How Web 2.0 improves Business Intelligence: showcase of emerging technologies

Berta Buttarazzi, Mirko Mechilli, Luciano Polinari
University of Roma "Tor Vergata", Via del Politecnico, 00133 Rome, Italy

Abstract. This paper explores the principles and technological processes useful for Business Intelligence 2.0 starting from the hypothesis that the integration of Web 2.0 in conjunction with new approaches and technologies may lead to relevant benefits. The goal is to help developer to identify methods and technologies (Wikis, Widget, Mashup) for improving architecture where more than ever, BI will have an impact on businesses seeking to grow and gain new sources of revenue while cutting costs and saving time.

Keywords. Business intelligence, web 2.0, Mashup, Widget, Wikis

Introduction

Business intelligence is a recent technique and discipline. It is not yet a science though it will probably soon become one, but it must already face a complex process of structural and tool related revamping, which will transform the features that were so typical of it up to now.

According to the research [1] carried out by T.P. Gerrity, and to the paradigms of Kimball [2] and Inmon [3], up to the pragmatic approaches of the entire community of operators, BI owes its success to two fundamental factors:

- the multiplication of information available to the companies due to internal factors: extension of the automation of management processes, reorganisation of corporate structures (fusions, acquisitions, mergers, etc…), the revision of internal IT systems, etc…; but also to external factors : globalisation of information, revolution in the access and distribution of information (just think about the spreading of the Net), reduction of purchase costs, etc…;
- the incredible technological evolution of hardware and software systems over the last few years. The hardware structure is now available to memorise enormous volumes of data, compared to the past, and represents a valid

support to the process of data acquisition, whilst the exponential growth of the average processing capacity of a PC makes it possible to process and access a vast quantity of information in a short amount of time.

BI was created with a specific requirement in mind, according to principles defined through a practical experience adding value every day.

Due to this change, one remains linked to the evolutions of the conditions of the socio-economical, communication and human systems that commonly interact with our way of operating.

The aim of this document is to describe the characteristics and main components of a BI system, and to identify those that may become the necessary tools to integrate the significant amount of non structured data, sent from sources that are external to the IT systems of a corporate entity.

The focus is on the network and on the evolution towards Web 2.0, which is the prominent channel for collection and diffusion of non structured information, with the aim to identify logical and operational processes that are essential to the acquisition, standardisation and organisation of simple and complex data and information sent from the latter.

The first chapter describes the main features of Business Intelligence systems and the objectives they pursue. The second chapter shows a real case for the applicability of cognitive systems, highlighting their peculiar characteristics. The third chapter explores the world of Web 2.0, highlighting the distinctive elements of a truly innovative way of operation. The fourth and final chapter discuss the possibility of using Web 2.0 tools to publish information dynamically within decision support dashboards.

1. Business Intelligence systems as Decision support tools

Quoting Gerrity, we can define a Decision support system (DSS) as "… an efficient merger of human intelligence and technologies which strongly interact amongst each other to solve complex issues." [1]. There is no better definition to allow us to grasp the theme being discussed, the scenario in which a DSS takes shape, the motivations and the targets at which it aims.

Decision support tools are designed and implemented using the analytical capacity of the human brain and technological resources that are presently at hand when tackling

this activity. But let us not forget that, to commit to such a challenge, we clearly need an objective, a defined mission, and a finish line to reach. Very often, the scope is defined inside an economical scenario, rather than a social or scientific scenario.

In the current economic and social context, the true challenge is to make decisions quickly through timely and meaningful information to maintain a competitive edge.

Having access to a tool that makes it possible to increase analysis efficiency, improve the capacity to assess pending risk situations, develop management strategies and organisational structures, means that the "decision maker" can face the hard task of managing a company.

At this stage, the concept of Business Intelligence (BI) is clear: a process of analysis and understanding of logical rules which characterises the production process of a company.

Exploration, in depth examination of the behavioural path of a company, the dynamic evolutions of the production cycle, verifying that the objectives defined by the management, but also the process of transition towards those, must follow an appropriate trend.

All of this represents the BI, envisaged as a true and real corporate process, which focusses on data and information collection to analyse them and to derive assessments and estimates on the corporate context and surrounding market.

The sources of information are mostly internal, coming from the corporate information systems, integrated amongst each other and often completed by information coming from the outside.

The input comprises, on the one hand, of the data inside the information applications of the company, and of the output of the management processes embedded in the software products. The role of the BI system is to transform raw data into information, and the information into knowledge.

On the other hand, it comprises of applications that are not present in the company, mainly a network that is able to provide a large amount of information.

Here is a simple equation[1]

INFORMATION = DATA + MEANING

[1] The word data can be defined as a simple illustration of the occurrence of a phenomenon being observed, whilst information is the notice of a knowledge that related to the occurrence of phenomenon being observed.

It shows us the relationship between a data and an information. Whilst a data is only the expression of the measured value of the event (done) being considered, the information provides this value with a clear and recognisable meaning, it is the knowledge expression of the occurrence of the phenomenon measured by this value.

We are talking about the meaning that must be attributed to the data, in order to convert it into information. The meaning is the logical position of a measurement inside a predefined observation model. The following step aims to organise all the measured information, extracting the knowledge that is necessary to represent the context in which we work[2].

The classification and organisation of information takes place by defining the reference framework, the scenario, the relevant model required to identify logical rules and metadata of the corporate function (business process) under analysis. The BI aims to transform simple data in knowledge through a representation of the relationships that link the variables involved, inserted in an operational scheme that describes the functional corporate environment.

The objective is therefore to classify values derived from measurement of the phenomenon involved, and to organise them carefully (through a specific method) according to a series of rules (business rules). These rules are specific to the phenomenon itself and the organisation takes place in an information structure (database) that is set up appropriately. The data structure takes its name from the data warehouse and covers all the source data that went through the foreseen transformation:

- formal control
- harmonisation and reconciliating
- normalisation and cleaning

whilst the data migration and integration process is called ETL (extract, transform e load).

The data warehouse is the basic structure for information analysis processes[3], which contribute to decision support, in order to setup an interactive and iterative procedure

[2] More and more often, the concept of business intelligence appears to be associated to that of knowledge discovery in databases (KDD).

[3] Now we are talking about information, not data anymore, since the ETL process, through transformation operations, implements the highlighted and summarised logical rules that are represented in the analysis structure, the data warehouse.

aimed at extracting hidden knowledge and unknown behaviour patterns that were unknown before and potentially useful [4].

The final user must always be involved in the definition of a decision support system, or analytical system, since he is not only the user of the product but also the main actor of it.

The application scenarios of the Business Intelligence process can be multiplied, surely in the scientific and medical fields but also in the economical and social fields. We provide a case study below, in which we were directly involved, to demonstrate the complete practical application of the aforementioned concepts.

2. GIACO LS architecture

The environment that we are about to describe is typically economical and representative of a company that operates in a specific operational field, the field of insurances.

The claim payment process is strategic for insurance companies since it represents accurately the outflows of money and the cover policy (reserve) set aside by the company to cope with expenses.

The ever closer link between the company and the client means that the claim payment process takes on a different meaning: from a "meeting between strangers" to "a service for the client", where the service to the client is under critical scrutiny of the consumers. Therefore the perceived service quality and efficiency will have a growing impact on the insurance business and companies will be able to manage claim cost factors, claim payment times and claim fairness. The ideal claim payment system must be able to ensure:

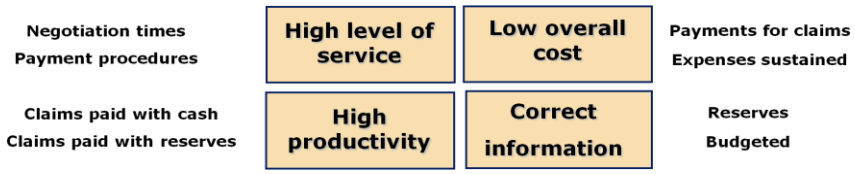

Figure 1. Claim payment system

and make it possible to make decisions in connection with the definition of service quality parameters, with organisational efficiency, with profit management and with process phase control.

To define an efficient operational procedure, a decision support system is designed and implemented to monitor the entire process. At this stage of work, the knowledge manager's contribution plays a key role in defining the steps needed to portray the entire functional process under observation. To enrich the scenario which represents a business function is essential the user intervention, who is the entity that may interact with the BI system, an intelligent user who will be discussed in detail in the next section.

The objective is to measure the performance on claim payment management with a product that is able to provide specific indications on the activity of the insurance company in connection with:

- Indicators of workload affecting the structure (inspectors, claim payment centre, area, etc…);
- Payment progress indicators which, through appropriate comparisons, give information on the level of efficiency reached by the organisational structure;
- Payment cost indicators which aim at monitoring the evolution in order to define appropriate correction measures;
- Payment speed indicators to control the internal process efficiency and the way this is perceived by the insured / victim;
- Indicators related to the process stakeholders out the company such as experts, legal advisors, etc…, to confirm the efficiency and trustworthiness.

A management system of knowledge based analyses is an ideal tool to monitor the management of claim payments as a measurement of correctness, efficiency and payment speed and makes it possible to analyse information organisation in the right way to support strategic decision to be made by the management of insurance companies (fig.1).

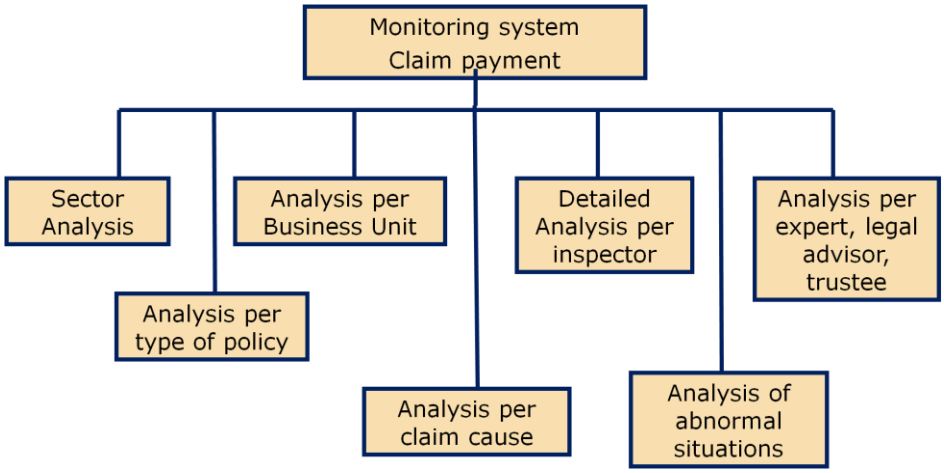

Figure 2. Organisation of information

The product, called GIACO LS, has the typical characteristics of a Business Intelligence product, such as:

- a Datawarehouse (a denormalised structure), built on RDBMS, able to contain the information foreseen;
- an ETL module to define the data filling process on the DWH;
- a series of structures and procedures, for the organisation and sorting of data, required for a correct release of information;
- a highly innovative tool to use the information through a web dashboard that is able to build informative frames summarising the information;
- all of these, for the aforementioned characteristics, making it possible to build a system:
- scalable, for compatibility with future demands (complete the data warehouse with information coming from other corporate functions);
- compatible with existing applications, in complex environments and architecture and different technologies;
- flexible, to be able to satisfy specific customisation demands in various functional environments and areas.

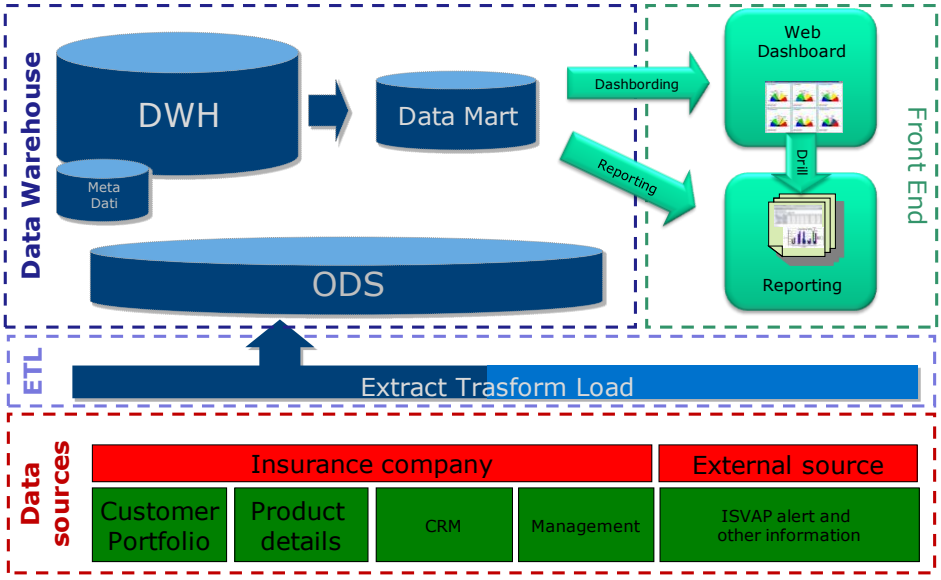

Figure 3. Applicative architecture of system GIACO LS.

The information managed from GIACO LS describes the entire life of the claim payment process and comes from:

- the client and product portfolio
- the corporate CRM system
- the administrative and management system for production of registers and balance sheet attachments
- the third party liability insurance claim alert system to ISVAP.

But it would be preferable to complete it with other information coming from external sources (behaviour of clients, territories, statistics, etc…) adequately prepared, integrated and standardised.

The process concludes with the publication of information through the production of a report and a dashboard, built on a web platform, that is usable by users through access to the Internet network or the corporate Intranet.

Figure 4. Decision support dashboard.

Hereafter, we provide some of the information frames that may be built from the system based on defined indicators.

The total number of claims attributed to each inspector with proof of claims followed, open and closed, and the navigation per branch.

Figure 5.Workload for each inspector and drill down per branch.

Evolution in time of claim communications.

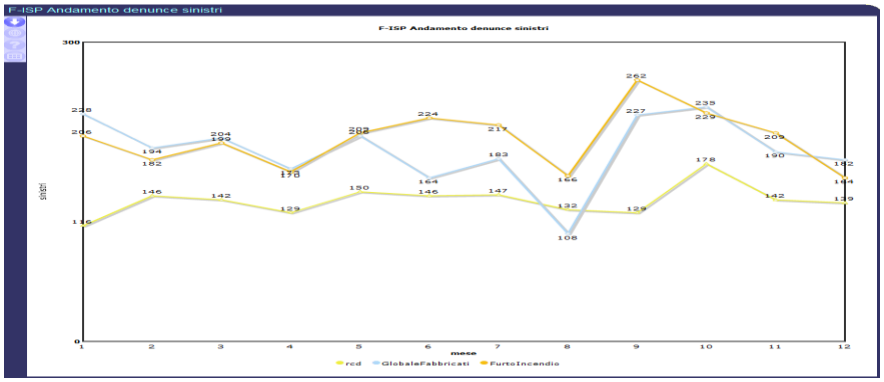

Figure 6. Evolution of claim communications.

The speed of open claim payment during the current year in percentage of paid compared to communicated, without follow-up. The green line is the trend.

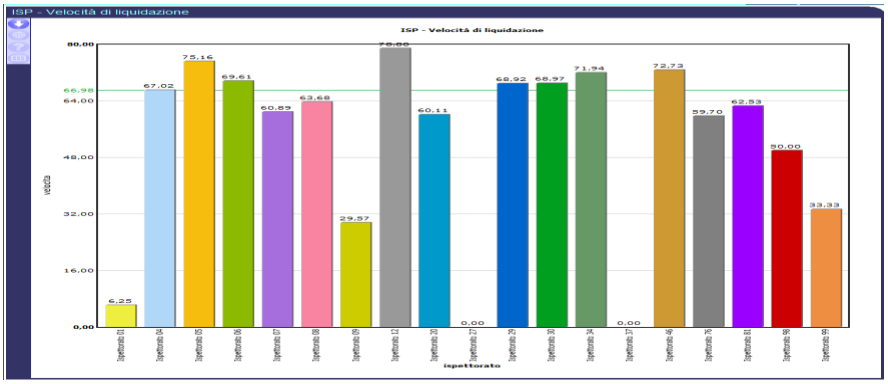

Figure 7. Payment speed.

The ratio between paid claims and reserve on closed claims attributed to the inspectors, with drill down per branch. The green line is the trend.

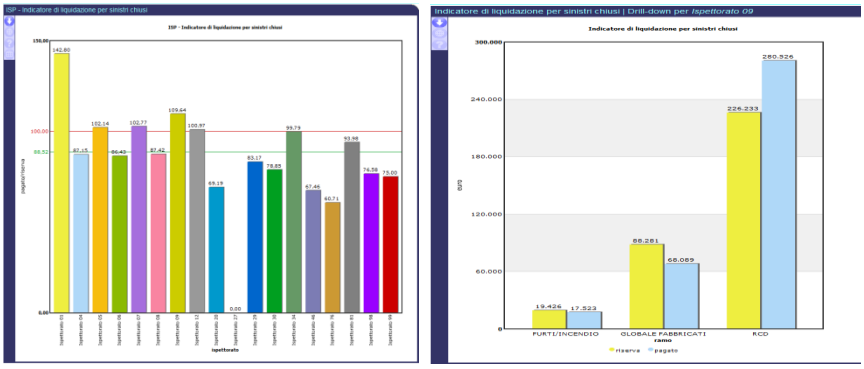

Figure 8. Paid vs reserve on closed claims.

3. Web 2.0 as an environment of non structured data

The word Web 2.0 is used to indicate broadly the level of evolution of the Network (i.e. the World Wide Web) compared to the previous status; to define clearly and widely a complex and constant innovation of the principles of digital communication on the web; to indicate all the approaches and usage methodologies of the network in an innovative way, and therefore of all the online applications that make it possible to obtain a high level of interaction between the site and the connected user (blog, forum, chat, systems such as Wikipedia, Youtube, Facebook, Myspace, Twitter, Gmail, Wordpress, Tripadvisor etc.).

The Web 2.0 is a real innovation, particularly for the approach that it proposes, even if the technologies and tools of the web may seem unchanged (such as forum, chat and blog, which were already there in the web 1.0). In fact the usage modes of these are the aspects that open new scenarios, which imply an active participation by users in the creation and modification of multimedia contents.

This reflects a true digital platform, which makes it possible for the inexperienced user to have a central role, a real interaction and sharing process inside an information exchange community. This is about a way to present the web with non technical dissemination tools, able to describe the evolution that takes place in the social communication, and about the way this communication takes place on the Web.

The Web 2.0 platform is the answer to the proprietary software systems which were typical of the 1.0 era. It solves enormous issues of integration amongst the various systems, which need to communicate amongst each other, but it also orientates the activities of the users inside a controller path with minimum interaction possibility. The scenario changed, and "communication oriented" systems such as the Internet, if considered as platforms, require an interoperation[5], whilst creating an ownerless system that is able to operate through a series of standard protocols and collaboration agreements.

It therefore becomes necessary to provide tools and technologies that support the new Web 2.0 platform, as well as approaches and methods that enable an effective interaction between all the users that are connected to Web. In effect, with Web 2.0, the web becomes a development platform.

The examples show where and how it is possible to navigate in a really interactive way, and illustrate the central communication role of the single user. We could insert many examples but this is not the aim of this document. However, we feel that it is useful to provide two cases that are representative of the way the situation is changing compared to the past.

The first one is about Google. It identifies a real example of web 2.0, born as web-made application, never designed as a package for sale, but provided as a service, with clients who pay directly or indirectly, for the use of this service. No distribution, only use and in any case, free access to the basic service (with possible advanced payable services). The truly innovative factor in Google is the capacity to manage data: the platform does not just provide a set of computer systems with which to interact but a system that is specialised in the management of data. Any computer system, without data, is not particularly useful, it is nearly useless, "…the value of the software is proportional to the scale and dynamism of data that the it helps to manage."[5]

Google does not host the contents that its search service can find. It is neither a server nor a browser, but it provides an interpretation of the intermediary, the guide, the language, the communication tool between the connected user and the community of people and the content waiting for her/him on the Web.

The second case is Wikipedia, an online encyclopaedia born from the idea that any item may be added by any user and modified by any other, provides a truly new communication experiment in which the life cycle of the information (from intuition to use) hides in the platform itself, managed by the technology itself. The information is used in the environment in which it was born. The Wiki technology (leading to Wikipedia, its most famous application, but not only[4]), intended as a collaboration procedure to create a software application, is the target of content management, in that it implements all the principles.

The objective is to create an open space, common to all, in which it is possible to share, exchange, accumulate and optimise knowledge in a cooperative way.

[4] A number of public and private institutions and organisations created applications based on the Wiki principles and methods, to create internal Business Glossary systems where semantic contents are linked to production processes and to the economical context in which they operate.

The Web 2.0 platform makes it possible to acquire an incredible amount of information, of simple data hidden in complex scenarios. It shows the way towards an immediate and free exchange of communication experiences just as in a social exchange.

An advanced Web gives an opportunity to concentrate and move masses of data, in quantities that are sometimes hard to assess, and extremely relevant for the construction of the information assets of people who are clever and experts in succeeding to structure it.

Whilst Web 2.0 represents a very powerful communication tool, it exposes the information that are presented to the user in a non articulated and disorganized way.

The information is structured according to criteria that reflect an organization of published documents that is therefore unable to use the enormous communication potential of the Web. All the information that is presented appears to be the result of a productive cycle that is external to the Web (it is nearly always the case), before being brought back inside the site. The capacity to reorganize and restructure all the available information is missing, although there is a wish to distribute this information according to principles that show its informative characteristics. At this level of analysis, there is a need to find a method that is able to manage and organise information (any kind of information) being disseminated over the Web.

4. The tools and technologies for BI 2.0

If the Web is to success in becoming a powerful knowledge dissemination tool, this has to go through the creation of a system that is perfectly able to organise all that is hidden in it, whilst making it possible to manage all the information in a complete, exhaustive and flexible way, consistent with objectives.

An unusual approach to the problem is needed. It must not be the result of a study carried out in an external environment and then brought back in, but the result of a method aiming at the explanation of characteristics of the Web, which is able to extract the data from the mass of available information.

In practice, we need an instrument and a technology that are adapted for the organisation and restructuring of data (not yet considered as information, according to the equation expressed in section 1.2), in order to build the process of acquisition inside a Business Intelligence system.

The necessity and opportunity to indicate appropriate tools for the acquisition of information from Web 2.0 leads us to the introduction of the concept of Text mining.
Text mining consists in the implementation of non structured data mining techniques (forum, blog, web pages, e-mail and in general, any type of text document) in order to:
- structure and classify documents in predefined categories;
- discover hidden associations (links between arguments or authors, time trends, ...);
- extract specific information (names of companies, names and categories of products, information on purchases, etc…);
- extract concepts for the creation domain ontologies.

The process of Text Mining involves mainly an iterative process, starting from the selection and preparation of data; from the analysis of these to the observation of outputs that may become inputs for the following refining processes, up to the point where results consistent with fixed objectives are reached.
Most of all, it is crucial to identify and separate text information from metadata (or meta-information in the document) in order to simplify positioning of the document in the appropriate scenario[5].
On the text part of the document, it is necessary to implement a grammar analysis and "information extraction" techniques to identify the function of each single section and to define the semantic structure. This makes it possible to reduce the ambiguities and to increase the meaning convergence, whilst introduction classification and sorting criteria, which make it possible to organise documents in thematic groups on the base of a controlled terminology and providing an overview of contents.
The procedures described above appear to be particularly useful when it comes to exploring a large amount of documents to extract summarised and content-representative information, or when you want to sort text without having a consolidated classification system.
In order to assess the important of a tool like Text Mining, with the aim to insert Web derived information inside a Business Intelligence system, you just have to think that

[5] As mentioned above in connection with the simple data coming from any application, we come back to the concept of reference context or scenario in which to position the document, with the aim to attribute a precise meaning to the data included in it. However, it is appropriate to consider that a document is a set of sections, connected to each other, a structure of complex data that may have various meanings linked amongst each other.

the proportion of documents, between “free texts” and “structured texts” reaches a threshold of 4 to 1 in favour of free texts versus structured texts.

The appropriate use of techniques such as Text Mining represents a convincing possible solution to the issue of non structured information hidden in Web 2.0. applications, but it must accompanied by operational and technological tools.

A technological and methodological support may lead to domain ontologies, intended as the definition of logical concepts for the definition of a semantic Web[6]. They provide support to the activity of identification of rules and other specialised operations for interpretation of Web documents, through the construction of a network of relations and connections between the documents, according to logics going beyond the simple hyperlink.

In reference to the aforementioned environment, domain ontologies clearly represent an extremely useful tool to create a semantic glossary and extract taxonomies, with which it is possible to translate non structured date present in a large number of Web 2.0 specific applications and to represent them in the information framework of a Business Intelligence system built for a specific purpose.

Once the information is recovered and riorganised in the data warehouse, as indicated in paragraph 1.2., it will be possible to edit them and analyse them.

In the Web 2.0 platform, the main actor is the user, who defines and publishes information and therefore decides how to view the available data. The user interacts by selecting the information to be compared and the ways to analyse it.

The technology identified to build such a system is mash-up programming, which makes it possible to build an heterogeneous and complex web application able to dynamically include information from various sources, to use the original content of several source system to create a completely new service.

Mash-up obtains information from other systems through API, through feed or simply by writing a Javascript code which allows multiple users (non connected between each other) to combine data from various sites in innovative ways. Moreover, it is rather

[6] Wikipedia, through the phrase semantic web, which was invented by its creator, Tim Berners-Lee, is intended as the transformation of the Web in an environment where the *documents* published (HTML pages, file, images, and so on) are associated to information and data (metadata) which define its semantic context in a format that is relevant to the querce, interpretation and more widely, to automatic processing.

easy to create the structure of a mash-up application , which uses web services[7] as technological reference points. These are tools that make it possible to build a single large container for the exchange or visualisation of information from various source or for construction of business processes.

The mashups, which can be mainly divided in three broad categories, represent the ideal support to ensuire the merger of two or more services combined together in a single graphic template (consumer mashup), in order to make it possible to merge various heterogeneous information sources in order to create a single data container (data mashup) but most of all to combine data from various sources (both internal and external to the application involved) with the creation of an integrated framework in mind. This last process, defined as business mashup, may be implemented for the analysis of specific corporate processes such as sales monitoring or staff and cost management inside a company, or to build marketing strategies. To summarise, a business mashup is the combination of consumer and data mashup aimed at building enterprise applications.

Subsequently, the information publishing system, which currently appears in Business Intelligence systems as a large frame, predefined for the analysis of previously defined information, would see its meaning entirely reviewed, no as a corporate static and predefined frame but as a dynamic system both in terms of contents and in terms of a configuration allowing any user to create his / her own “mashboard” using off-the-shelf tools such as widgets (tiny interactive preset applications which any user may search, select, organise or tailor).

To conclude, a business intelligence system needs to integrate an ever growing and ever more detailed set of data, able to describe the information aspect of corporate processes in a comprehensive way. Typical Web 2.0 applications represent a mine of strategic information for decision support systems and this is exactly the reason why an ever growing number of companies is investing in the creation of analytical systems or so-called Business Intelligence 2.0 systems.

[7] Web services can be defined as a collection of data standards and formats which make it possible for web applications to communicate with each other and to share information. They represent applications that were developed in various programming and implementation languages on various platforms.

5. Conclusion

After presenting the principles, rules, processes and objectives that regulates the life of a business intelligence system, a proposal was made for a real use case that shows its real potential and possibilities for application in all sectors of economic, scientific, social activity, etc. At the same time, we have highlighted the ways of operation of the user, in a cognitive system (BI), and the need that it operates and interacts directly with the tool in a way as flexible as possible. At this point, we would just introduce and describe the Web 2.0 that represents a new computing platform extremely innovative also in its operational criteria, where the user is the true interpreter of the exchange and the dissemination of information. Finally, combining the methodological and operational ideas from Web 2.0 with processes, rules and above all with the goal of achieving a system of business intelligence, we introduced the possibility of using technologies and methodologies of the "mashup programming" to achieve highly innovative instrument in order to publish the information in a cognitive system (BI) that constitutes a decisive phase for the distribution of data between knowledge workers.

References

[1] T.P. Gerrity, Design of man-machine decision systems: An application to portfolio management. *Sloan Management Review 12* (1971), pp. 59–75.

[2] Ralph Kimball, The data warehouse lifecycle toolkit (1998).

[3] William H. Inmon, Exploration Warehousing: Turning Business Information into Business Opportunity (2000).

[4] Richard J. Roiger, Michael W. Geatz, Introduction to data mining (2003) pag. 95.

[5] Tim O'Really, Design Patterns and Business Models for the Next Generation of Software (2009) pp. 1-2.