# MUSING Annual Public Report (2010)

Next Generation
Business Intelligence

http://www.musing.eu/

In April 2010 MUSING was coming to an end. In this last annual report, we will thus certainly focus on the final dissemination event of MUSING, which has been taking place within the context of the 20th International XBRL conference[1], but we will also give an overview of the main topics dealt with by the project in the four years it has been lasting. As a large Integrated Project (IP), MUSING has also been confronted with project internal issues, like for example partners having to withdraw from the project, and to be replaced by new partners. And in the last year it was the project coordinator, the company METAWARE S.A. that had to leave, due to its liquidation. In this case the consortium opted for an internal solution and one partner took over the role of the coordination in order to conduct the project through the final dissemination event and through the final review of the project, which took place on the 1st and 2nd of July 2010. In this the consortium has proven its ability to react to a critical situation, but we have to stress that this achievement would not have been possible without the very efficient support of the Commission in the person of the project officer of the project.

In the following, we will first summarize the general goals of MUSING, and then present the activities that have marked the last year of the project.

## Goals and Motivation of MUSING

MUSING was an FP6-funded Integrated Project on semantic-based knowledge management applied to Business Intelligence (BI).

A main goal of MUSING has been to respond to actual needs of BI and to anticipate further developments in this field considering the development and use of new technologies resulting from the integration of Semantic Web, Natural Language Processing and Quantitative & Qualitative Statistical approaches in the broader fields of Company Information and Risk Management, in order to support for example business decision procedures. Through knowledge management technology, advanced predictive analytics and intelligent access to third party data, MUSING solutions increased the precision, consistency and agility of business decisions.

While business decisions are a critical source of value, making the best decision was beyond the capacity of most business systems, when decisions must be made faster, leveraging more data from different sources, under greater regulatory demands and competitive pressures, and with more complicated constraints and trade-offs. Pressure points on business decisions we were considering at the beginning of MUSING were:

1. the data available to make decisions has ballooned, but there are challenges to using all that data – in particular the unstructured data – effectively;
2. organizations must comply with more new regulations, stricter and more complex rules, shorter deadlines and greater consequences for non-compliance;
3. decisions that once took days, now have to be executed within rapidly shrinking timeframes;
4. business objectives involve trade-offs between risks, resource constraints, opportunity costs and other factors.

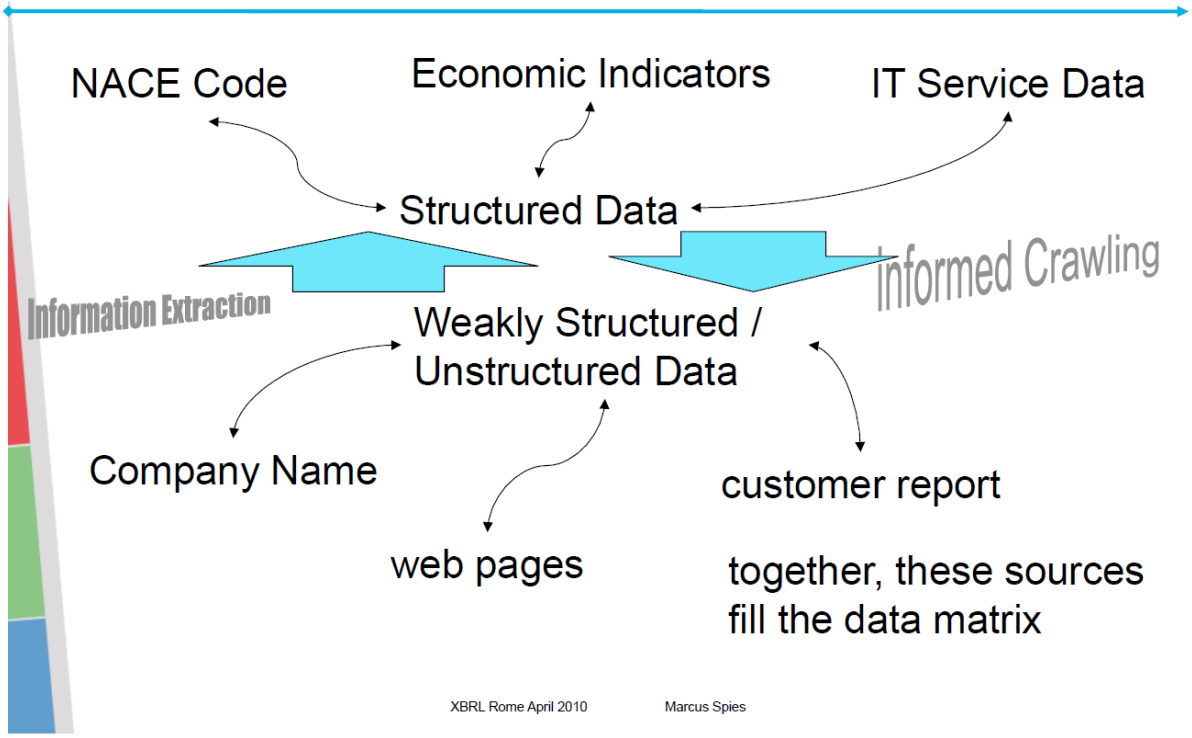
## Results of MUSING

In the course of the project the consortium could develop an integrated framework that offers possible R&D solutions to the points 1.) 3.) and 4.) above, while the topic under 2.) turned out to be more specific and relies less on available data to be transformed onto semantic facts, but on a more

[1] See http://20thconference.xbrl.org/, for which MUSING acted as Silver Sponsor

complex transformation of lengthy regulation and law documents onto a logical representation. As such we have to say that this is an unsolved issue for now, but that in MUSING nevertheless the problem has been addressed and that we assume that there can be solutions to "upgrade" regulation texts onto a (partial) semantic framework that can be "digested" by machines.

MUSING thus explored ways to automatically extract relevant information from heterogeneous data, including unstructured data, and to automatically classify, analyze and to integrate it, on the basis of the use of semantic web technologies – mainly ontologies and reasoning procedures – to automatically discover and characterize trends and patterns in it, transforming thus data into information and knowledge that can be used in concrete decision scenarios. Combined efforts in research fields such as statistics, data mining, natural language processing (NLP), information extraction (IE), knowledge engineering and machine learning have been contributing to those results. The figure below, which is taken from a presentation given by the Scientific & Technological Coordinator of MUSING at the 20th XBRL conference, in Rome (April 2010), shows the integration and merging of various types of information in MUSING:

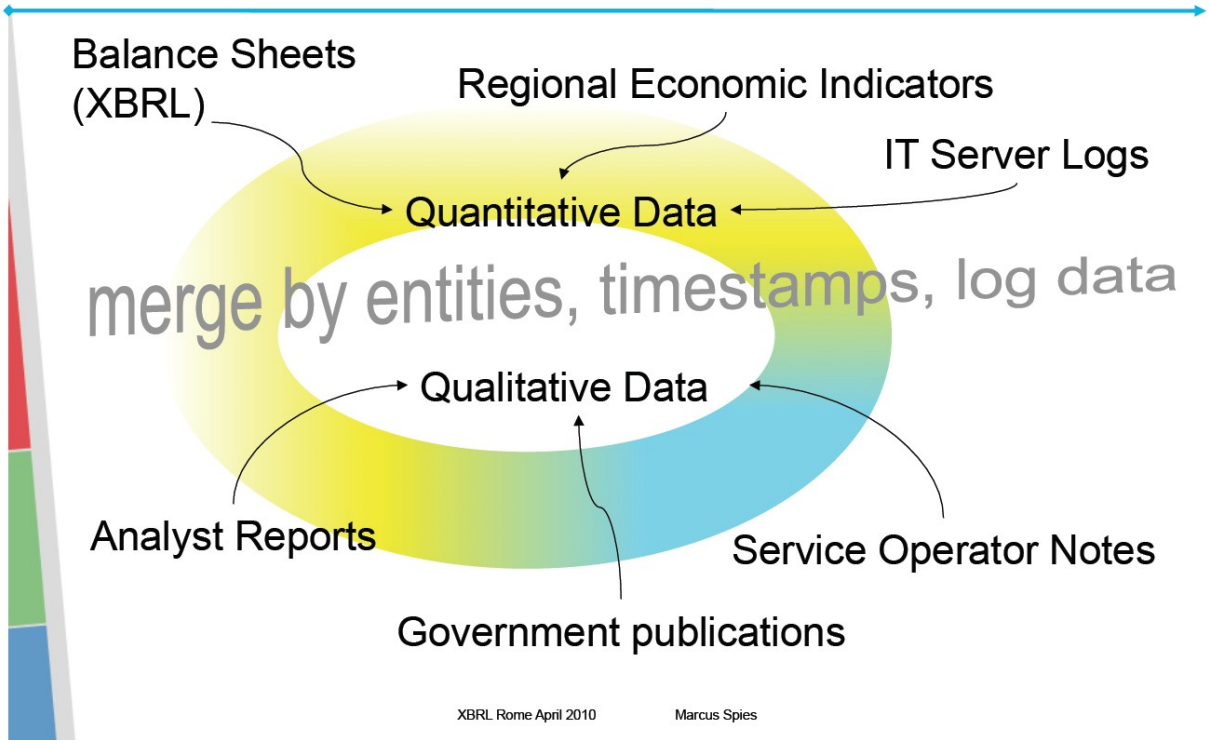



Besides the R&D aspects, MUSING addressed the use of such results in concrete services to enable perceptive Business Intelligence (BI) and enterprise's self assessment capabilities and reducing the risk in business decision making. For this, various batches of pilot applications have been specified and implemented in the context on three-so-called Vertical Streams, whit their specific domain of application:

- **Financial Risk Management (FRM):** Development and validation of next generation (Basel II and beyond) semantic-based Business Intelligence solutions, with particular reference to Credit Risk Management and Access to credit for enterprises, especially Small and Medium Enterprises (SMEs).
- **Internationalisation (INT):** Development and validation of next generation semantic-based internationalisation platforms supporting SME internationalisation in the context of global competition by identifying, capturing, representing and localising trusted knowledge.
- **IT-Operational Risk & Business Continuity (IT-OpRisk):** Semantic-driven knowledge systems for IT operational risk measurement & mitigation, in particular for IT-intensive organisations. Management of business continuity & operational risks of large enterprises and SMEs impacting positively on the related user communities in terms of service levels and costs.

To enable knowledge and tool re-use in the pilot applications of these 3 streams, MUSING developed an integrated Service Oriented Architecture platform, which delivers B2B services that business partners can incorporate in their own tailored application; in this view, the platform delivers its services via Web Services, the present standard for systems interoperability.

All solutions in integration and in vertical streams were based on the state-of-the-art technologies. Here again a figure taken from the presentation by the Scientific & Technological Coordinator of MUSING at the 20th XBRL conference, in Rome (April 2010) in order to visualize the type of integration MUSING has been performing for the data available across the 3 vertical streams:




## Summary of Activities

The MUSING project started in April 2006 and, after its last year of activity, it has delivered and validated a third batch of intelligence solutions (services) described during the specification phase.

From an industrial point of view, the past two years brought substantial progress in implementations of pilots from all vertical streams enabling in the final year the validation procedure as defined according to the GQM (Goal Question Metric) approach that was specified and implemented in MUSING. Key pilot applications are all accessible via web applications from the MUSING platform, which uses a business process execution language (BPEL) engine for the service integration layer as a common representation of information processing steps in all pilot applications.

Relevant activities carried out by MUSING are detailed below.

### 1.1 Market and Context for MUSING services

MUSING offers BI tools and solutions based on semantic technologies that can be offered via web-services, opening the immediate market and take-up to organisations of all sizes – i.e. SMEs and micro-businesses, as well as large enterprises – and Public Administration. BI is carried out in a variety of applications, from financial services and insurance, to internationalisation and IT operational risk measurement, management and mitigation.

In the field of **financial risk management (FRM)**, MUSING has developed and validated semantic-based BI solutions, with reference to Credit Risk Management and Access to credit for enterprises, especially SMEs that might not have enough human resources, or the capital required to hire them, for collecting and storing the requested complex information on financial risk management. MUSING's solutions are useful both for SMEs and for Financial Institutions that need to gain valuable information from other channels, to evaluate the financial health of SMEs (through external credit merit

assessment, BI support to Basel II-compliant rating systems, automatic population of XBRL taxonomies from financial prospects) as is the case for large enterprises.

MUSING has approached the **Internationalisation (INT)** market through the development and validation semantic-based internationalisation platforms. Their main goal is to support internationalisation of SMEs (through partner search and analyses, target market analysis and selection), in the context of global competition, and to provide an economically viable alternative to today's highly priced personalised services.

In reference to **IT Operational Risk Management (IT-OpRisk)**, the project addressed the development and validation of semantic-driven knowledge systems for risk measurement and mitigation, with particular reference to those operational risks faced by IT-intensive organisations. The MUSING semantic-based BI platform helps in measuring and controlling the operational risks of large enterprises, as well as SMEs, impacting positively on the community, both in terms of service levels and costs.

In the last year of the project special emphasis has been put in making the solutions in those 3 streams interoperable among each other, as far as possible, also on the base of the sharing of information available at the semantic level. So for example the risk ontologies have been distributed between the FRM and the IT-OpRisk domains.

## 1.2 Technology Outlook and Innovative Features

BI has typically been used to present data so that report consumers can try to figure out what happened after the fact. MUSING solutions support information professionals to make better-informed decisions that are more aligned with corporate objectives, where decision making will no longer relies solely on knowing “what happened” and will be supported by comprehensive intelligence about “what’s happening now,” and also, by extension, “what is likely to happen.” In addition, MUSING was about tapping into relevant content, wherever it is and whatever its format, using content analytics tools that take advantage of the latest technologies and are fully web-based for wider enterprise reach and availability.

In particular, integration of semantics and statistics gives business information management systems (data warehousing, content management, log systems) the intelligence needed to operate on the “efficient frontier”, where risks, costs and losses are minimized, while efficiency and compliance are maximized. That’s exactly where MUSING took off to the development of innovative BI solutions that integrates:

- Merging of information extracted from heterogeneous sources
- Knowledge management & reasoning, including temporal reasoning
- Combined use of declarative rule-based approaches and statistical methods in semantic technology, for ontology population and ontology learning
- Bayesian integration of qualitative and quantitative knowledge elements

The technological development in MUSING is taking place at various levels, and in general, it can be said that MUSING is leveraging SOA (Service Oriented Architecture) and BPM (Business Process Management) technologies and languages for achieving a highly flexible BI platform architecture that integrates services from the different technology areas mentioned above into highly dynamic composed services. This enables MUSING platform users to address BI service demands from customers or specific markets in a rapid manner.

At a lower level, MUSING is dealing with data collection and basic data analysis. This includes for example the uploading of balance sheets, as structured data, and their mapping into a standardized format like XBRL, or a basic analysis of system logs. In periodic reports of companies there are normally annexes to the balance sheets, in an unstructured textual form. The basic analysis steps of those documents are concerned only with the extraction of metadata and the linking of text parts to elements of the structured balance sheets. Semantic analysis of both the structured and unstructured data takes place at a higher and semantic level of processing, which allows abstracting over the particular format of the data.

There, relevant information is extracted from the various data sources, merged and mapped into instances of ontology concepts/classes for ensuring interoperability of the extracted data/information. Here is the knowledge representation level, the knowledge being encoded in ontologies, which encompass beyond generic knowledge of the world all the relevant knowledge of the strategic domains, whereas the instances gained from the analysis of the documents represent particular

instantiations of the concepts/classes, and are said to *populate* the ontologies, At this level, MUSING also provides means to access the knowledge, to update it and to check consistency.

While continuous evolution of the ontologies is possible (and inevitable), the foundational infrastructure for dealing for the various types of data, information, knowledge and the associated methods and tools have been laid in such a way to allow further refinement without dramatic changes in the structure, resulting in immediately usable ontologies. On top of that, models exploiting this knowledge, for example for supporting decision procedures, are now in an mature state of development and statistical and data mining models are being used to exploit additional semantic features that ontologies provide, as opposed to the typical lower level raw features.

One major usage of ontology here is as a source of structured knowledge for the process of extraction of meaningful information from textual data. With the aid of specific natural language processing steps, ontologies can support complex information extraction tasks like named entity recognition, identification of evaluative statements about a company or a product, classification of reported events in customer relationship management data etc.

Ultimately, integration of the MUSING components has started based on the requirements defined by the numerous pilots defined. So, for example, a framework for the representation of temporal information has been developed in MUSING, including as well a time ontology, which has been integrated in the domain and upper level ontologies. And since not all knowledge can be gained from the documents, but quite often on the base of reasoning, an integrated reasoning architecture has been designed and implemented, using and combining already available reasoning components, exploiting their strengths and avoiding as far as possible their weaknesses.

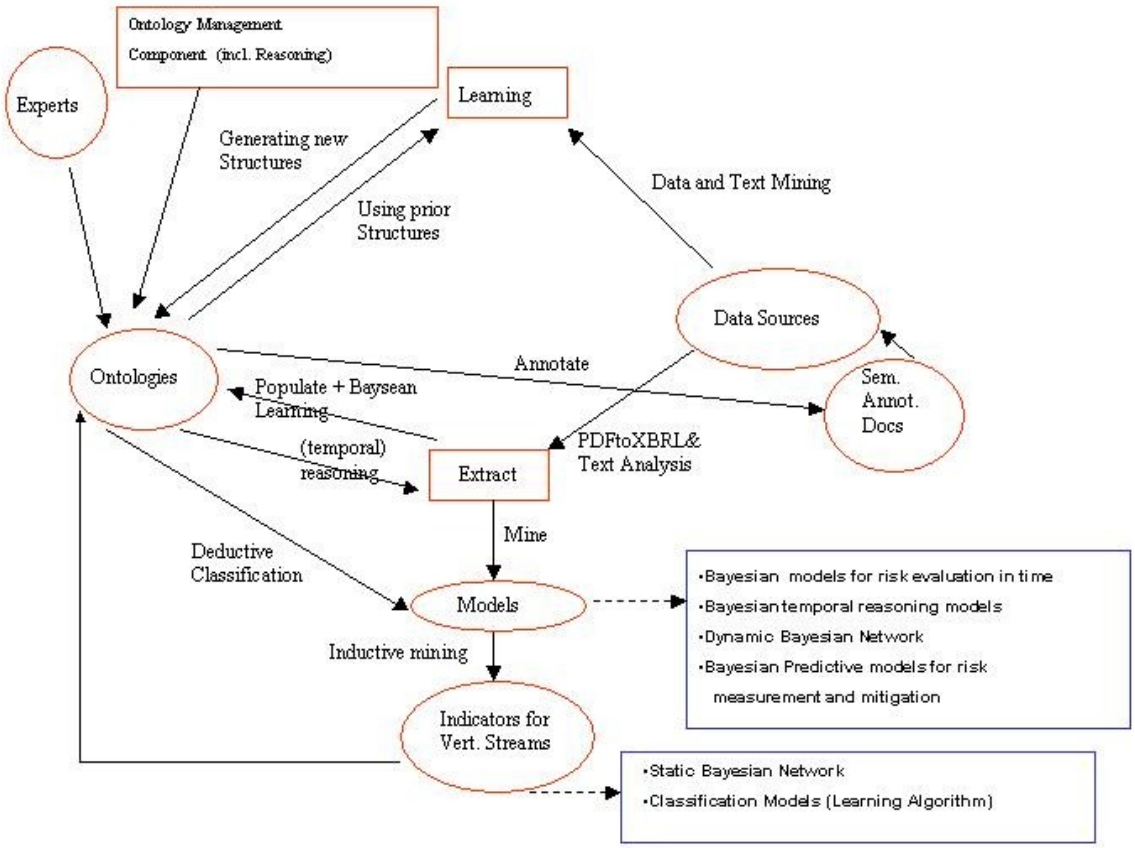



**Figure 1: Technological Foundation Components**

## 1.3 Results of Field Tests and Services Delivered

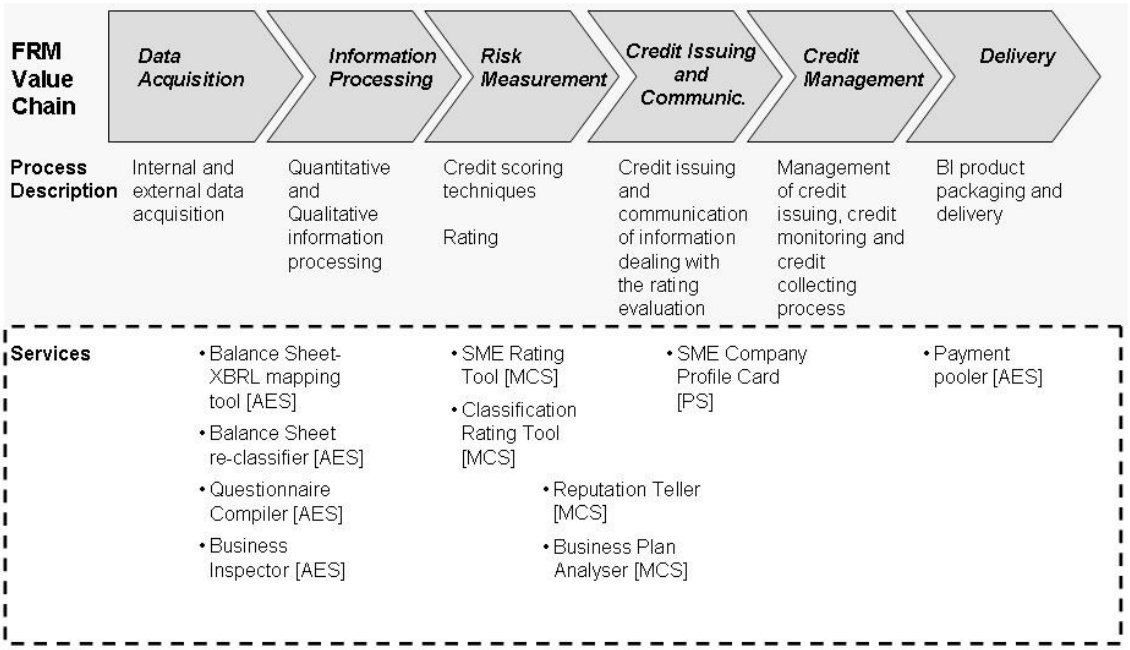
Focused, industry-driven pilots have been carried out by MUSING throughout the 4 years of project

activities. As a direct result, a number of pre-competitive services have been consolidated on three domain-specific platforms, as indicated below:

1. Financial Risk Management
    a. *BS-XBRLmapping tool*: Automation tool to populate XBRL taxonomies from typical Balance Sheets (BS) of enterprises, typically provided in PDF format (but also available sometimes in proprietary XML format, or in HTML documents).
    b. *Balance Sheet Re-classifier*: Tool to automatically extract, upload and reclassify (i.e., re-read the items according to agreed-upon principles utilised by the financial institution to assess the B.S.) quantitative information from a Balance Sheet in pdf format, ready to be inserted in financial institution internal systems (typically, Basel II-compliant rating systems).
    c. *Business Navigator*: Visualisation service to provide business information related to a specified company or person through intuitive navigation. The Business navigator represents an innovative and powerful visualization tool that allows users to interact with data, showing relationship patterns or drilling down to details, and so on, on-the-fly. With this tool users can navigate a large variety of information about companies and ownerships, with particular focus on – but not limited to – the connection between companies and persons.
    d. *Business Plan Analyzer (BPA).* An merit-centred service for supporting both the self-assessment of Small and Medium Enterprises (SMEs) in the analysis of their Business Plans and the decision making process of Financial Institutions and Investors in judging business proposals. It is a semantic-assisted evaluation tool that merges a classic classification technique with prior (domain) knowledge. The basic idea is to learn from past cases, reporting on the characteristics of the company and its business plan and on its successes or failures, thus producing a classifier capable of forecasting whether a new case is likely to succeed or not.
    e. *Non Linear rating statistical model.* Company rating is one of the most common activities of Business Intelligence and considering the attention placed on SMEs in the new Basel Capital Accord, we propose a set of Bayesian and classical longitudinal models to predict SME default probability, taking unobservable firm and business sector heterogeneities as well as analysts recommendations into account. The purpose of this quantitative model is not just that of providing such a rating functionality, but to do so by exploiting more information than usual, and cutting-edge rating algorithms. A more general, quali-quantitative nonlinear rating model for SMEs is under development.
2. Internationalisation
    a. *Region Selection*: A model to select the region of a country where a company is expected to be more successful doing business. The idea is that basing on characteristics of a firm and of the operation it is seeking to perform abroad (provided by way of a questionnaire), and of the regions of the selected country, the model suggests which region would be the best choice, and the characteristics of the region that were more important in determining its score, also basing on historical data about other companies. The region Selection Model accesses data about the regions, fetched from various locations, mostly the web, and annotated.
    b. *Partner Search*: Semantic-based services to support an intermediate company in the partner search activity for clients in the process of internationalising their business: the aim of the pilot is to show how the MUSING services allow an Enterprise to analyse and search for a partner in a country different from the one in which it usually operates; so far, the case India has been analyzed.
    c. *International Enterprise Intelligence (IEI)*: the goal of this pilot is to develop a BI solution that offers, to a variety of users, a knowledge-base on European enterprises, with progressive level of information available to users.
    d. *Counterpart monitor*: The monitoring service extends the standard IEI service of providing business press information. It is controlled and managed by IEI. The pilot deals with the integration of automated monitoring services between a business information provider and a financial institution as an aggregator of internationalising enterprises.
3. IT - Operational Risk
    a. *Addressing IT – OpRisk at a Telco*: the application of MUSING will be in producing risk assessments and mitigation strategies for a Telco company.

b. *Assessing The Risk of A Customer's Business to a VNO*: This pilot will concentrate on assessing the risk, taken on by a Virtual Network Operator (VNO), when it decides to take on a new customer, as a result of that customer's IT-OpRisk. The VNO, in addition to getting a better view of its own risks, will now use MUSING capabilities to estimate the IT-OpRisks of its potential customer as well as their financial risks, in order to arrive at a better informed decision regarding the acquisition or the maintenance of the customer by the VNO.
c. *Business Continuity Management*: clearly, Business Continuity Management is inexorably connected with Operational Risks in general and with IT Operational Risks in particular. Thus, MUSING will combine these two adverse-scenario management approaches into one comprehensive approach, and will provide facts-based solutions to various potential business disruptions that may be caused by information technologies employed in the business.

A schematic representation of the value chains of the Vertical Streams investigated by MUSING is reported in Figure 2 below. The positioning of the MUSING services (those already available as well as the to-be ones) is emphasised.



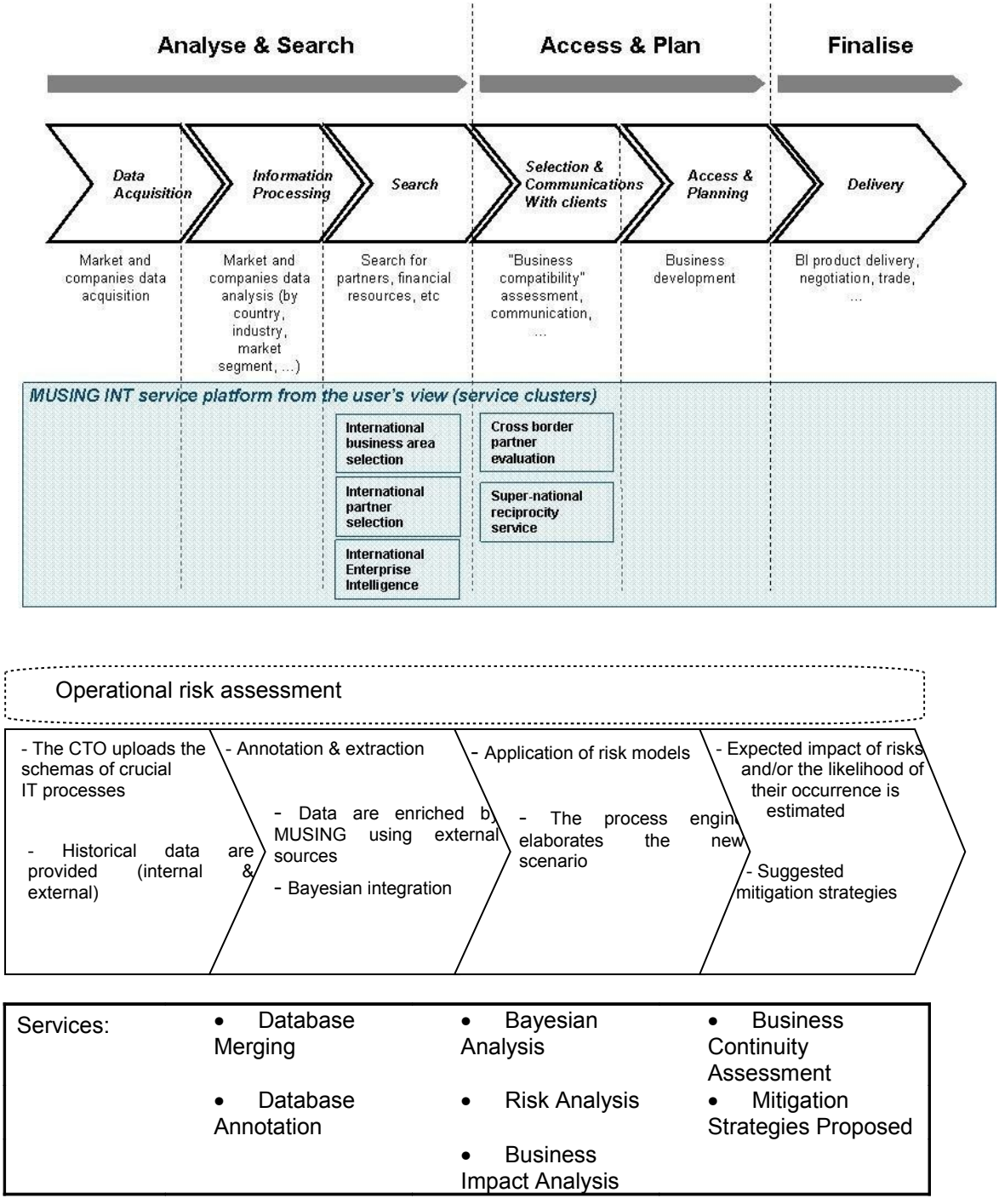



**Figure 2: Value-chain representation of typical to-be MUSING BI services**

## User Involvement, Promotion and Awareness

Since the beginning of the project, the MUSING Consortium partners have been particularly active in disseminating the work done to both scientific and industrial audiences. As a result, a large and steadily increasing number of high-quality, mostly peer-reviewed scientific publications and tutorials on aspects of MUSING has been produced, reflecting the multiple scientific fields and disciplines that are represented within the project. Presentations and scientific publications by MUSING Consortium Partners are made available on the dissemination website at http://www.musing.eu and on the web pages of the partners.

In this actual report, we stress only the outstanding dissemination events that have occurred in the last year of the project.

**The MUSING final event**

The decision of the MUSING consortium was to have the final event of MUSING in the context of the 20th XBRL conference (http://20thconference.xbrl.org/). Due to the importance of XBRL international in all fields of reporting and the new direction adopted by XBRL towards Business Information in general, this turned out to be an excellent opportunity for a wide dissemination of MUSING, and for presenting our final results to a public of experts in fields related to most of the topics of MUSING.

In spite of the volcanic ash cloud from Iceland, four representatives of MUSING managed to make the trip (by night and by train for some of them).

As a silver sponsor of the XBRL conference, MUSING had a large booth, and we could there distribute information to all interested participants. The MUSING logo and a link to the project web page were also added to the web page of the XBRL conference. Additional advertisement material has been printed on leaflets, bags, large posters etc. of the conference.

MUSING has been also presented within 4 tracks:

1. Track 4: Finance: Business Registries, Capital Markets & Banking
2. Track 5: XBRL Technology (Specification & Implementations)
3. Track 8: Academic: "Celebrating a Decade of Contributions to XBRL by Academics around the world. Where Do We Go from Here?"
4. Track 9: XBRL Solutions

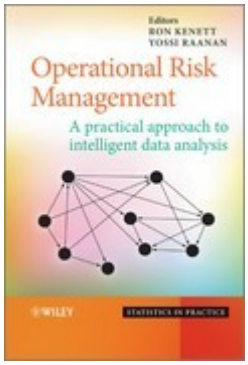
The main contributions for MUSING were in Track 5 and 9, where the S&T Director of MUSING gave respective talks on ***Towards an ontology model of XBRL reporting taxonomies*** and ***Towards the semantics enabled next generation Data Warehouse – Results and Industry applications of the EU MUSING project.***

Both talks have been extremely well received and a long term impact of MUSING within XBRL is to be expected within the XII committee on best practices, which is taking an orientation towards business information (instead of “only” reporting”), and which is taking into account many of the contributions of MUSING.

All in all we can say that with the organisation of the final event as a silver sponsor of and within the context of the 20th XBRL conference, MUSING has been able to raise a very wide awareness towards the potential of semantic- based approaches to business intelligence.

**The MUSING book**

The MUSING partner KPA manage to finalized the edition of a book dealing with many aspects of MUSING and involved nearly all the partners of the project: “**Operational Risk Management** A practical approach to intelligent data analysisStatistics in Practice.” See http://www.wiley-vch.de/publish/en/books/ISBN978-0-470-74748-3




**Other dissemination results in the last year of MUSING**

- Monika Jungemann-Dorner (VVC) and Marcus Spies (S&T Director) have been contributing to an extensive presence of MUSING at the European TDWI conference in Amsterdam (2009). This conference is sponsored by MUSING, with the additional benefit of having the MUSING logo and introductory text visible on the web page of the conference and on other material of

the conference. The MUSING partners report on good contacts with visitors. The interest of the DW community to the topics of MUSING has been brought to us by the S&T director, who is continuing, after project lifetime, to present topics of the project to this community.

- DFKI presented the B3.4 Pilot (Monitoring) at a DFKI "Hausmessse" (House Trade Fair), 2009, and at Cebit (2010), the biggest IT trade fair in the world. Interesting contacts resulted from this presentation, and the idea is to extend the MUSING use case to the web presence of small towns in an e-Government application. This is a significant result, since the re-use of semantic-driven information extraction seems to be re-usable to other domains of application.
- Christian Leibold (UIBK) presented a MUSING paper at a W3C and XBRL workshop, and it turned out that this presentation has been quite influential on the position paper of XBRL for future developments. This work is also very relevant for the contribution of MUSING to standardisation activities.
- Diana Maynard USFD), invited talk, "GATE: Bridging the Gap between Terminology and Linguistics", TIA (Terminologie et Intelligence Artificielle),Toulouse, France, November 2009. http://www.irit.fr/TIA09/ProgrammeEN.html#keynote2
- Thierry Declerck was invited research at the Université of Toulouse (1 month), also due on his work in MUSING. He is teaching there also on MUSING related topics.
- Diana Maynard (USFD) and Johanna Voelker, tutorial, "Natural Language Processing for the Semantic Web Tutorial", LREC, Malta, May 2010. http://www.dcs.shef.ac.uk/~diana/courses/lrec-nlp-semweb-tutorial.html
- Two modules of the Second GATE Training Course (Sheffield, May 2010) included material from MUSING: "Advanced GATE Applications" used our MUSING IE applications as examples of customizing ANNIE and using sophisticated conditional processing; "Advanced Machine Learning" included our work on sentiment analysis from MUSING.
- The same two modules of the Third GATE Training Course (Montreal, August & September 2010) will include material from MUSING.
- Thierry Declerck was invited speaker at a APA-IT event in Vienna (12 November). APA-IT is the IT Company working for the Austrian Press Agency. Part of his talk was about MUSING.
- DFKI had a MUSING presentation at LREC 2010
- Various Dissertations and Bachelor or Master thesis (UPI, UPV, DFKI, S&T coordinator), with a clear MUSING background has been terminated during the life-time of the project. We name here only some dissertations:

    - Andrea Bellandi, "Extending Ontology Queries with Bayesian Network Reasoning", IMT Institute for Advanced Studies, Lucca, 2008
    - Barbara Furletti, "Ontology-Driven Knowledge Discovery" ", IMT Institute for Advanced Studies, Lucca, 2009
    - Andrea Romei, "XQuake: an XML-based Knowledge Discovery Environment", Dottorato di Informatica Università di Pisa (2009)
    - Valerio Grossi, "A New Framework for Data Streams Classification" ", Dottorato di Informatica Università di Pisa (2009)

    - Mihaela Vela, Extraction of Ontology Schema from Financial News (to be read in the Winter 2010-2011), University of Saarland.

# Future Work or Exploitation Prospects

Some of the topics of MUSING are going to be further investigated within R&D projects co-funded by the Commission and with participation of former MUSING partners, like

- FIRST : Large scale information extraction and integration infrastructure for supporting financial decision making (http://cordis.europa.eu/fetch?CALLER=PROJ_ICT&ACTION=D&CAT=PROJ&RCN=96468)
- MONNET: Multilingual Ontologies for Networked Knowledge (http://www.monnet-project.eu/), with the participation of XBRL-Europe.

Many partners have their specific exploitation plans and cooperation among some partners is envisaged.

The consortium also agreed to make available all the resources that are not specific to a company. This will consisted mainly in all the ontological resources of MUSING and some annotated corpora.