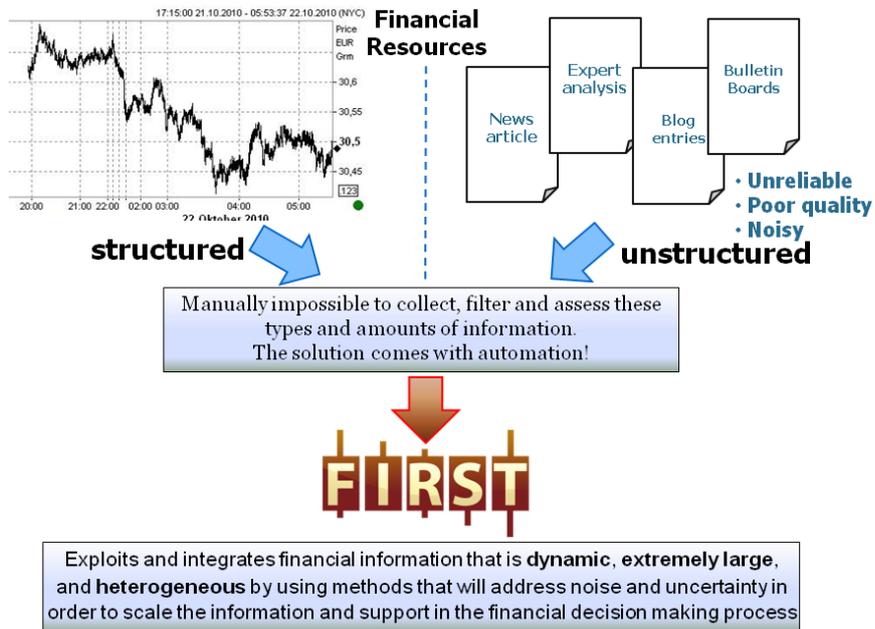# FIRST

Large scale information extraction and integration infrastructure for supporting financial decision making.

How to focus on the relevant while making financial decisions in a world of information overkill?

Ask FIRST first

http://project-first.eu/

FIRST aims to provide an information extraction, information integration and decision support infrastructure for information management in the financial domain. This area is extraordinarily faced with the challenges of extremely large, dynamic and heterogeneous sources of information. The daily work and the business success of all decision makers in this industry depend on the availability of highly trustable, easily acquirable information.

Within the international financial industry not only individual market participants are negatively affected by information overflow and information asymmetries. Financial flows and financial stability within entire industries of the global economy are affected as well. This may have also severe consequences as has been observed in the current financial and economic crisis. Information is amongst the most valuable assets in the financial industry and reducing information asymmetries and increasing transparency by providing a fast, real-time, automatic and more comprehensive information base can help preventing false decisions.

Important areas of work

By extracting and analyzing unstructured financial information from the Web, FIRST is basically trying to provide 'yet another view' on the situation in the financial markets that could be tailored for different users' needs. The figure below shows the main steps that FIRST follows to fulfill this view:

During the first year, the project delivered preliminary but already working prototypes of:
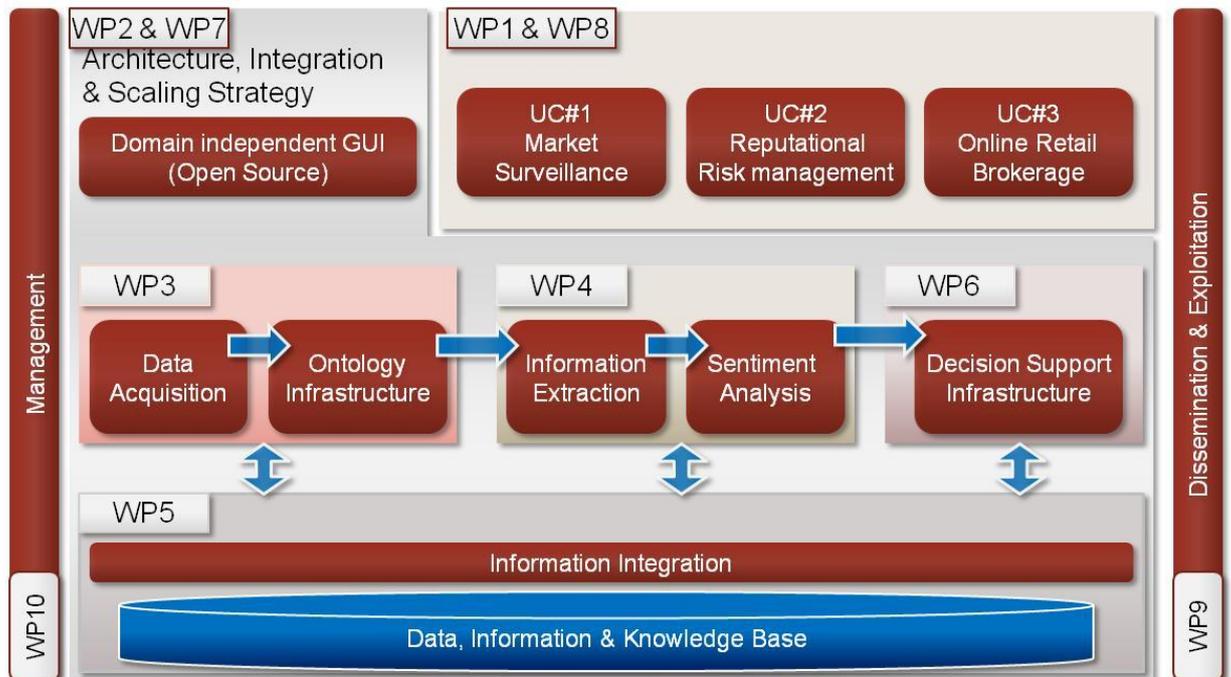
- **Data Acquisition Infrastructure** for acquiring and accessing massive amounts of heterogeneous historical information and live feeds of unstructured or semi-structured information.

- **Ontology Infrastructure** with tools for manually and semi-automatically capturing and maintaining the domain knowledge models.

- **Semantic Information Extraction System** for extracting information and financial sentiments from texts with respect to specific objects of interest.

- **Information Integration Infrastructure** for integrating and consolidating information from heterogeneous (unstructured, semi-structured and structured) information sources.

- **Decision Support Infrastructure** for integrating information from the knowledge base into financial event detection models, visualization models, decision models and for scalable execution of these models.

One of the FIRST main results to be developed within the second year of the project is an **Integrated Financial Market Information System** based on a pluggable open architecture framework for non-ICT skilled end-users for on-demand information access and highly scalable execution of financial market analyses. Furthermore, FIRST aims at demonstrating domain-independent as possible applicability and Graphical User Interface (GUI).

The FIRST project is validated through three compelling use-cases. These use-cases have been defined in detail during the first year and show a huge potential to evaluate the project's results:

1. **Market surveillance**: Financial markets are often influenced by rumours. Some of them are true and some others false. Sometimes false rumours are even intentionally placed in order to distract and mislead other market participants. In both cases FIRST assists by providing a plausibility checking methodology that identifies hoaxes and assesses the consistency of new information with other reliable sources.

2. **Reputational risk management**. Due to the increasing relevance of reputation in Financial Services, combined with higher-frequency dynamics of the market, reputational risks is a major issue for many financial institutions. In particular, there is no off-the-shelf product that takes into account unstructured information to assess reputational risks. Therefore risk managers have to deal with large and unstructured amounts of market and company information. They have to manually collect, filter, assess and, most importantly, interpret the heterogeneous information about third parties' reputation that might have impact on their own reputation. By using FIRST, this process will be automated and objectivised and hence will be faster, more reliable and less time-consuming.

3. **Online retail brokerage**: The focus lies on private investors who, unlike the investment manager, do not have the ability or the time to collect and analyze new information about financial markets in real time.

The project is structured as depicted in the following figure:

Summary of activities

FIRST started in October 2010 and runs for three years. The first six months of the project were characterized by the overall organization of the project and the setting of the objectives, analysis of technology and reusable components coming from external projects or partners' developments, collection of users' and technical requirements and specifications of the use cases. The second half of the year was mainly devoted to having a set of preliminary working prototypes focused mostly (but not only) on the areas of data acquisition and sentiment analysis.

During this first year project partners also work on depicting use-case scenarios and evaluation metrics, and decided on the overall architecture, scaling strategy and storage facilities. A significant effort has been devoted also to the preparation of a golden standard for sentiment annotation of web documents, which will be useful for future evaluation and is an outstanding contribution to the scientific community in itself. A first version of the dissemination and exploitation plans has been also provided.

It is worth noticing that from the very beginning of the project partners have started a plan towards having early prototypes more advanced in some cases than foreseen contractually.

In summary, the main activities carried out during 2011 are the following:

- Extraction of relevant **user requirements** from domain experts.
- Extraction of relevant **technical requirements** and survey of the state-of-the-art on related technologies.
- Setting up **computing infrastructure**, tools, and previous work.
- Developing the **digital corpus** of manually annotated documents, as a golden standard on **sentiment analysis**.
- Defining the scope and developing the first **early software prototypes**.
- Defining the **architecture, integration and scaling** strategies.
- Annotated document **corpus for sentiment analysis** (gold standard)
- Initial set up of the **dissemination and exploitation strategies**.
- Submission and presentation of **academic papers** to several conferences
- FIRST Web site on-line (http://project-first.eu).
- Established and put to use various **social media channels**

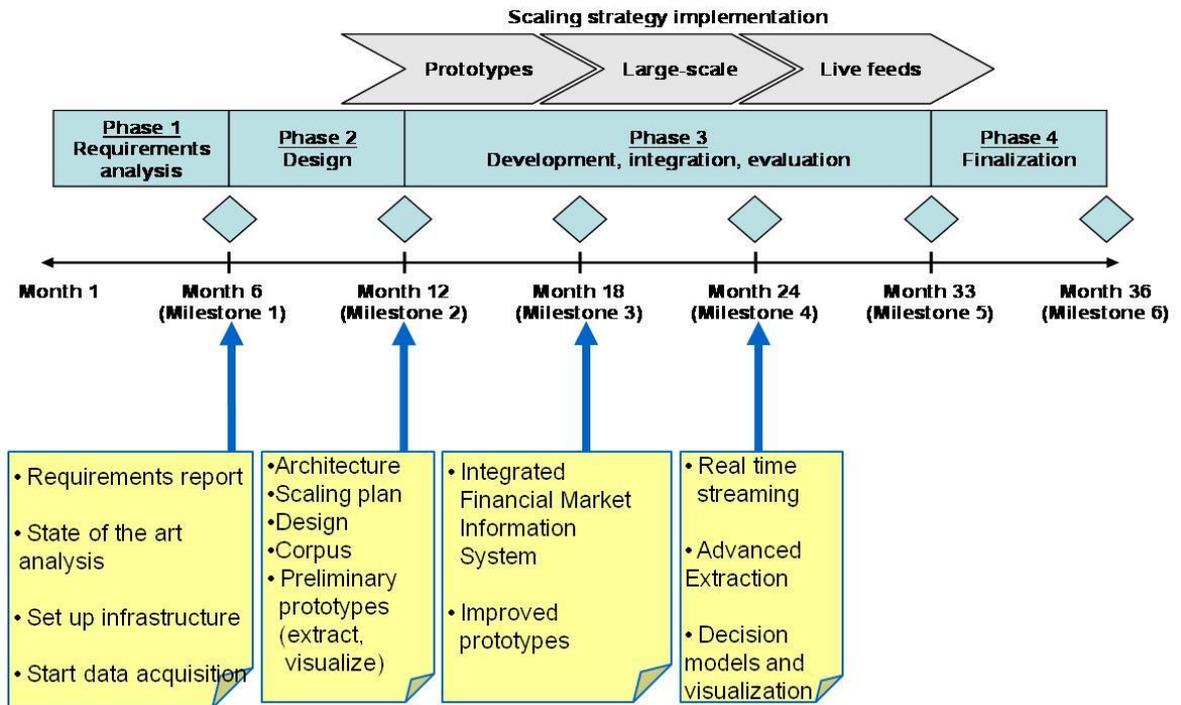| Channel	Link	Objective
	http://first-fp7.eu/blog	Post about interesting results of the project (1 page articles)
	http://twitter.com/projectfirst	Short project updates
	http://www.facebook.com/projectfirstfp7	Informal project updates and links to the RSS website
	http://www.youtube.com/user/projectfirstfp7	Videos produced by the project partners
	http://www.linkedin.com/groups/FIRST-Project-3998776	Project updates directed at professionals in the field

Future work

In the second year, FIRST will continue to work on the development of the prototypes, improving their functionalities and mainly focusing on the following issues:

- Use cases implementation and validation
- Maximise pipeline throughput
- Develop glass-box decision support models and visualizations
- Focus on scalability of the solution
- Define the details of the open source and dual-licensing strategies
- Initiate the industrial uptake (1st FIRST workshop foreseen for March 2012)

We also aim to pay special attention to the dissemination and exploitation aspects, by organizing two industrial workshops which main goal will be the introduction of industrial partners inside and outside the consortium to the developed solutions and their functionalities. To this extent, we are discussing in several alternatives to show the results achieved so far in FIRST project.

The expected work during the project is structured in several milestones as depicted in the figure below. We have already reached milestones 1 and 2. 2012 will see new releases of software prototypes, more robust and scalable.

Further information

PARTNERS

FIRST is co-funded by the European Union with a duration of 36 months and an overall budget of 4,574 million Euros. The FIRST Consortium consist of **9 partners** drawn across the European Union and covering the entire value chain of large-scale information management in the financial domain, with a broad spectrum of companies, financial institutions, and academia focusing in assessing the benefits of the productivity, competitiveness, and profitability of the FIRST solution.

- Atos Spain (Project coordinator – Spain)
- University of Hohenheim (Germany)
- Jozef Stefan Institute (Slovenia)
- Goethe University Frankfurt & E-Finance lab (Germany)
- B-Next Engineering GmbH (Germany)
- Banca Monte dei Paschi di Siena SpA (Italy)
- Interactive Data Managed Solutions AG (Germany)
- Boerse Stuttgart Holding GmbH (Germany)
- Georg-August-University Göttingen (Germany)

CONTACTS

Project Coordinator	Technical Coordinator
Tomás Pariente Lobo	Achim Klein
Atos Spain	University of Hohenheim
tomas.parientelobo@atosresearch.eu	achim.klein@uni-hohenheim.de