

MUSING Annual Public Report (2009)



Next Generation
Business Intelligence

<http://www.musing.eu/>

Business decisions are the next critical source of value, but making the best decision is well beyond the capacity of most business systems today, 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 are: a) the data available to make decisions has ballooned, but there are challenges to using all that data – in particular the unstructured data – effectively; b) organizations must comply with more new regulations, stricter and more complex rules, shorter deadlines and greater consequences for non-compliance; c) decisions that once took days, now have to be executed within rapidly shrinking timeframes; d) business objectives involve trade-offs between risks, resource constraints, opportunity costs and other factors.

So, we must find ways to automatically extract relevant information from unstructured data, to automatically classify this information, to automatically analyze it, to automatically discover and characterize trends and patterns in it. In a nutshell: getting the most value we can – namely, effective knowledge - from our data. Researchers in areas such as statistics, data mining, natural language processing (NLP), information extraction (IE), knowledge engineering and machine learning are strongly contributing to this field. MUSING delivers next-generation knowledge management solutions and services to enable perceptive Business Intelligence (BI) and enterprise's self assessment capabilities and reducing the risk in business decision making. Through knowledge management technology, advanced predictive analytics and intelligent access to third party data, MUSING solutions increase the precision, consistency and agility of operational business decisions.

Next generation business intelligence

Information professionals don't just need content, they need the tools that can meld content and context, creating a synthesis of information and analysis that can guide business decision making. And this synthesis is key to next generation Business Intelligence (BI).

BI has typically been used to present data so that report consumers can try to figure out what happened after the fact. Analysis lets them further explore that data to determine trends and probabilities or uncover root causes of successes and failures.

In addition to making BI, reporting and analysis more proactive and more usable and more available across the enterprise, the next generation of BI also will be 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 natural language processing 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 takes 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

MUSING is an FP6-funded Integrated Project on semantic technology enabled knowledge management applied to Business Intelligence. Through knowledge management technology, advanced predictive analytics and intelligent access to third party data, MUSING solutions increases the precision, consistency and agility of business decisions. In particular, MUSING addresses specific target vertical industries:

- Financial Risk Management (with particular reference to Credit Risk)
- Internationalisation Services (with particular reference to location and partnership selection)
- IT Operational Risk Measurement and Mitigation.

Summary of Activities

The MUSING project started in April 2006 and, after its third year of activity, is about to deliver on its third batch of practical solutions defined during the specification phase.

During the first phases of activity, much attention was posed on carefully defining the specifications for the technology components to be used for the MUSING project services; in particular, great emphasis was given to the definition of the global structure of the developed and used ontologies that provide the knowledge schemas supporting BI activities in the three strategic domains (“Vertical Streams”) currently dealt with by MUSING. Moreover, an in-depth and thorough market study was carried out, in order to describe current players and potential users of the MUSING services, to understand their requirements, and to acquire knowledge of the current solutions adopted by them.

A further step was the definition and specification of several application scenarios, in order to delineate and build the framework where MUSING will act, and to begin giving a flavour of what the final services will be.

During the third year of the project lifespan, all activities planned were successfully carried on and increasingly focussed on achieving tangible results in the end user communities of the MUSING vertical streams.

From a technological and scientific point of view, the Consortium mainly worked on the core components that are part of the MUSING platform. An important group of activities was related to technical / scientific integration in order to prepare a common runtime environment for MUSING services (in particular, the usage of web service composition for integrating statistical processing, NLP and ontology management services). This runtime environment, also referred to as the MUSING platform, features a Service Oriented Architecture (SOA) approach to next generation BI. We envision future BI as part of an Internet of Services building on flexibly composable functional building blocks, integrating analytical applications building on a wide variety of approaches (Bayesian networks, natural language feature detection, reasoning, etc). From the implementation point of view, MUSING uses a business process execution language (WS-BPEL) engine for the service integration layer. However, integration of all MUSING services does not come for free. Significant activity from the MUSING

research partners was needed to enable interoperable interfaces for providing their services in the format required by WS-BPEL (as partner web services).

A further key activity was the implementation of a “house-of-quality” supported process of scientific/technical prioritization of the MUSING last batch of application pilots in the vertical streams. This activity was supported by the completion of the validation report on the previous batch of pilots: the results from the validation process were integrated with the business based preferences of the vertical stream in order to define a coherent set of pilot applications that are ensured to be supported by the MUSING platform.

The validation analysis has revealed that coherent component development, implementation of interfaces and technical documentation are crucial for the realization of the service oriented architecture as envisioned by the MUSING project plan and to be realized in the next development phase. The MUSING phase of pilot developments has considerably extended the portfolio of technology demonstrators of the general objectives in this project. With respect to the outcome of this pilot validation phase, the key priorities for upcoming next pilot development are therefore:

1. Platform / service integration according to principles of service oriented architectures
2. Usage of ontologies throughout all pilots, including massive ontology population requiring advanced ontology management and versioning
3. Selection of the most promising domains for information extraction, building on and extending domains investigated already
4. Definition of pilots that together form a reasonable portfolio for demonstrating all key MUSING technologies across MUSING vertical streams.
5. Defining the business model approaches for every pilot and the Musing platform respectively.
6. Appropriate technical documentation of all service interfaces and platform scripts that allow reuse of the components by the MUSING community itself and beyond the MUSING project.

As a visible result, the key pilot solutions are now accessible via web applications from the MUSING internal server. Some additional services that are not yet part of a pilot have been nonetheless deployed as web services. The usage of the web services based process execution language WS-BPEL as a common representation of information processing steps in all pilots has been fully implemented.

From a dissemination point of view, all the MUSING consortium actively worked on disseminating the work done within the project to both scientific and industrial audiences. A large and steadily increasing number of high-quality, mostly peer-reviewed scientific publications and tutorials on aspects of MUSING has been constantly produced. The public web portal (<http://www.musing.eu>) plays a pivotal role in disseminating the project and its initiatives, distributing public documents, and creating a community of professionals, organizations, and institutions interested in MUSING, now including over 170 members. In January the first issue of the MUSING newsletter was released, addressing the mailing list built through the public web portal; since the first issue another 3 newsletters have been released for Spring, Summer and Autumn respectively.

Finally, the Consortium sponsored the 6th Annual European Semantic Web Conference (www.eswc2009.org/) and held the yearly MUSING Industry Day participating with a booth and a talk at the TDWI (The DataWarehouse Institute) conference, the premier event for Business Intelligence and Data Warehousing, in Munich (15th-17th June). Furthermore, MUSING also agreed to participate in the TDWI Conference in Amsterdam

(16-17th November) the project will also host a booth at this event and MUSING representatives will be on hand to answer any questions about the MUSING project.

Progress in MUSING research

The MUSING consortium comprises partners contributing advanced research in the areas of natural language processing and information extraction, semantic technology and ontology management, as well as advanced methods in financial statistics. In this section, the key steps of progress in these research areas are reported.

The foundational backbone for the integration of knowledge and project results is formalized in ontologies which themselves serve as schema for the MUSING knowledge base repository. Research by MUSING took up the challenge by constantly improving and monitoring the conceptual model and ontological commitments of the MUSING semantic BI. This includes the release of a second version of XBRL representation in the ontology language OWL with extended possibility to map from/to XBRL compliant data to knowledge available in the MUSING ontologies.

Major progress has been also made in the reuse of information contained in the ontologies, while supporting applications and research results achieved by the other consortium partners (e.g. facilitation of information extraction). Another result is the provision of a probabilistic meta layer, allowing access to specific statistical information as distribution of values from observations and how they change over time.

A platform for the persistent storage and update of semantic data compliant to the ontology schema of MUSING is now fully implemented. All the partners of the project – and all potential users – can query and update this storage on the base of advanced semantic queries facilities. But since end-users should not be trained in the use of those complex update and query mechanisms, graphical user interface are now in place for specific pilots of the vertical streams of MUSING.

Major progress has been achieved in the field of Ontology-based Information Extraction (OBIE), in the context of two pilot applications of MUSING. Ontologies are giving the background knowledge, instances of which should be detected and annotated in different data sources, being structured and unstructured, in various languages. Data sources are for example imprint pages of companies, company profiles available at the web pages of stock exchanges, newspapers, etc. An efficient merging tool has been implemented that performs a semantic fusion of the information extracted from the different sources. Only the merged information is then submitted to the ontology storage layer for performing ontology population. On the top of the OBIE tools, a monitoring service has been implemented, which detects changes in the information. The monitoring service encodes the detected changes in the ontology with the help of the temporal representation schema developed in MUSING, and sends notice of such modifications to the end-user, which can then update her/his own data management system.

MUSING has developed and deployed a number of information extraction applications which target ontologies in 3 vertical streams. The tools are designed to extract information from text (structured, semi-structured, unstructured) in various languages (English, Italian, German) which is used to populate a knowledge repository and thus used by various business intelligence applications. MUSING has also made progress in the areas of cross-source coreference, producing state of the art systems based on semantic-based clustering. Progress has also been made in the area of opinion classification and mining, in particular supervised machine learning has been applied to business sources obtaining very competitive systems.

In particular, in the context of evaluation and estimation of the reputation of a firm there is not a wide literature since the novelty of the theme. Thus the activity of UPV has been

mainly dedicated to explore among the main statistical models more suitable, improving and modifying them in the more appropriate way. The principal innovative aspect of the problem consists in the availability of variable derived from textual source. Thereby we needed to incorporate such information into a self contained statistical model.

Another research topic addressed by MUSING in the third year of the project focused on how to extract knowledge about probabilistic relations among concepts hidden in populated ontologies. The basic intuition is that object properties and datatype properties maybe looked at, under certain conditions, as establishing either networks or set of rules, which can be transformed into either Bayesian networks or weighed inference rules by computing the frequencies of the corresponding instances.

The research activity has followed two lines: the extraction of Bayesian Networks and the extraction of weighted inference rules.

The extraction of Bayesian network is based on the idea of exploiting the differences between ISA relationships and other object type relationships to consider the TBOX of an ontology as a network, the nodes of which are ISA hierarchies of concepts and the arcs of which are object properties between hierarchies. Such structures are then compiled into a two-level Bayesian networks by computing the probabilities analyzing the instances. A Bayesian query language is then provided for different form of reasoning on the network. A typical query is “which is the likelihood of default of a company given that it is limited and it has branches outside Europe?”

The second approach looks at the TBOX as if it were a network of concepts bound by generic arcs corresponding to the object properties. The network is searched to find out the most relevant concepts and the relations among them with methods similar to the ones used for the analysis of the web.

The set of rules found this way is further specialized by performing mining on the corresponding instances. The results are rules that involve concepts, their datatype properties, the object property that binds them and a weight factor resulting from the frequencies computed in the instances.

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**, MUSING is developing and validating next generation (Basel II and beyond) semantic-based BI solutions, with reference to Credit Risk Management and Access to credit for enterprises, especially SMEs in order to face the general concern that the new international Accord on Capital Adequacy will cause banks to redirect credit away from distressed markets and from SMEs that might not have enough human resources, or the capital required to hire them, for collecting and storing the requested complex information. MUSING's solutions will be 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 is approaching the **Internationalisation** market through the development and validation of next generation 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**, the project addresses 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. There is a need for flexible operational risk management systems able to synthesise, share and present large amounts of data on operational losses, as well as to identify sources of operational risk, especially related to IT systems. Naturally, once those risks are identified and presented, a mitigation strategy for handling them is required – and that will be another service provided by MUSING. 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. The added view, recently incorporated into this VS, is the combination of IT-OPR with the financial risks facing the service provider (VNO or a similar enterprise). In doing so, the IT-OPR VS is affording the decision makers a much broader view of the potential risks associated with an existing customer or with a prospective customer. That is because one of these risk measures alone does not give the complete view of that customer, as far as the decision maker is concerned. There may be cases where the IT-OPR is small, tempting the decision maker to accept the risk – but the chances of collecting on the contract are small, making this transaction unprofitable. Of course, a symmetrical situation may exist where the IT-OPR is fairly substantial (resulting in a high charge to the customer) but there is virtually no financial risk, so again a different decision may be reached. And, finally, a whole spectrum of intermediate situations exists, making the decision maker's job difficult and uncertain. Thus, providing a single, combined, measure of risk provides a valuable aid to the decision maker in arriving at the final disposition for each customer.

Technology Outlook and Innovative Features

The technological development is taking place at various levels (refer to Figure 1 for an overview of the main MUSING technological components). 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, 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 of those documents is concerned mainly 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 level of processing.

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.

After continuous evolution of the ontologies in the former period, the foundations have reached a state 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 in an advanced state of implementation; statistical and data mining models to exploit additional semantic features that ontologies provide are now in place and tested in the MUSING pilots. Ultimately, integration of these components has been successfully implemented, following the requirements defined by the numerous pilots defined in the successive

batches. 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, which allows information extraction processes to provide intelligent novelty detection based on temporal information. This feature is used in the MUSING pilot for enterprise monitoring (hosted by MUSING partner VVC as part of an enterprise monitoring service).

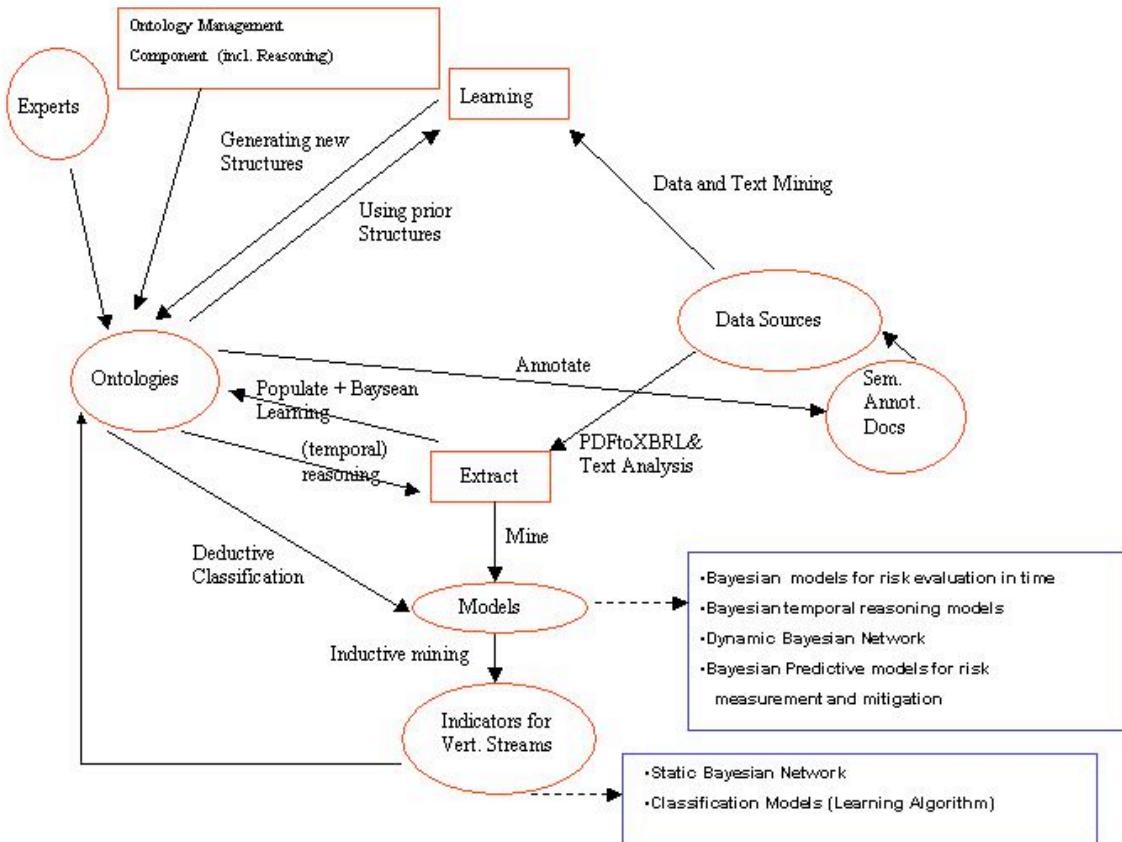


Figure 1: Technological Foundation Components

Demonstrators and Results of Field Tests

The MUSING demonstration phase started in Month 27. In order to begin field tests of the first MUSING modules, a number of technical “pilots” have been set up in the three Vertical Streams, and are already part of the prototype of the MUSING service platform, namely:

1. Financial Risk Management
 - a. *Balance Sheet Analyzers* are a set of web services conceived for XBRL balance sheet instances analysis. These tools are: Pdf 2 Xbrl, Balance Sheet Annex Visualiser, Balance sheet reclassifier and Financial Indicators Computer. **Pdf 2 Xbrl** turns Balance Sheets in PDF file into the corresponding XBRL instances (both turned into a stream for interface reasons). Ideally this tool would be parametric with respect to the XBRL

taxonomy, which might change in dependence of the national legislation underlying the corresponding XBRL taxonomy. Pdf 2Xbrl currently handles Italian and German Balance Sheets. Intrinsically, the component cannot guarantee that the XBRL instance generated is 100% correct with respect to the source document. This implies that when this component is to be deployed where absolute correctness is required, it needs to be followed by a manual revision of the instance. The **BS-Annex Visualiser** is a tool for an ergonomical visualisation of a Balance Sheet (i.e. an XBRL balance sheet instance and a PDF file containing annex text). This service displays the annex text in Html and creates anchors from the actual values in the Balance Sheet the description refers to. This is useful as most of the time the annex is read in order to find explanation or detailing of some of the Balance Sheet content. The BS-Annex Visualiser depends on the taxonomy and on the structure of the annex, although these are not presently separate from the component itself. The **Balance Sheet Reclassifier (BSR)** “reclassifies” an XBRL instance of a balance sheet. Reclassification is a different representation of a balance sheet (in our case rendered in XML format), which is usually the starting point of balance sheet analysis. As additional resources, it relies on the mapping between taxonomies. This is declaratively stated, and should ideally be hosted by the ontology in future revisions. Reclassification mappings are currently defined over the Italian jurisdiction. The reclassified taxonomy is “common sense”, but not standard. This will be improved by generating an XBRL instance conforming to a “reclassification” taxonomy yet to be agreed upon. Reclassification is not 100% definable from the sole information found on a balance sheet. Approximations and rules-of-thumb are therefore used in the mapping definition. Last, the **Financial Indicators Computer**, that performs computation of common Financial Indicators over an XBRL instance of a Balance Sheet, outputting a set of <attribute, value> pairs, where the attribute is the financial indicator (such as Return on Investment), and the value is the computed value for the indicator. This tool accesses an external definition for the financial indicators, and uses MUSING XBRL-libraries to manipulate the instance. The financial indicators are defined over a reclassified version of the balance sheet against a harmonized and simplified business reporting taxonomy frequently used the European banking sector (the BACH taxonomy).

- b. *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.
- c. *Business Plan Analyzer* (BPA) provides services for supporting both the self-assessment of Small and Medium Enterprises (SMEs) in the analysis of their Economic Plans and the decision making process of Financial Institutions and Investors in judging business proposals in order to grant a credit. It is a semantic-assisted evaluation tool that merges a classic classification technique with prior (domain) knowledge. The basic idea is to learn a classification model from past cases and then makes a prevision for new unseen cases. In this context the past cases consist of a set of

answers to a qualitative questionnaire reporting on the characteristics of the company and its economic and financial plan and a value concerning the obtained score. The prevision is about the qualitative score a company may obtain on the basis of its answers w.r.t. the past cases and the prior knowledge.

- d. *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 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.
- e. *Reputation models*: We extended the text analysis to a larger number of documents, including full lengths articles of IlSole24Ore, financial tickers from Radiocor (a service offered by IlSole), web pages of the Italian stock exchange, and merge this into one set of information, which is computing the opinion expressed on a company.

2. Internationalisation

- a. *Market Scan for India*: The pilot provides businesses (SMEs or Large Enterprises) planning to expand into markets of India with a list of regions where the investment seems to be most promising. In addition, it supplies up-to-date information on the most favourable region(s), the region's business potential, and the probability to succeed in view of the firms' given objectives of the market entry. By way of a questionnaire, a firm provides all the relevant information – the firm's characteristics and the operations it is seeking to perform. The model elaborates this information together with the regions of the country, and then suggests which region would be the best choice, as well as the characteristics of the region that were more important in determining its score, also basing on historical data about other companies. The Market Scan model accesses data about the regions fetched from various locations in various formats mostly in the Internet.
- b. *Enterprise monitoring*: 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. Enterprise Monitoring is based on semantic extraction of business information from unstructured data sources (Internet, articles etc.). It provides companies with monitoring ("International Enterprise Intelligence") service, using proprietary and publicly available data sources. It is also capable to extend the detection and monitoring beyond the case of personal changes to the cases of "change in product" and "capital changes". Company information is monitored over a period of time - updates are registered as monitoring events (triggered by a change of business information, e.g. a change in personal, capital, business rating or address.) This allows user to get up-to date information on relevant changes in status of businesses.

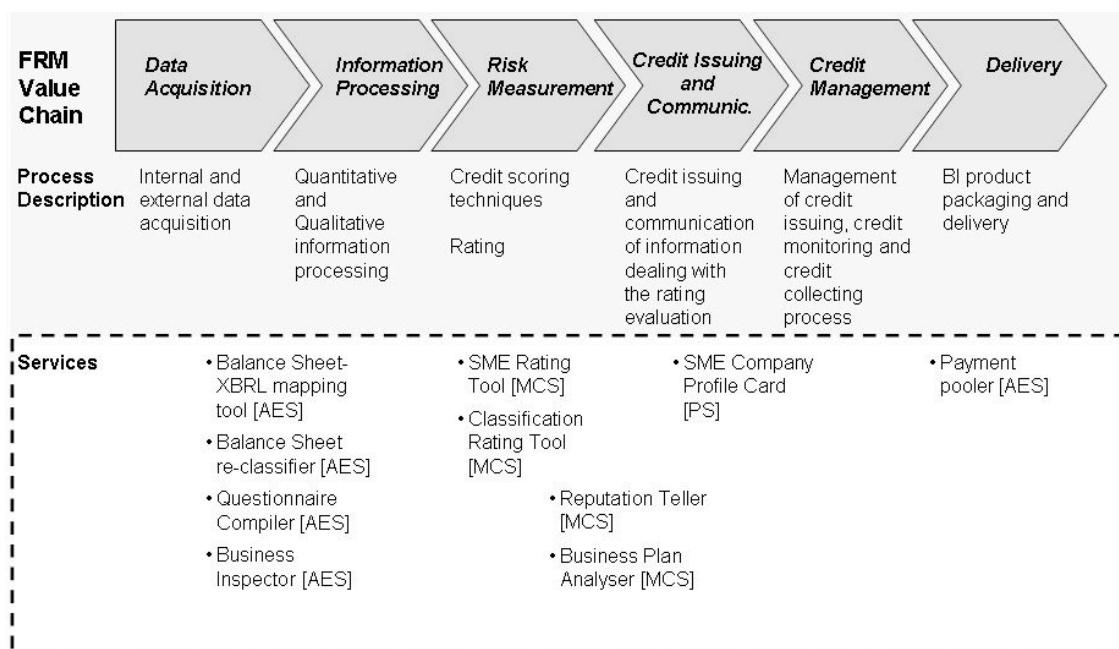
3. IT - Operational Risk

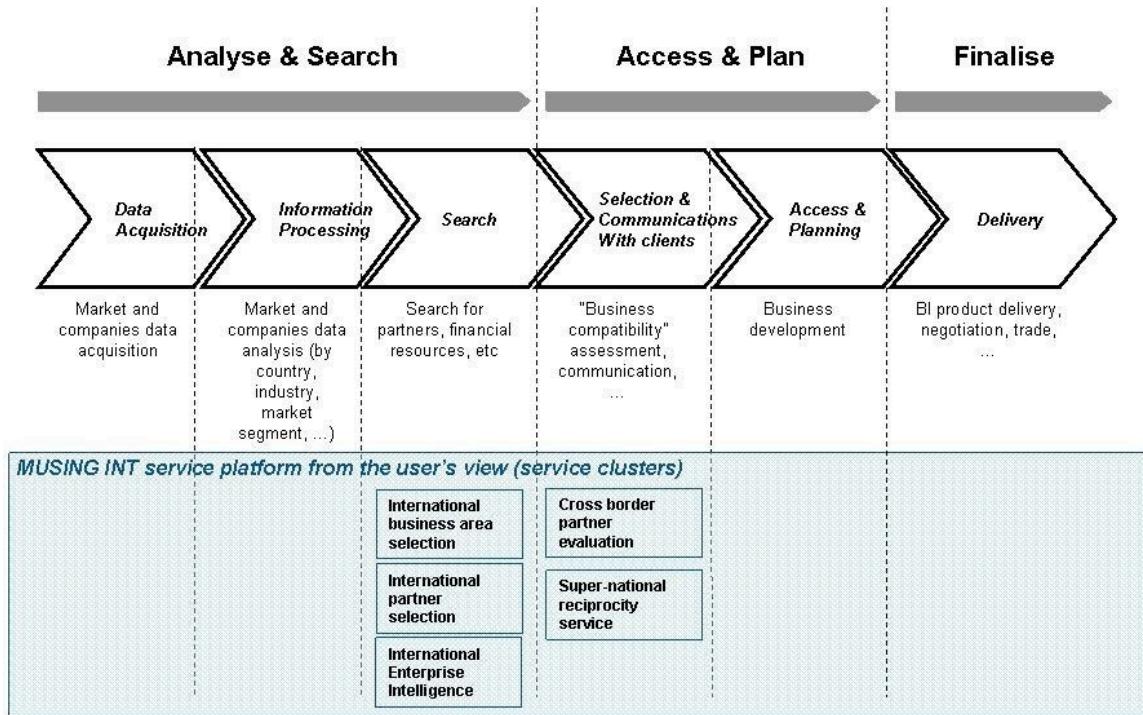
- a. *Addressing IT – OpRisk at a Telco*: the application of MUSING will be in producing risk assessments and mitigation strategies for a Telco company.

On the basis of the customers' profiles (new or old ones), this tool produces a bunch of on-line & off-line analysis to improve and enrich the quality of the provided service. This objective is accomplished through the employment of innovative statistical tools in the context of IT Operational Risk. In particular Bayesian networks are profitably built taking into account different type of data, even produced by heterogeneous sources. The problems and interruptions occurred to clients are first of all recorded on a log-file, from which useful information can be derived and introduced inside the network. Moreover, if the occurred problem is not solved by the call centre, an expert technician is devoted to a physical problem solving activity and all the information concerning the intervention are collected in an excel file. The Bayesian network can exploit the relations existing between the typology of problems occurred to the customers and the level of registered severity (seriousness of the problem). On the other hand some explorative methods are employed as well, such as association rules. They allow the analyst to have a quick and explorative look to the possible relation between problems, severity and type of clients.

- b. *Assessing The Risk of A Customer's Business to a VNO:* This pilot concentrates 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. Bayesian networks and association rules have been usefully implemented and tested. In fact the service can be interrogated by VNO which can consequently take the best decisions in terms of evaluation of a new customer, or decide on the continuation of an existing customer's contract.

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.





Operational risk assessment

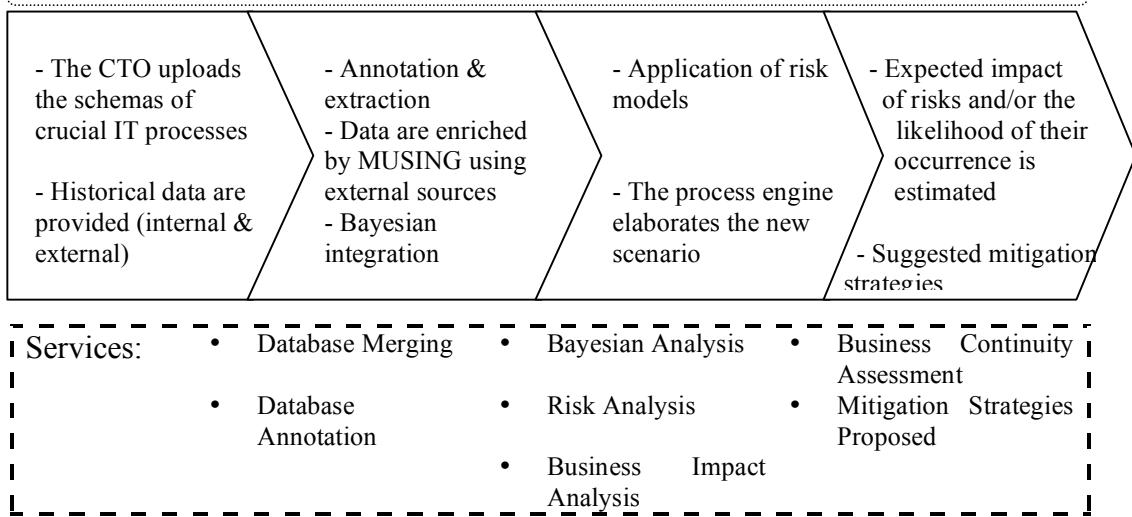


Figure 2: Value-chain representation of typical to-be MUSING BI services

User Involvement, Promotion and Awareness

During the first three-quarters of the project, a large number of scientific papers and presentations have been submitted and accepted, to transmit the large RTD opportunities arising from state-of-the-art advancements achieved by MUSING and from the application of semantic based business intelligence to the three vertical streams.

A selection of presentations and scientific publications is available at <http://www.musing.eu>, including for example: P. Giudici P. Cerchiello and E. Bonafede (UPV), *Statistical models for Business Continuity Management* (in *Journal of Operational Risk*, Vol. 2, N° 4, pag 79-96, 2007), M. Yankova (USFD), H. Saggion

(USFD), H. Cunningham (USFD), *A Framework for Identity Resolution and Merging for Multi-source Information Extraction*, presented at the Language Resources and Evaluation Conference (LREC) 2008, 28-30 May 2008, Marrakech, Morocco; T. Declerck (DFKI), H.-U. Krieger (DFKI), H. Saggion (USFD), M. Spies (UIBK), *Human Language and Semantic Web Technologies for Business Intelligence Applications*, presented at LangTech 2008, Rome (IT), 28-29 February 2008; V. Grossi (UPI), A. Romei(UPI), S. Ruggieri (UPI), *A Case Study in Sequential Pattern Mining for IT-Operational Risk*, presented at the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD), Antwerp, Belgium, 15-19 September 2008, *Extending Ontology Queries with Bayesian Network Reasoning* presented at the IEEE 13th International Conference on Intelligent Engineering Systems 2009, Barbados, 17 April 2009, *Quality, Risk, and the Taleb Quadrants* presented at the 2009 Quality and Productivity Research Conference, IBM T. J. Watson Research Ctr., Yorktown Heights, NY, 3-5 June 2009, A threshold based approach to merge data in financial risk management to be published in Journal of Applied Statistics. With the publications of these papers MUSING has managed to achieve a widespread audience allowing the project to show what it has to offer from each aspect and area.

The dissemination of scientific results, especially in the form of peer-reviewed papers published or presented at significant national or international events, will continue to be a fundamental aspect of the overall dissemination work. Whenever possible, scientific papers and other selected publications are made available for download in a specific section of the main communications website of the project on <http://www.musing.eu>.

MUSING has been steadily present with high-quality papers and presentations at the most important international events in the relevant fields, including for example LangTech2008, Rome (IT), February 2008; LREC 2008, Language Resources and Evaluation Conference , Marrakech (MA), May 2008; ECAI 2008, European Conference on Artificial Intelligence, Patras (GR), July 2008; ISBIS-2008 - International Symposium on Business and Industrial Statistics with special emphasis on Quantitative Analytics for Banking, Finance and Insurance, Prague (CZ), July 2008; MMIES-2, Multi-source, Multilingual Information Extraction and Summarization Workshop, Manchester (UK), August 2008; ENBIS8, Annual Conference of the Network for Business and Industrial Statistic Conference, Athens (GR), September 2008; HIS 2008, 8th International Conference on Hybrid Intelligent Systems, Barcelona (ES), September 2008; SEPLN 2008, 24th Conference of the Spanish Society for Natural Language Processing, Leganès (ES), September 2008; ECML PKDD 2008, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Antwerp (BE), September 2008; ISWC 2008, International Semantic Web Conference, Karlsruhe (DE), October 2008. Events that hosted presentations by MUSING later in 2008 include EIBA 2008, 34th European International Business Academy Annual Conference, Tallin (EE), in December 2008.

In 2009, MUSING was presented at the World Bank Intellectual Capital conference, the conference of the Italian Statistical Society, also presented at the IEEE 13th International Conference on Intelligent Engineering Systems 2009, Barbados, 17 April 2009, World Conference on Intellectual Capital for Communities, Fifth Edition, Paris (FR), 28 May 2009, the Annual Conference of the Network for Business and Industrial Statistics, ENBIS2009, Goteborg, Sweden, 21-23 September 2009, and the W3C / XBRL International "Workshop on Improving Access to Financial Data on the Web", Washington DC, 5-6 October 2009, where MUSING member UIBK as well acted as program committee member. Overall MUSING was able to continue its global presence through these presentations and more. MUSING intends to until the end of the project ensure that it remains in the public eye and is accessible for all interested parties be it through events, publications or presentations.

Furthermore, MUSING has devoted substantial effort to the organisation of events such as workshops or conferences to enhance the awareness of the project in the relevant professional circles. The events organised in 2009 include, Seminar "Business Intelligence for Knowledge Management and International Business" Helsinki (FI), 27 Jan 2009, also as previously mentioned MUSING hosted booths at the TDWI, through the annual MUSING industry day and also in Amsterdam at the TDWI 10th European Conference.

The main communications website functions as the point of convergence of the initiatives carried out at the national level, providing the public and the EC with complete and up-to-date information on the various dimensions of the project; it offers services to support the sharing of information (e.g. news, document downloads, feedback collection, surveys) and the organisation and running of events. The website will continue to play a major role for the dissemination of MUSING, and will be improved throughout the project's lifecycle to reflect new research developments and emerging marketing needs. Furthermore, measures are continuously taken by both the website administrators and the other MUSING partners to promote the website and encourage visits.

Training activities are mainly addressed to two distinct targets: the Researchers (with training focussing on illustrating the applicable, technological achievements to date on RTD and concentrate on deploying the results and issues surrounding the technical implementations, standards, interoperability and regulatory issues) and the Users (for whom the training focuses more on demonstration based cases illustrating a more product-oriented approach). Training topics covered particularly the areas of document and text annotation, risk management, and internationalization services, targeting financial institutions/managers (1 course), Scholars and Academia (1 courses), SME's managers (2 courses) and Business consultants and CPAs (2 courses).

Moreover, since June 2008, MUSING is a member of XBRL International with the aim of contributing to the effectiveness of this standard, whose potential to transform internal and external business reporting is considerable, and to point out what future research must be done to develop analytical applications with a high degree of *intelligence* and very low reaction time based on XBRL and Business Intelligence.

Future Work or Exploitation Prospects

Pilot applications results represent the cases for developing a methodology of usage for the approach and tools that are objectives of the MUSING project – that may involve the adoption of knowledge engineering methods - and demonstrating how existing methodologies can be utilized in the BI domain and new settings ("hosted pilot applications").

Pursuing its multi-industry potential, MUSING is seeking for collaborations with information and content providers to consolidate and further expand its application domain. In its final phase, MUSING deploys pilots to key end users in large enterprises in all vertical stream areas (Banca Monte dei Paschi di Siena, Il Sole 24 ore, Verband der Vereine Creditreform, TBSI) as well as SMEs (...)

MUSING content analytics can be provided in terms of content mining applications (e.g. information extraction, text analytics or social network identification tools) expanded beyond what foundational information access applications provide. The hosted pilots may also entail "serialization" of those analyses to create a succession of processes that converge to produce a very specific answer to a specific question. Exploitation within the context of the hosted pilot campaign is intended to:

1. Identify content sources that contain the relevant information needed to perform multifaceted analyses and that can be mined to answer critical business questions;
2. Produce prototype analyses that will give results with identifiable business benefits, as well as best practice toolkit for taxonomies and content analytics

(how content analytics helps to bridge the diverse entities of the content continuum, enterprise information management and metadata).

Further exploitation of project results is planned for the time after the end of the project. Funding for developing industrial prototypes as well as pre-products building on the MUSING technologies is planned. Following the financial crisis, MUSING technologies enable enterprises to better comply with ever more complex regulations as well as to detect regularities yielding exploitable business informations themselves.

Further Information

For further information, please visit <http://www.musing.eu> or contact our Coordination Office at info@musing.eu.