chosen.

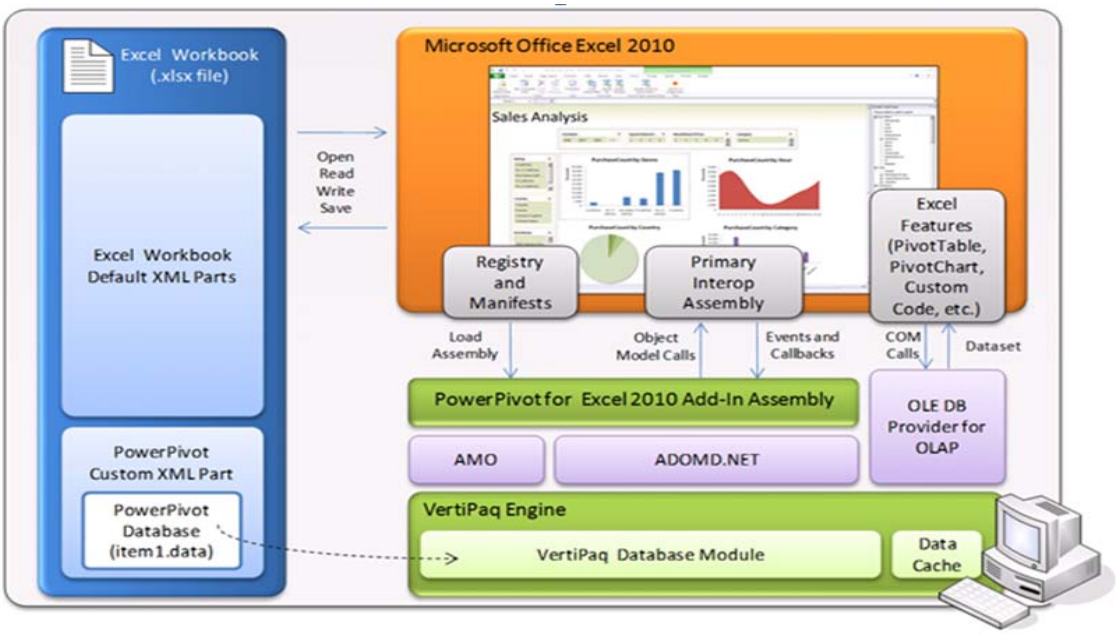

Fig. 4. Semantic Business Intelligence Model [11]

In this solution, there is a **layer of metadata** that describes the concepts (entities) and connections (relationships) between them. This layer was thought by developers of BI solution to be *user-oriented* [23], in that it shows what can they gain significance, for example it shows and what are the tables and relationships in the real world.

One of semantic solutions Microsoft is PowerPivot, which is nothing but a collection of applications and integrated into Excel 2010 and SharePoint 2010. The novelty of PowerPivot occurs due to the fact that instead of cubes MOLAP, ROLAP and HOLAP classic components of traditional BI solutions, PowerPivot has his own way of data storage called VertiPaq, a database vertically organized.

Thanks to a special layer called "layer on relationship" can be integrated, process and established relationships between data by mapping data columns containing identical or similar, even if them come from different sources as if all would come from the same place, and included both data from other Excel worksheets and other types of advanced databases.

SAP BusinessObjects BI 4.0 - SAP directs one's efforts towards to semantic integration in SAP BusinessObjects SAP BI 4.0 application by adding a new semantic layer in OLAP. This semantic layer is an abstraction layer located between the database and users and allows them to access, process and analyze data, regardless of origin and regardless

of the complexity of data structures and technical details.
Semantic layer comprises the following elements [15]: parameters for connection to the database, universes, panel query, query generator, computer, and the local cache (represented by a microcube). For all these parts listed, showing the new elements is "universe".
A universe is an organized collection of metadata objects including dimensions, measures, hierarchies, attributes, predefined calculations, functions and queries. Object metadata layer called business layer is built on the relational database schemas or OLAP cubes so that the objects to be mapped directly onto the SQL database structure or MDX expressions, includes identifying those connections to data sources achieving queries.
The role of he universe is to simplify the user's work because tries to give its objects semantically understandable (e.g. Client, Quarter, Sales, etc) without having to know the technical terminology, physical or logical structure application or data source.
As it can see, each presented applications intends to use a layer data to integrate the unstructured data. As will appear on market BI solutions of other major manufacturers certainly will occur new semantic technologies, in addition to those presented in the material.

7 Conclusion

SBI field has not yet been defined with sufficient precision being enough fragmented and volatile. Meaning given in the literature is also quite broad and imprecise. However, as we have seen, from the presentation of two new solutions, Business Intelligence Semantic Model and Business Objects SAP BI 4.0, it contains new components and technologies, different components of applications encountered before in market.
Using new technologies like FCA, vertical databases, data universes and triple stores and lead us to the conclusion that we are talking about another level of BI solutions. Thus, the use of semantic technologies in BI validates the hypothesis that we start up according to which, Semantic Business Intelligence is a new generation of business intelligence solutions.
Even if SBI's developing solutions only on the beginning, from the materials analyzed can be seen easily that speak of a new level of applications. Using new technologies that FCA and triple stores, leads to conclusion that we talk about a new generation of BI, generic called Semantic BI, thus validated the hypothesis that we started.
The appearance of applications to process semantic data is salutary but it remains an open question referred to specialists "What will happen with BI solutions currently used in companies?" Considering the fact that acquisition involved traditional BI outstanding financial efforts from companies is hard to believe that will discontinue them to acquire other advanced technologies.

Acknowledgements

This work was supported by the European Social Fund in Romania, under the responsibility of the Managing Authority for the Sectorial Operational Program for Human Resources Development 2007-2013 [grant POSDRU/CPP 107/DMI 1.5/S/78342].

References

[1] D. Airinei and D. Homocianu, „The Mobile Business Intelligence Challenge", Informatica Economica [Online]. 10(1/2010). Available at: http://www.economyinformatics.ase.ro/content/EN10/01%20Airinei%20Homocianu.pdf , pp. 5.
[2] T. Berners-Lee, (1999, August, 17), "Weaving the Web", Orion, Cambridge. Available at: http://www.w3.org/ People/Berners-Lee/Weaving/
[3] T. Berners-Lee, (1998, September, 14), "Semantic Web Road map". Available: http://www.w3.org/DesignIssues/Semantic.html
[4] T. Berners-Lee, J. Hendler and O. Lassila, (2001), "The Semantic Web - A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities", *Scientific American Journal* [Online]. Available:

http://www.scientificamerican.com/article.cfm?id=the-semantic-web

[5] D.-A. Berta, „Business Intelligence in education", Leveraging Technology for Learning, *in Proceedings of The 8th International Conference eLearning and Software for Education*, vol. II, Bucharest, 26-27 April 2012, p.62.

[6] D.-A. Berta, "From Marketing Viral to Social CRM", Analele Universității Ovidius, Seria: Științe economice, Nr. Vol. XI, nr. 2/2011, Editura Ovidius Press, 2011, pp. 89-90.

[7] S. C. Buraga, „Web03 Semantic Web: realizarea inferentelor via RDF(S)", Available at: http://www.slideshare.net/busaco/web03-resource-description-framework-presentation

[8] Ph. Cimiano, (2005, August, 04) „Formal Concept Analysis", Available: http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume24/cimiano05a-html/node3.html

[9] M. Daconta, L. Obrst and K. Smith,"*The Semantic Web: The Future of XML, Web Services, and Knowledge Management*". John Wiley Inc., June 2003, p.1.

[10] R. Fikes and D. McGuinness, (2001, December, 2001). „ An Axiomatic Semantics for RDF, RDF-S, and DAML+OIL". Available at: http://www.w3.org/TR/daml+oil-axioms

[11] K, Jonge, (2010, August, 1)," Share your PowerPivot file with Excel 2007 users". Available at: http://www.powerpivotblog.nl/share-your-powerpivot-file-with-excel-2007-users

[12] D. Hudson. (2011, May, 16). "The Semantic Social Business Phase of Enterprise 2.0". Available at: http://webtechman.com/blog/2011/05/16/enterprise-3-0-semantic-social-business/

[13] T. Lachev. (2010, November, 13). "Business Intelligence Semantic Model – The Good, The Bad, and the Ugly". Available at: http://prologika.com/CS/blogs/blog/archive/2010/11/13/business-intelligence-semantic-model-the-good-the-bad-and-the-ugly.aspx

[14] S. Mattison, "Microsoft Business Intelligence", *Conference at TECH-ED North America, New Orleans*, Louisiana., 2010, Available at: http://blogs.adatis.co.uk/blogs/simon/default.aspx

[15] D. Mazoue, E. Immand and A. Astito. (2010, Octomber, 18). "Questions and Answers: SAP BusinessObjects BI 4.0 and the New Semantic Layer for OLAP". Available at: http://www.sdn.sap.com/irj/scn/go/portal/prtroot/docs/library/uuid/b08143a1-eea3-2e10-da94-f6c5b122e0f4?QuickLink=index&overridelayout=true&51887500377508 pp.4-5.

[16] MUSING, „Multi-industry, Semantic – based next generation business Intelli-Gence" – Short Presentation, Available: ftp://ftp.cordis.europa.eu/pub/ist/docs/kct/musing-presentation_en.pdf

[17] M. Napoli, F. B. Ghisi, J. L. Pacheco and R., Todesco, "SBI: A Semantic Framework to Support Business Intelligence", *OBI'08*, Karlsruhe, Germania, Octombrie 2008.

[18] R. Paquet. (2010). "Technology Trends You Can't Afford to Ignore". Available: http://www.gartner.com/it/content/1503500/1503515/january_19_tech_trends_you_cant_afford_to_ignore_rpaquet.pdf

[19] Ouksel and A. Sheth, A Brief Introduction to the Research Area and the Special Section, *Special Section on Semantic Interoperability in Global Information System, SIGMOD*, Vol 28(1), 1999.

[20] R. Paquet. (2010). „Technology Trends You Can't Afford to Ignore. Available at:http://www.gartner.com/it/content/1503500/1503515/january_19_tech_trends_you_cant_afford_to_ignore_rpaquet.pdf Accessed 10.03.2012

[21] U. Passant, J. Bojars, Breslin and S. Decker, "Semantic Enterprise 2.0 - Enabling Semantic Web Technologies in Enterprise 2.0 Environment, *Semantic Technologies Conference, San Jose, SUA*. Available at: http://semantic-conference.com/session/1783/

[22] R. Pettey, van der Meulen, "Gartner Says Worldwide Business Intelligence, Analytics and Performance Management

Software Market Surpassed the $12 Billion Mark in 2011". Available at: http://www.gartner.com/it/page.jsp?id=1971516

[23] T. Piasevoli, "Business Intelligence Semantic Model", *Microsoft Local WinDays 11 Conference*, Rovinj, Croatia, 2011. Available: http://www.mswindays.com/ , p.13;

[24] D.J.,Power, 2007, „A Brief History of Decision Support Systems". Available: DSSResources.COM, World Wide Web, http://DSSResources.COM/history/dsshistory.html

[25] U. Priss, "Formal Concept Analysis in Information Science", [Online]. Available:http://www.knu.edu.tw/lecture/%E6%AD%B7%E5%B9%B4%E6%95%99%E5%AD%B8%E8%B3%87%E6%96%99/2010%E5%B9%B4%E7%A0%94%E7%A9%B6%E6%96%B9%E6%B3%95%E5%A4%8F%E4%BB%A4%E7%87%9F/2010-8-17/Formal%20concept%20analysis/conceptAnalysis-Soft/Formal%20Concept%20Analysis%20in%20Information%20Science.pdf, p.1;

[26] D. Sandu. (2008). „Operational and real-time Business Intelligence", Revista Informatica Economica Journal [Online]. 3(47). Available at: http://revistaie.ase.ro/content/47/06Sandu.pdf, pp. 2-4.

[27] M. Uschold, (2001, June), "Where are the Semantics in the Semantic Web?" *Autonomous Agents Conference*, Montreal, 2001, [Online]. Available: http://lsdis.cs.uga.edu/SemWebCourse_files/WhereAreSemantics-AI-Mag-FinalSubmittedVersion2.pdf, pp. 3-4 ;

[28] R. Wang. (2010, June, 28). Research Report: Rethink Your Next Generation Business Intelligence Strategy. Available:http://blog.softwareinsider.org/2010/06/28/tuesdays-tip-rethink-your-next-gen-business-intelligence-strategy/

[29] W3C. (2001). „Semantic Web Activity Statement". Available at: http://www.w3.org/2001/sw/Activity;

[30] D. Zaiț and A. Spalanzani, "*Cercetarea în economie şi management – repere epistemologice şi metodologice*", Bucharest: Editura Economică, 2007, pp. 145-146;

[31] D. Zaiț and A. Spalanzani, "*Cercetarea în economie şi management – repere epistemologice şi metodologice*", Bucharest: Editura Economică, 2007, pp.177-178;

[32] http://dexonline.ro/definitie/semantic

[33] http://rhetoricaldevice.com/articles/TripleStore.html

[34] http://www.studyblog.net/2008/10/business-intelligence-group-assignment-semester-2-2008/

[35] (2010, February, 25) „Data, data everywhere", *The Economist Journal*, Available at: http://www.economist.com/specialreports/PrinterFriendly.cfm?story_id=1557443

[36] http://www.cubist-project.eu/uploads/media/CUBIST_factsheet_short.pdf

Dinu AIRINEI has graduated the Faculty of Economic Sciences from "Alexandru Ioan Cuza" University of Iaşi (UAIC) in 1978. He holds a Ph.D. diploma in Economics from 1997 and he had gone through all didactic positions since 1982 when he joined the staff of the Faculty of Economic Sciences, teaching assistant in 1982, senior lecturer in 1990, assistant professor in 1998 and full professor in 2000. Currently he is full Professor of Business Information Systems within the Department of Economics, Quantitative Analysis and Information Systems, at Faculty of Economics and Business Administration (FEAA) from UAIC. He is the author of more than 22 books and over 65 journal articles in the field of Business Information Systems, Decision Support Systems, Data Warehouses and Business Intelligence.

Dora-Anca BERTA has graduated the Faculty of Economics and Business Administration from "Alexandru Ioan Cuza" University of Iaşi (UAIC) in 2008 and Faculty of Automatics and Computer Science from "Gh. Asachi" University of Iaşi in 2011. Since 2010 she is a Ph. D. student on "Alexandru Ioan Cuza" University of Iaşi on Cybernetics and Economics Statistics and associate teaching assistant on Faculty of Economics and Business Administration Iaşi. She is focused on Decision Support Systems and Business Intelligence.