*Research project*

# MasterICT

***Social Web, Semantic Technologies and Semantic Web to enhace Collaborative Innovation and Technology Watch Processes***


*PhD Student*

## Alain Perez

*Supervisor*

## Dr. Felix Larrinaga

*Co-Supervisor*

## Dr. Juan Ignacio Igartua


mu_grande.pdf

MONDRAGON
UNIBERTSITATEA

**Mondragon Goi Eskola Politeknikoa**

**December 2013**

# Contents

# List of Figures

# Acronyms

**AI** *Artificial Inteligence*

**API** *Application Programming Interface*

**CI** *Competitive Intelligence*

**CMS** *Content Management System*

**FOAF** *Friend of a Friend*

**GRDDL** *Gleaning Resource Descriptions from Dialects of Languages*

**HTML** *HyperText Markup Language*

**HTTP** *Hypertext Transfer Protocol*

**ICT** *Information and Communications Technology*

**IMS** *Idea Management System*

**LD** *Linked Data*

**LOD** *Linked Open Data*

**NLP** *Natural Language Processing*

**OSS** *Open Source Software*

**OWL** *Web Ontology Language*

**POWDER** *Protocol for Web Description Resources*

**RDF** *Resource Description Framework*

**RDFa** *Resource Description Framework in Attributes*

**RDFS** *Resource Description Framework Schema*

**RSS** *Really Simple Syndication*

**RSS** *RDF Site Summary*

**SKOS** *Simple Knowledge Organization System*

**SPARQL** *SPARQL Protocol and RDF Query Language*

**SQL** *Structured Query Language*

**SVM** *Support Vector Machine*

**SW** *Semantic Web*

**TM** *Text Mining*

**TW** *Technology Watch*

**URI** *Uniform Resource Identifier*

**W3C** *World Wide Web Consortium*

**XBRL** *eXtensible Business Reporting Language*

**XHTML** *eXtensible HyperText Markup Language*

**XML** *eXtensible Markup Language*

**XSLT** *eXtensible Stylesheet Language Transformations*

# 1. Introduction

There is no doubt that in this new era of emerging economies innovation is the valid alternative for achieving competitiveness, the only chance to survive. This is precisely one of the bases of the Lisbon Strategy[?], "innovation as a driving force for change in business". Today, organizations that remain are those that adapt to change.

Moreover, researchers have worked on guidelines that describe what are the key factors to consider in the implementation of innovation processes. Recent investigations indicate that collaborative processes mixing both internal and external agents[?][?][?][?][?] are key. Innovation needs to be based on experimentation and collaboration between organizations, universities, public sector, and of course, users[?].

The innovation process and the technology watch process are closely related. In any Innovation process, both technology watch and competitive intelligence processes are key instruments, and require the creation of a collaborative environments. The amount of ICT tools and platforms to support those processes is extent but each of them focus on specific data or areas of the process. There is no data interoperability among the tools and platforms to support information flow among processes and their stages. It is urgent to create the conditions to foster an environment capable of capturing and sharing collaborative information, in order to capitalize information and knowledge, from both inside and outside the company. This knowledge will be used later when making strategic decisions, which could guide the organization in the right direction. ICT provide tools and techniques to model those data flows and to enhance interoperability among heterogeneous systems.

The challenge is how to promote and facilitate such participation and collaboration, how to provide platforms and systems that share information efficiently (interoperate) and determine which *Information and Communications Technologies* (ICT) can help in the context of Technology Watch and Innovation. Conducted research on the field identify Social Web technologies as tools to boost collaboration and participation. On the other hand different studies point to the Semantic Web as the techniques to represent or model processes and their data flows and consequently provide the frameworks to enhance interoperability. Semantic technologies complement Social Web and Semantic web by providing techniques to classify, filter, discover, or associate data, information or knowledge. Thus, the focus of the research is aimed at Social Web, Semantic technologies and Semantic Web.

This document is organized as follows.

Chapter 2, *Foundations*, presents the state of the art in the fields and technologies addressed in the project. It begins by describing the basics of Innovation and Technology Watch. It follows with a brief analysis of the ICT proposed by the project classify in three groups; the Social Web, Semantic Technologies and the Semantic Web. It continues by describing the application of those technologies in both processes nowadays. Finally, it presents the conclusions and *opportunities* extracted from the state of the art.

Chapter 3 presents the objectives of the thesis and the hypothesis for it's development.

Chapter 4, shows the methodology followed for the thesis, the metrics to test the results, the work schedule for the four years of the project and the tools used for it's implementation.

Chapter 5 working done and the results so far. It describes the already developed work, the collaboration with universities and companies, the scientific publications published, projects involved during the research for founding and findings gathered in developed case studies.

Chapter 6 gathers all the conclusions from the research project.

# 2. Foundations

This chapter focuses on the study of Innovation and Technology Watch processes. The general concepts related to them are outlined first. The first section presents an historical approach to the different perspectives of the innovation concept. The existing types of innovations, the stages that compose the innovation process and the issues to be solved are also presented. The next section focuses on the Technology Watch concept, the stages that compose it and the issues to be addressed.

The chapter continues outlining the technologies selected to address the identified issues of both processes. This section is divided in three parts. First, the Social Web and its benefits for collaborative processes is presented. Next, the Semantic Technologies capable of supporting task automation are described. Finally, the Semantic Web and its ability of interlinking different systems is outlined.

The following section presents the state of the art ant on the application of the technologies outlined in the previous section on the Innovation and TW processes. The first two subsections gather the application of Social Web and Semantic Web in the Innovation process. Next, Semantic Technologies and Semantic Web applications for Technology Watch processes are described.

Finally, taking into account all the previous topics, the conclusions of the study are outlined.

## 2.1 Innovation

In the twentieth century (1934), Schumpeter [?] made one of the first definitions on innovation. From its traditional definition, innovation encompass the following five cases:

1. Market introduction of a new good.
2. A new method of production.
3. The opening of a new market in a country.
4. The conquest of a new source of supply of semi-finished products or raw materials.
5. The implementation of a new structure in a market.

Half a century later, Padmore, Schuetze and Gibson [?] summarized Schumpter's definition by saying that innovation is any change in inputs, methods, or outputs that manages to improve the trading position of a company and it's new to the its current market.

Gee [?] and Pavón and Goodman [?], incorporate the concept of *process* to these definitions, where from an idea, invention, or recognition of a need, a useful product, technique or service is developed in order to make it commercially accepted. They even consider innovation the improvement of a product to meet market needs. In the same line, and with reference to technological innovation, Cantisani [?] defines innovation as "the sequence of activities to generate new techniques with the help of science and its method ". Amabile [?] incorporates the nuances of creative ideas as a source of organizational innovation, while Porter [?] identifies innovation as "a new way of doing things, which is marketed".

Galanakis[?] defines these issues as the use of "new or existing scientific or technological knowledge... to generate ideas that give rise to innovation (something new) ." Innovation is understood as an idea, process, system, method, service, product, policy, etc. characterized as new or improved and commercially accepted. There is a change from a linear view, where activities take place in a sequential manner, to a new one, where activities overlap and have multiple feedback loops.

Both concepts, *innovation* and *innovation process*, have incorporated these latter aspects, the partition of different actors throughout the process and decision-making aspects, in order to reduce development time of the innovation process, and key aspects such as the incorporation of new technologies and train network to accelerate the process of knowledge capture and transfer of learning in a collaborative environment with other agents that ensures mutual benefit [?].

The first reference to the models of innovation is the one that today is known as the linear model of innovation. People like Godin [?] studied this concept and made a review of its origin and historical evolution. Although used, criticized and improved by several authors, this model has rarely been cited as an original source. Some authors place that source in Bush [?], but others disagree [?]. The mentioned linear model only provides a new product market perspective, but it is not the only one. Although it is still in use, other models have appeared throughout the twentieth century. Rothwell [?] mentioned four generations of models on the evolution of innovation and predicted a fifth where networking becomes important. Below those generations of models on the evolution of innovation are listed:

1. Technology push.

2. Demand pull.

3. Interactive Model.

4. Integrated Model.


Figure 2.1: Innovation process baseline stages.

Hobday [?] took Rothwell's approach and presented the five generations of innovation models also indicating their advantages and weaknesses. More recently, Cantisani [?] analysed the different generations, focusing on the first three, and collecting contributions of various authors like Bush [?] or Stokes[?].

After 30 years of research in relation to innovation models (1977-2006), Errasti, Oyarbide & Zabaleta [?] conclude that most of them agree on a common baseline, and that the main difference is the adaptation of each to a particular case. That common baseline process stages can be found on figure 2.1 and each stage is described below:

1. *Idea generation:* this stage gathers new ideas. This stage can also gather comments or ratings of involved agents.

2. *Idea analysis:* this stage is the first filter where the ideas are analysed by experts and some of them are set aside.

3. *Idea enrichment:* ideas that pass the previous stage are enriched by experts so they can make a deeper study and add valuable information into them.

4. *Idea selection:* after enriching the idea, experts analyse them again and rate the ideas according to some criteria. This way, the second filter is performed and only feasible selected ideas are taken into account, becoming projects.

5. *Idea development:* idea developing planning is approached in this stage. Studies are conducted on many issues; market, technology, business plans, risks, possible collaborations, competitors, prototypes...

6. *Idea implementation:* the last stage is concerned with implementing the idea and bringing it to the market.

The economic challenges that arise require new models and concepts around innovation. Among the new models, *Open Innovation* concept is more emphasized, coined by Chesbrough in 2003 [?] and later studied by Christensen, Olesen, & Kjær [?]; Chesbrough & Crowther [?]; Almirall [?]; Dodgson, Gann, & Salter [?]; Enkel & Gassmann [?]; Fredberg, Elmquist, & Ollila [?] and the European Commission[?]. The Basque Government has also internalized this new concept of open innovation. This is reflected in the *Science and Technology Plan* underway since 2007 [?], [?], [?].

Open innovation is based on the following principles [?]:

- Not all the smart people work for us - we need to leverage external knowledge.
- The R&D outsourcing can generate significant value for us.
- The research does not have to originate from our own work to make it profitable for us.
- A robust business model is more important than being first to market.
- Both internal and external ideas are essential to winning.
- We can capitalize on our intellectual property and we should buy the other when we need it.

There are several ways to practice open innovation. Enkel and Gassmann [?] suggest some examples:

- Integration of customers and suppliers.
- Through listening, as innovation clusters.
- Applying innovation across industries.
- Buying intellectual property.
- Investing in the creation of global knowledge.

The first open innovation models have been studied in the Open Source Software (OSS) development industry and later were transferred to more general practices of open innovation. West and Gallager [?] identify three main threats (motivation, integration and innovation exploitation) and define four generic strategies open innovation:

- Combined R&D - shared R&D (requires a change in culture).
- Spinouts - an escape from the bureaucracies of big business.
- Sale of accessories - accept the commercialization or development of differentiated products based on commodities.
- Accessories donated - general purpose technologies that are sold so that users can develop differentiated products (eg, user folders).

Open innovation is assuming several changes, one is the ability to be able to collaborate with many people. Surowiecki [?] called it *The wisdom of crowds*, assuming that the collective intelligence exceeds that of a few people, both in terms of ideas and knowledge. The use of the power of crowds to increase the capacity for innovation is closely linked to community-based innovation. In addition, there is a general assumption that even though open innovation increases the potential creativity in the innovation process, also increases the complexity involved in managing the process.

Another tendency linked to Open Innovation is Collaborative Innovation. It involves the participation of multiple actors in the ecosystem of organization, ranging from employees to the competitor. Different authors see collaborative innovation from two perspectives:

1. *Firm-centric innovation:* Innovation within the organization.

2. *Network-centric innovation:* Innovation included in an extended organization concept.

Firm-centric innovation focuses on both internal resources as a source of acceleration of innovation processes, while the network-centric innovation, the process extends beyond the boundaries of the organization. In reference to these different levels of openness, there are three sub-models:

1. *Ecosystem innovation.* This concept addresses the classical extension of the value chain of an organization. Usually, some stakeholders have more knowledge in certain areas of the value chain than others. Within this process of opening innovation, both the organization and the different stakeholders benefit from sharing knowledge. Google is one of the best examples of this innovation model[?].

2. *User innovation.* It is a concept introduced and developed by Prof. Eric Von Hippel [?], [?]. User innovation concerns the innovations achieved by end users and producers. Perhaps the most documented case study is the Lego [?]. Lego engineers had been working for seven years in the development of Lego Mindstorms robotic game and only three weeks after its release there were thousands of hackers working on new developments of robots. From there, Lego knew organize various initiatives whose main base line has been the optimization of the product. Currently, there are more than 20,000 Lego fans who organized an online community of innovation.

3. *Crowdsourcing.* This term was first introduced by Jeff Howe [?] and is defined as the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call'. Procter & Gamble and the *Connect and Develop* platform is one of the best examples of crowdsourcing[?].

When talking about innovation processes, it is associated with the research and organization development (R&D) and numerous studies have focused on analysing different aspects of this activity in the organization, the R&D effort, how to organize this activity and the innovation results achieved [?], [?], [?], [?], [?], [?], [?], [?], [?].

In reference to the operational part of the innovation process, innovation requires a flow of ideas, obtained through formal and informal processes [?]. This process is more effective in organizations that combine these two features: first, the control directs initiatives at the different layers of the organization and second, the great commitment of the participants of the organization with respect to this process [?]. In addition, staff of organizations tend to be highly trained in innovation and knowledge commonly shared with scientists from different disciplines and departments, ensuring a direct connection between business opportunities and organizational capacity and production.

However, there are many problems when it comes to face the innovation process. One of the biggest problem is the *interoperability*. Many systems contain relevant information for innovation, but it is not easy to reuse it from other systems. Most of the times users have to jump from a system to another and lose time in order to get the needed information. The company could for example use the information gathered in the technology watch process (see section 2.2), offering that data to the innovation process users helping them generating richer ideas. But the interoperability problem does not affect only to different information systems, there are problems to link data from ideas in the process itself. Data should be interoperable between different stages of the process too, relating for example ideas with other similar ones. As it will be shown in section 2.3.3, ICT can be used to create links between the data and generating an interoperable data space where related information can be found and exploited. Igartua, J.I.[?] confirmed on 2010 the relationship between the use of Innovation Management Tools and innovation activity, showing the need of ICT tools in the process. Also in 2010, Westerski et al.[?] wrote about innovation, interoperability and linking. They described how ontologies can be used for fully describe the domain in pair with other existing ontologies. That way, they claimed to extract idea metrics from the links and discover new related ideas.

Another problem can be the large amount of data. If a company has a large idea flow, they could spend too much time managing. Using semantic technologies (see section 2.3.2) to automate this work could reduce this non-productive time.

In conclusion, the most critical phase of the innovation process is *the management of ideas*. Once the idea is selected, the next stage is to manage the project. Thus the main difficulties are in the early stages of the innovation process: those ranging from the creation of the idea, its analysis, the enrichment of this and finally the selection.

As previously mentioned, R&D is closely associated with Innovation process. Therefore, Technology Watch process and its implication within the innovation has been studied. Next section (2.2) presents the state of the art for that process.

## 2.2 Technology Watch

As reflected in the norm UNE 166006:2006 Ex Management of R+D+i (Technology Watch System), Technology Watch is an *organized*, *selective* and *permanent process*, to capture information from outside and inside the organization about science and technology. All of this in order to select, analyse, disseminate and communicate the information. To turn it into knowledge, making decisions with less risk and anticipating change. This way, technology watch represents a key tool in the R+D+i process.

The information that must be watched comes from different domains and formats:

- Patents, utility models, industrial designs (national, European or global). Often the time of the presentation is important, other times, the expiring time.
- Legislation and Regulations that may affect the activity of the company, its customers or suppliers.
- Socio-economic situation in the home or target countries of the company.
- Scientific and technical news on specialist journals, symposia, conferences and other scientific events.
- Doctoral theses and scientific and technical publications from universities, research centres and agencies.
- Sector news (without neglecting other sectors that can have positive or negative interference with the business of the company).
- Information on grants and subsidies.
- Products, prices, quality and sale conditions of competitors.
- Trade Shows: emerging industries, new competitors, distribution strategies, new products, etc.
- Direct personal contacts with competitors, suppliers, research centres, universities, etc.

The whole process of well analysed information capturing that becomes into knowledge for the company and its use within the organization, is a practice known as *Competitive Intelligence* (CI). CI analyses the factors influencing the competitiveness of the company in order to generate competitive strategies and to act successfully in the generation of innovation in the global environment of Business Intelligence.

There is almost universal agreement in relation to Technology watch and CI structure. In its simplest form, it is a process of adding value to information, analysing and producing knowledge in an intelligence way [?]. Society for Competitive Intelligence Professionals (SCIP), the most representative association worldwide in CI field, identifies five steps in the process. Those stages can be seen on figure 2.2 and are described below:

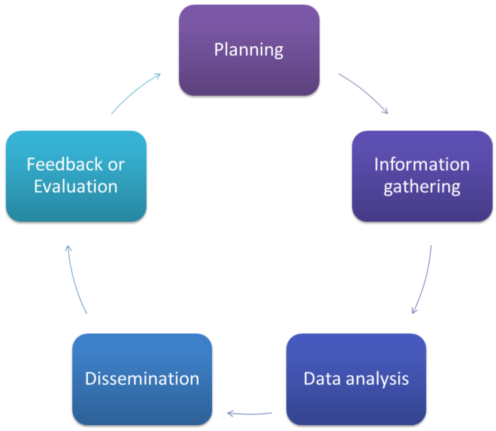


Figure 2.2: Five stages of the Technology Watch process.

- Planning: work with decision making agents to discover and define the TW requirements.
- Information gathering: includes source identification; keyword or taxonomy definition, search, data access and data extraction.
- Data analysis: data interpretation and compilation of recommended actions.
- Dissemination: deliver the findings to decision making responsible.
- Feedback or evaluation process: taken into account the response of decision makers and gather the new requirements to continue with the process.

The term "watch" was first applied to technology and was part of the management models of innovation and technology [?], [?]. TW was understood as a function for analysing the innovative behaviour of direct and indirect competitors, exploring all sources of information (books, grey literature, patent offices, etc..), examining the products on the market (built technology analysis) and attending trade shows, conferences to position over competitors and gather knowledge and technologies that will dominate the future [?].

Rouach [?], leaving aside the technological field, writes about the function of watching, in general, and describes it as the art of discovering, collecting, processing, storing information and relevant signals (weak and strong) that will guide the future and protect us from competitors' present and future attacks. Technology watch is all about external information transfer to the inside of the company, trying to gather relevant information and send it to the right people in the right time.

Palop and Vicente [?] clarify that companies must make a systematic and organized effort to enable them to observe, capture, analyse and disseminate information from the economic, technological, social or business environment in order to make appropriate decisions with minimal risk.

In 2006, Durán et al. [?] agree with the process for TW, and claim its benefits. They identified the importance of the classification of the information and proposed to use list of controlled terms, classified and grouped according to different points of view. In 2009, Fernandez et al. [?] also proposed to undertake technology watch as a *process*, indicated that it is the starting point of the innovation process and identified the need of establishing key words in order to tag the information gathered.

The TW system must be a collaborative process. All members must have access to a common platform so they can enter information that could be of organizational interest. Given the establishment of some filters for access to information that will be defined by the principal in charge of the organization, this action can become into a collaborative knowledge management strategy by creating feelings of cohesion and making the members feel actors of management strategies of the company.

One of the most important issues about the TW process is the amount of time spent on some stages that can be automatized (see section 5.2.4 and figure 5.19). This happens usually because of the great amount of data collected in the process. Therefore, Semantic Technologies can be used in order to automatize as many tasks as possible and making the process faster (see section 2.3.2).

Interoperability is also an issue that opens research opportunities in this process. Much information can be gathered by Technology Watch process, but if it is not exploitable from other systems (such as the previously mentioned Innovation process), the value of that information decreases. Thus, Semantic Web can be used in order to store that data in an interoperable format and link it with other systems (see section 2.3.3) for its exploitation.

Next section (2.3) describes the selected technologies to solve those issues and the ones that were identified for the Innovation process (see section 2.1).

## 2.3 Technologies to support Innovation and Technology Watch

Emerging ICT have revolutionized the TW field by offering new options when seeking and gathering information. Internet has been a huge source of information where it is not difficult to get large amount of data but it is hard to obtain relevant information. Collaboration has been defined as key for the Innovation process, and Social Web is one of the big step for interaction in the internet, using those technologies multiple agents can simultaneously access to the process and participate in it. But both processes can benefit from the interaction between them, therefore, they need some technologies that link their data and enables interoperability. Thus, this section also analyses Semantic Web Technologies and its interoperability capabilities.

Below, the three key technologies selected for innovation and technology watch will be presented: the Social Web, with which you can manage the collective knowledge of people; the Semantic Technologies, that can help with task automation for reducing the amount of non-productive time of the experts; and the Semantic Web, that can help managing global knowledge with data extracted from the network and link different type of data.

### 2.3.1 Social Web

Social Web is the evolution of the first Web, transforming the plain information web into a collaborative web. This subsection describes the Social Web, some tools and advantages it can bring for the Innovation and Technology Watch processes.

The term Web 2.0, that some use to refer to the Social Web, is attributed to Tim O'Reilly [?]. It refers to a second generation in the history of the development of communications in the Web environment. Using the Internet as a platform, the information is always available and it can be collected at any time. In the first Web ages, the information was provided in a static way. This second generation focuses on providing some tools in order to have dynamic web pages where non expert people can add information to the web.

What prevails in Web 2.0 are the users, the client. Hence the concept of *Collective Intelligence* or *Wisdom of crowds* [?]. The sum of intelligences is superior to each individual product. Exchanging ideas generates more ideas; share, contribute, have attitude to create content is the philosophy of Web 2.0. Therefore, Innovation processes can benefit from this collaboration among agents, giving them the chance to work together.

In order to achieve that, Social Web is composed of 5 tool types in order to manage on-line communication:

- Publishing Tools (Drupal, WordPress, Sharepoint).
- Tools that manage shared content (Wikipedia, MusicBrainz).
- Tools to share multimedia, such as videos, photos, etc.. (YouTube, Vimeo, Flickr).
- Social networking tools (Facebook, Twitter, Google+).
- Virtual worlds (Second Life, PlayStation Home).

Publishing tools can be used in order to have a common place to gather the knowledge of innovation agents. Those tools can be used as a base to generate a platform that gathers ideas, and also comments or ratings about those ideas.

Content and multimedia sharing tools can be used by Technology Watch and Innovation processes. They can have relevant information related to those processes, and enabling the interoperability between those contents, sharing tools can add valuable information to the processes.

Finally, Social networking tools and Virtual Worlds can be used in order to find new agents that can be interested on the Innovation processes, attracting them to boost it. This way, Social Web enables the Open Innovation, described on section 2.1.

There is also another Social Web term, Enterprise Social Software or Enterprise 2.0, that refers to the application of Social Web applications to enterprise environments. Most demanded functionalities for these applications are similar to those of Facebook or LinkedIn but with more control and governance.

As social networking started to grow in popularity a new breed of Web applications took on the market among enterprises; community platforms. Among the core features, community platforms offer all the functionalities inherited from social Web technologies like blogging, wikis or social networking[**?**].

A issue of the series of McKinsey reports on Web 2.0 adoption shows very positive results on the use of social technologies and a majority of respondents say their companies enjoy measurable business benefits from using Web 2.0[**?**]. The use of social webs in the context of enterprise is incipient.

Many companies started to use Social Web for internal communication, and coordination, and the adoption of such tools has seen a rapid rate according to field studies. As a consequence a number of commercial tools have emerged to provide Social Web infrastructure, such as: SocialText[1], Sharepoint[2], CONFLUENCE[3], and Yammer[4].

Apart from the promises of making the internal communication more efficient, and raising the awareness of the ongoing work in the company, the Social Web tools also help constitute a sort of memory of the companies life, and produces other beneficial side effects. Some companies have used the Social Web tools to leverage their social capital and find experts for certain projects inside the company. In particular Kolari[**?**], proposes to use corporate blogs for this task. Social bookmarking has also been suggested as a useful application of the Social Web in the enterprise[**?**].

Some Enterprise 2.0 tools related to the Innovation process are the *Idea Management Systems* (IMSs). They perform a systematic end to end process that goes from the generation

[1] http://socialtext.com
[2] http://sharepoint.microsoft.com
[3] http://www.atlassian.com/software/confluence
[4] https://www.yammer.com/

of ideas to its selection and launch. Such systems provide support for the idea funnel process[5] "where ideas are systematically filtered and assessed against criteria and only the most valuable ones are implemented and put into practice"[6]. Even if idea management systems have been in the market for a while there has been a lately increase in their popularity particularly with the advent of the social Web and the rise of social business software or community platforms. Idea management systems are sometimes seen as subset of such platforms.

Besides, the increasing interest in *Idea Management System* (IMS)[**?**] has pushed community platform vendors to integrate idea management and community software into a single type of platform that includes idea management functionalities and social technologies.

But many barriers are detected, above all organizational barriers such as no implication of managers, need of a cultural change or hierarchical structure. However, initial data show very quantifiable benefits and there is no doubt about the upward trend of adoption of these technologies. A high percentage of companies have planned to increase the investment in 2.0 technologies.

In conclusion, Social Web enables gathering information from many agents and even finding new agents to help on Innovation and Technology Watch processes. But one of the problems is that, sometimes, it carries a big amount of information that is hard to manage. Therefore some automation is needed in order not to treat all that information. Next subsection (2.3.2) describes Semantic Technologies and how they can help in order to solve this type of problems.

### 2.3.2 Semantic Technologies

Semantic Technologies help to derive meaning from information. Their main goal is understanding large or complex sets of data, without having any previous knowledge about it. Semantic Technologies are in many cases complements or supporting tools for the Semantic Web application.

This subsection describes some Semantic Technologies that can help in the Innovation and Technology Watch processes. *Natural Language Processing* (NLP) is proposed for machine understanding of human language; *Artificial Inteligence* (AI) is adopted for task automation; and semantic search is suggested with the objective of finding answers for the searches and not only answers with searched words.

#### 2.3.2.1 NLP

NLP's objective is to enable computers to make sense of human language.

[5] http://www.ifm.eng.cam.ac.uk/dstools/paradigm/innova.html
[6] http://www.think-differently.org/2007/06/what-is-idea-management-system.html

In 2003, Chowdhury[?] described NLP as "an area of research and application that explores how computers can be used to understand and manipulate natural language text or speech to do useful things". The human language has a structure, called grammar, and understanding that structure is one of the biggest efforts in NLP. It is a hard work though, because even humans have difficulties in some contexts. For example, trying to understand a short sentence as *Do you see the man with the telescope?*, could mean two different things:

1. You are using the telescope, and if you can see the man.

2. If you can see a man that is using a telescope.

Depending on the context of the sentence both can be right. So, in order for human language to make sense for computers, it is not enough to understand the grammar but also understanding the whole context.

NLP can be used for different purposes:

- **Entity extraction:** identifies proper nouns and other specific information from a plain text. Tries to map terms to concepts. For example, the text "Resource Description Framework" should map to the same concept as RDF.

- **Auto-categorization:** classify different documents by grouping them based on certain criteria.

- **Question Answering:** given any question, trying to understand it and give its correct answer. As an example, a tool called *Watson*[7] [?] can be found, developed by IBM. Watson, learning through interactions, delivers evidence based responses driving better outcomes. It generates hypotheses, recognizing that there are different probabilities of various outcomes. Watson "learns" tracking feedback, learning from success and failure, to improve future responses.

Several studies analyse the use of NLP as the supporting technique for Semantic Web, being *entity extraction* one of the main areas of interest [?], [?]. There are also some works that aim to extract text and map it to an ontology or semantic element [?], [?].

In conclusion, NLP can be used to identify concepts from a text, enabling the identification of the elements appeared on it. That way, *entities* can be identified and linked between them, relating text with those entities. This enables interoperability without explicitly defining those entities.

#### 2.3.2.2 Artificial Intelligence

Artificial Intelligence (AI) or Computational Intelligence “is the study of the design of intelligent agents” [?]. An agent is something that acts in an environment and an intelligent

[7] http://www.ibm.com/watson

agent is an agent that acts intelligently, doing something appropriate for its circumstances and its goals. It is not necessarily something intelligent, but something that follows a logic. AI can be used for different purposes, below, three examples that can help Innovation and Technology Watch processes on task automation are mentioned:

- **Classification:** *Support Vector Machine*s (SVMs), can be trained giving them some examples as training set. It can learn from some of the features of a document and its classification. Then if a new document is analysed, the SVM can predict its classification. This way, some classification tasks that are made in the Technology Watch process can be automatized, reducing the amount of time spent on them. SVMs is a very good technology for classification, that can adapt to the amount of features (amount of words) and training set size (number of documents). It can adapt to those changes using different type of *kernels* that change the way it learns. Therefore, there are kernels (*Linear Kernel* for example) that performs very well with big amount of features and reduced training sets, a very common case in text classification tasks.

- **Similarity predictions:** giving some similar and less similar examples, a *similarity and metric learning* machine can create a similarity function that can analyse different objects and determine if they are similar or not. This way, related content can be found, avoiding for example repeated ideas on the Innovation process.

- **Pattern recognition:** giving some inputs to an *Artificial Neural Network*, different patterns can be found. It is inspired on the brain and how it works. It can be used for clustering, filtering or even classification.

#### 2.3.2.3 Semantic Search

The most used type of search in the internet is called *keyword search*. It means that the search engine returns a bunch of documents that contain the entered text.

In contrast, rather than returning anything that contains the text, the semantic search tries to understand the context of the search and return a meaningful document.

For example, if you type *What time does the plain from Bilbao to Dublin take off?* into a search engine, probably you are not searching for a document containing that text, but a document containing the answer or the answer itself. So if the search engine returns the document with Bilbao airport departures would be a nicer answer than a document with the text.

Companies such as Google or Facebook recently have implemented some semantic web features on their tools. Google developed *Knowledge Graph*, and according to their official blog, it "contains more than 500 million objects, as well as more than 3.5 billion facts about and relationships between these different objects"[8]. Facebook developed *Graph Search*, that

[8] http://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html

combines its users data and external data into a search engine. This way provides user-specific search results, sorted by *social proximity* that takes into account the publication of related people.

### 2.3.3 Semantic Web

This subsection describes the Semantic Web, and its ability for creating interoperable data spaces. Next it shows the Semantic Web technologies and the standards that can be useful for the Innovation and Technology Watch processes. Finally, it presents the *Open Data* and *Linked Data* and their principles.

Semantic Web (also known as *Web 3.0*, *the Web of Data* or *the Linked Data Web*) supposes the next major evolution in connecting information. The Semantic Web proposes the principles and standards for data interlinking with the objective of making data understandable by computers. This enables automation of sophisticated tasks. While semantic technologies are algorithms and solutions that bring structure and meaning to information, semantic web technologies are W3C technology standards that aim to bring semantics to the data in the web, and making easier to link different kind of data.

Berners-Lee's description [?] says that "The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation". As the Web was the way to link document with each other, Semantic Web tries to do the same with data. Christian Bizer [?] describes one of the problems that Semantic Web tries to solve: "Traditionally, data published on the Web has been made available as raw dumps in formats such as CSV or XML, or marked up as HTML tables, sacrificing much of its structure and semantics.". With the Semantic Web not only the documents are linked, giving semantics and structure to the data, they can be interlinked too.

Two are the main technologies for machines to comprehend linked data on the web: RDF and OWL. Those technologies combined try to make explicit descriptions of content on the web, whether catalogues, forms, maps or any type of documentary objects. This way, the content of the web is structured and linked as in any database, creating the *Linked Data* (LD). This information semantically structured allows content managers, and therefore machines, to interpret digital documents and perform intelligent search, capture and processing of information. Thus, Semantic Web is all about having linked and well defined data on the Web so they can not just display the data, but also: automate tasks, integrate and reuse data between applications.

The role of the technologies involved on the semantic web can be graphically seen on Figure 2.3. On the lower level we can see *the information* itself. It is a structured or semi-structured *dataset* exposed on the Internet. The second level contains the abstract formalization of data or information. It is built by triplets that formalize relationships be-

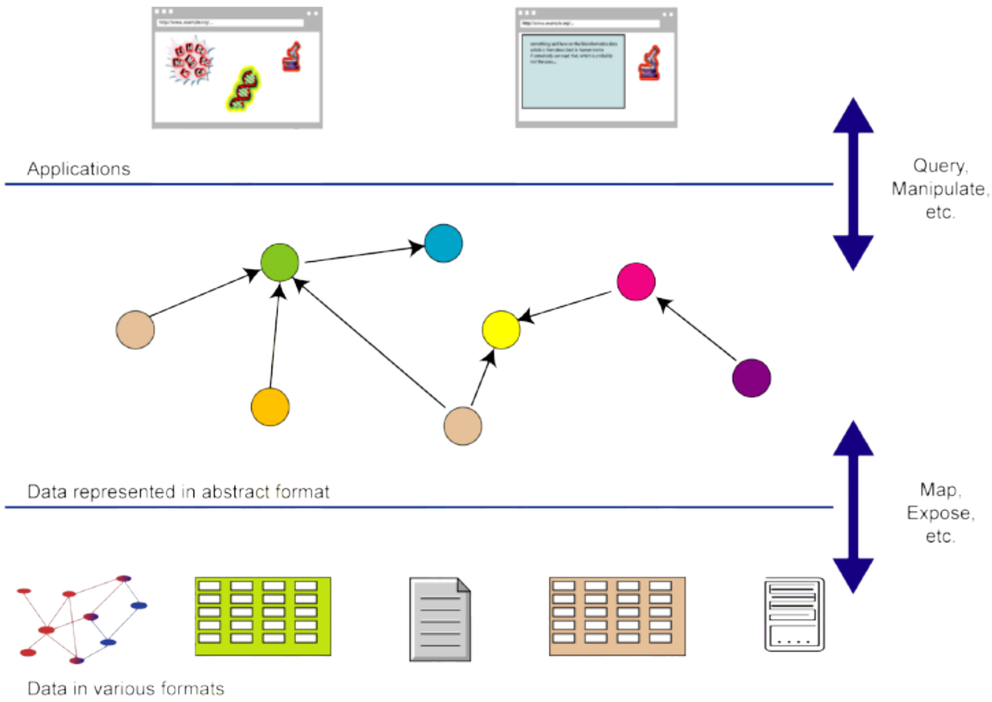


Figure 2.3: The role of the technologies involved on the semantic web ( by Herman, I.[?], page 74)

tween data, which can be represented by directed graphs, and by integrating additional meta-cognition from vocabularies, ontologies, classes, etc. The third level contains the applications or services, which leverage data sets formalized in the previous level to get the desired results. These applications may be based on the *query* of formalized datasets, from the previous level, in the same way that SQL queries a database. They can also be based on inference or logical calculations, but the computational complexity of this task may represent a practical limitation.

Most of the languages or technologies involved on the Semantic Web are being standardised an reviewed by the W3C, an essential task if it aims to be a common mechanism of representation, exhibition and exploitation of the information.

Besides of standardising technologies, the crux of the matter from the second strategic response produced lately in the Semantic Web is *linking the data*. This means that the goal is not merely the annotation or semantic mark-up of the information in isolated repositories, but is the establishment of the greatest possible number of interconnections between different repositories or datasets.

The Linked Open Data (LOD) project[9] is the leading exponent of this response. It has been named as *the seed* that will bring the authentic Web of Data. The *LOD cloud* gives support to some of the most attractive semantic web applications to date, such as DBpedia[10]

[9] http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData
[10] http://dbpedia.org

or the BBC's music portal [11].

Bellow, some Semantic Web technologies and concepts will be explained.

#### 2.3.3.1 Semantic Web Technologies and Concepts

Semantic Web technologies are a family of standard technologies from the World Wide Web Consortium (W3C) designed to relate and describe data from the Web and the enterprises. On 2000, Berners Lee [?] presented those technologies in different layers (see figure 2.4). The lower levels are related with the most *technical* aspects of the data (how to represent a entity identifier, how to describe de data...), while the upper ones describe the *logic* of that data (if a relation can exist, rules to infer information automatically...).

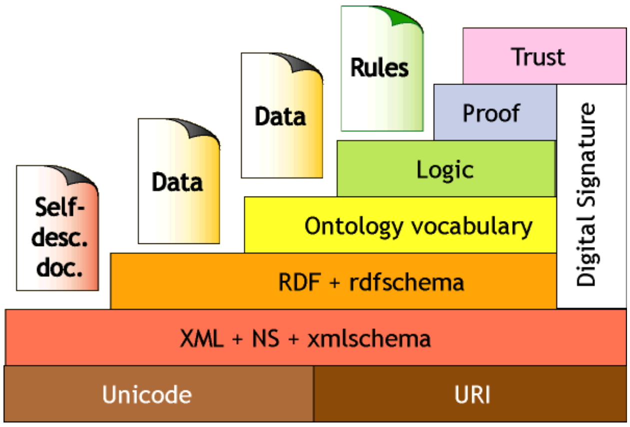


Figure 2.4: Tim Berners Lee's Semantic Web stack.

Next, previously mentioned LD's principles along with Semantic Web basic concepts and some of standards technologies are presented.

#### Principles of the Linked Data and its Publication

Since the announcement by Sir Tim Berners-Lee and its adoption as an official project of the W3C, the publication of new data sets has been massive. Thus, the acceptance of Linked Data principles [?] and the publication of new data sets has brought the Web to a global space that allows the connection of different data sources, leading to the first initiatives for search engines and indexers creation in order to exploit these data [?] [?] [?].

The four principles where the Web of Data should be supported are enumerated next. They were proposed on 2006 by Berners-Lee[12] and they are also known as Linked Data principles:

1. Use URIs as names for things.

[11] http://www.bbc.co.uk/music

[12] http://www.w3.org/DesignIssues/LinkedData.html

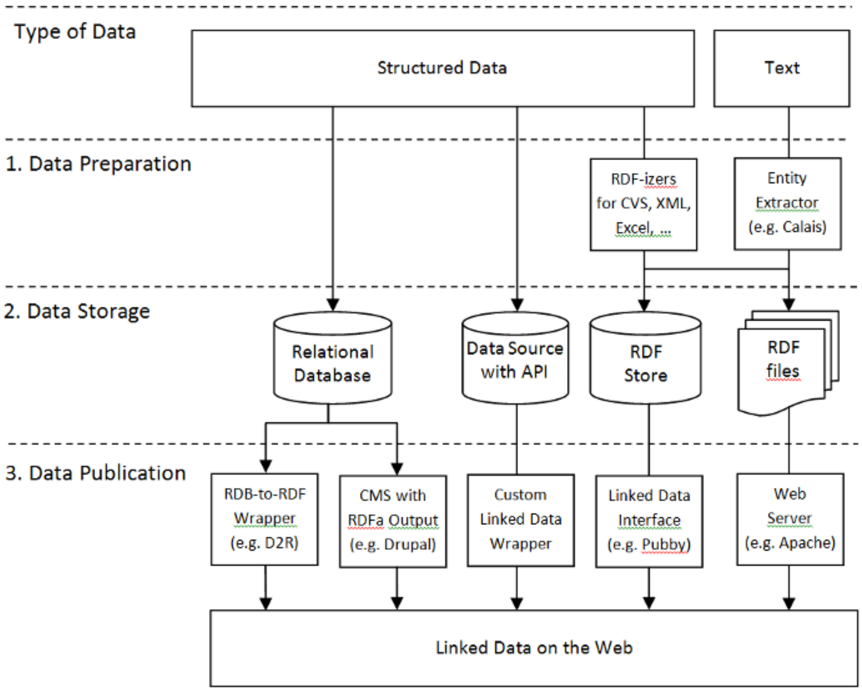


Figure 2.5: Linked Data Publishing Options and Workflows.[?]

2. Use HTTP URIs so that people can look up those names.

3. When someone looks up a URI, provide useful information using the standards.

4. Including links to other URIs so that people can discover more things.

The process of Linked Data publishing requires the adoption of those basic principles. The use of Semantic Web standards along with those principles make interoperability and reuse of data more efficient. However, this does not mean that current data management systems must disappear (such as relational databases). Adding a technology layer to interconnect such systems with the Web of data can be beneficial. There are various mechanisms for accomplishing this and they are summarized in a number of patterns reflected in Figure 2.5.

Berners-Lee[13] presented a classification that measures the level of commitment of a publication with the LD. Those levels are presented next and in figure 2.6:

- ⋆ : Available on the web (whatever format) but with an open licence, to be Open Data.
- ⋆⋆ : Available as machine-readable structured data (e.g. excel instead of image scan of a table).
- ⋆⋆⋆ : As the previous one plus non-proprietary format (e.g. CSV instead of excel).

[13] http://www.w3.org/DesignIssues/LinkedData.html


Figure 2.6: Linked Open Data publication classification according to Tim Berners-Lee

- ★★★★ : All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff.
- ★★★★★ : All the above, plus: Link your data to other people's data to provide context.

For example, Open Data Euskadi[14] could get between 3 and 4 stars on this classification. Some contents are in RDF format, but they are not linked between them.

### Vocabularies and SKOS

On the Semantic Web, vocabularies define the concepts and relationships (also referred to as *terms*) used to describe and represent an area of concern. Vocabularies are used to classify the terms that can be used in a particular application, characterize possible relationships, and define possible constraints on using those terms. In practice, vocabularies can be very complex (with several thousands of terms) or very simple (describing one or two concepts only)[15].

*Simple Knowledge Organization System* (SKOS)[16] provides a model for expressing the basic structure and content of conceptual schemes that are also called vocabularies: thesauri, classification schemes, taxonomies, hierarchies, lists of keywords, tags, folksonomies ... and any other similar purpose scheme. In this sense, by its profoundly theoretical-conceptual nature, SKOS probably represents a first approach to the descriptive characteristics of the ontologies.

According to W3C in basic SKOS, conceptual resources (concepts) are identified with URIs, labeled with strings in one or more natural languages, documented with various types

[14] http://opendata.euskadi.net/
[15] http://www.w3.org/standards/semanticweb/ontology
[16] http://www.w3.org/TR/skos-primer/

of note, semantically related to each other in informal hierarchies and association networks, and aggregated into concept schemes.

In advanced SKOS, conceptual resources can be mapped across concept schemes and grouped into labeled or ordered collections. Relationships can be specified between concept labels. Finally, the SKOS vocabulary itself can be extended to suit the needs of particular communities of practice or combined with other modeling vocabularies.

There is an interest in providing SKOS descriptions for all conceptual schemes, not only new ones but also to those already known for decades. The interest lies not only in its own formal description and/or publication, but in the linking that enables the web, and in particular the semantic web, that could help improving the interoperability between different conceptual schemas. This problem is known from the very existence of the classifications, but it sharpens precisely in the digital environment.

### Ontologies, OWL and OWL2

Actually, there is not a very clear difference between *vocabularies* and *ontologies*, but the second term is the one that it is more associated with the idea of the Semantic Web. According to W3C, the trend is to use the word *ontology* for more complex, and possibly quite formal collection of terms, whereas *vocabulary* is used when such strict formalism is not necessarily used or only in a very loose sense. Therefore, it can be said that *Dublin Core* is called a vocabulary because it defines a series of concepts or terms with some simple relations between them, while when thinking about an ontology it seems that these relationships will be much more complex. However, Dublin Core, FOAF, BibTeX, etc. are still ontologies.

*Web Ontology Language* (OWL)[17] is the language used to describe the ontologies and is designed for use by applications that need to process the content of information instead of just presenting information to humans. OWL facilitates greater machine *interpretability* of Web content than that supported by XML, RDF, and RDF Schema (RDF-S) by providing additional vocabulary along with a formal semantics.

There is a second versión of OWL[18] and many of its modifications are relatively simple (many of them syntactic helps), but there is one that is crucial: the existence of new 3 profiles[19]:

- OWL 2 EL is appropriate for applications where large ontologies are needed and where there are no high level response requirements.
- OWL 2 QL allows using the standard database query technology(SQL), is therefore suitable for relatively simple ontologies but which possibly apply to a very large set of data.

[17] http://www.w3.org/TR/owl-features/
[18] http://www.w3.org/TR/owl2-overview/
[19] http://www.w3.org/TR/owl2-profiles/

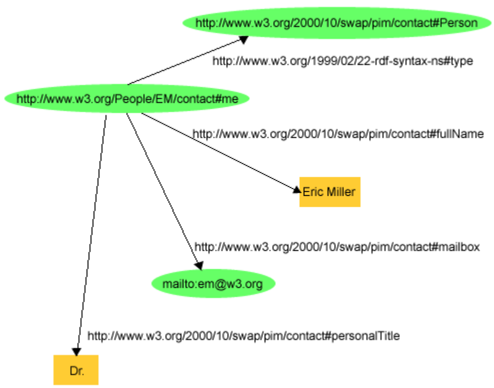


Figure 2.7: RDF graph example describing Eric Miller.[23]

- OWL 2 RL is suitable for operate with relatively simple ontologies, but directly in the RDF triplets.

The purpose of these profiles is to enable better ontology based practical applications, one of the less developed aspects in the Semantic Web by now.

### RDF

*Resource Description Framework* (RDF)[20] is a language to represent relation between resources on the Web. Conceptually, this relations can be represented using a *graph*, therefore sometimes it is said that RDF is a triplets expressing language. Each of the triplets links two pieces of data using a relation (using a graph terminology, a labelled connection between two nodes).

The specific syntax to write the triplets may vary, and there are mainly two varieties:

- RDF/XML[21], that uses XML to represent the RDF relations.
- Turtle[22], a more human readable language.

An RDF graph example can be seen on figure 2.7. The nodes are both green and orange. The difference is that the first ones are URIs, or internet resources, and the last ones *literal* values. Two nodes are linked with an arrow, and its label is the name of the relation. The generalization of this concepts is made this way:

- The origin node is the subject.
- The destiny node is the object.

[20] http://www.w3.org/TR/rdf-concepts/
[21] http://www.w3.org/TR/REC-rdf-syntax/
[22] http://www.w3.org/TeamSubmission/turtle/
[23] http://www.w3.org/TR/rdf-primer/#intro

- The labelled arrow is the predicate.

So an RDF triplet has this form:

**Subject** –*Predicate*→ **Object**

Finally, the triplet's three components are URIs, or syntactically valid URIs. The unique exception can be a literal object. It is not mandatory for the URIs to be an existing one in the *real* Web or on the Internet, they are used as identifiers. This is because the machines need a formal artificial syntax to manage the resources, so the URI syntax has been used, without having any capacity loss.

### Microformats and RDFa

Microformats and *Resource Description Framework in Attributes* (RDFa) are two tools with a similar purpose to RDF: to label semantically content of the Web. The main difference is that RDF describes the data and its relations in its own language and dataspace, while Microformats and RDFa describe that data inside of the traditional web content languages, such as HTML or XHTML. This enables semantic annotation without any additional infrastructure to it.

While Microformats define both the syntax for semantic labelling in XHTML and the vocabulary, RDFa[24] only defines the integration syntax. This makes RDFa vocabulary independent, making it possible to use any vocabulary, such as Dublin Core, BibTeX, FOAF, etc. Therefore, RDFa allows more freedom of labelling. On the contrary, microformats, based on a well-defined vocabulary, allow applications to use the tagged information more easily. This makes RDFa more flexible and Microformats easier to use.

The RDFa standard also describes the way to process the RDFa+XHTML pages in order to extract the corresponding RDF triplets, making it possible to use Web pages as RDF repositories.

### GRDDL

*Gleaning Resource Descriptions from Dialects of Languages* (GRDDL)[25] is a mechanism to obtain RDF triplets from documents based on XML and in particular XHTML pages. Authors may explicitly associate documents with transformation algorithms, typically represented in XSLT, using a link element in the head of the document.

For example, if a web page has been partially labelled using the *hCalendar* microformat, the existing transformation can be indicated as follows:

[24] http://www.w3.org/TR/xhtml-rdfa-primer/
[25] http://www.w3.org/TR/grddl-primer/

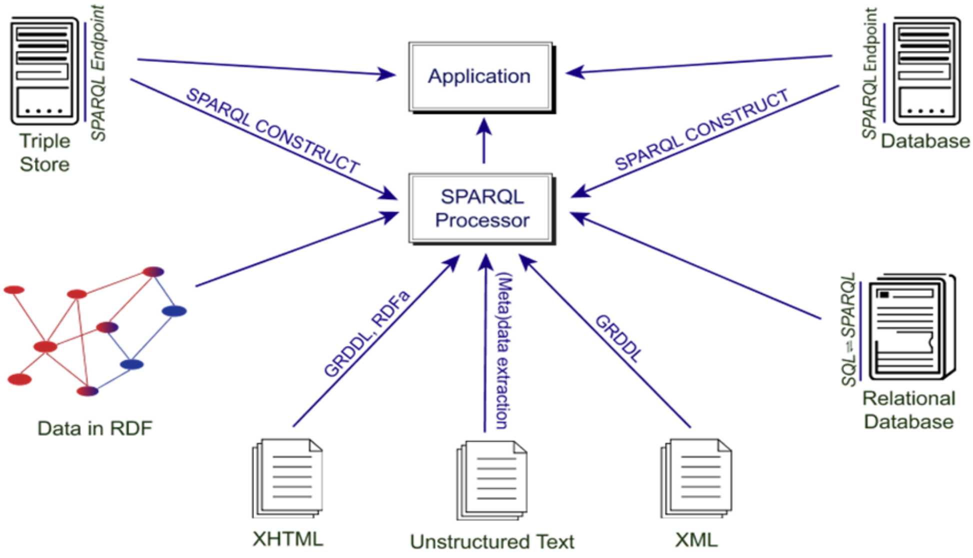


Figure 2.8: SPARQL as unification mechanism (by Herman, I. [?], page 140)

```
<link
    rel="transformation"
    href="http://www.w3.org/2002/12/cal/glean-hcal"
/>
```

This way, *glean-hcal.xsl* style-sheet will apply to the web page in order to obtain the corresponding RDF triplets.

## SPARQL

*SPARQL Protocol and RDF Query Language* (SPARQL)[26] is the query language for RDF triplets. It is also a protocol[27], because it works as a web service: using this protocol, a SPARQL query is converted to a web service request, where a SPARQL processor makes the query and returns the data in XML format.

SPARQL is a standard mechanism for semantic information exploitation. It has become into the unification point for applications' semantic information gathering, as indicated in Figure 2.8.

Thus, offering a *SPARQL endpoint*, major data repositories can expose their data in a common way, either RDF native or databases with other management systems. This way, *wrappers* such as GRDDL or RDFa can be used as data spaces, and in conjunction with SPARQL, they allow accessing to a big amount of data from the semantic web to any application.

[26] http://www.w3.org/TR/rdf-sparql-query/
[27] http://www.w3.org/TR/rdf-sparql-protocol/

## 2.4 Applications of the Technologies in Innovation and Technology Watch context

This section presents the state of the art on the application of the previously outlined technologies in the Innovation and Technology Watch context, outlining some existing tools. The study begins analysing the application of Social Web in the Innovation process and continues outlining the application of the Semantic Web in that same context. Next the application of Semantic technologies and Semantic Web in the different stages of the Technology Watch process is presented. Finally, the applications of those technologies in the enterprise is described.

### 2.4.1 Social Web applications for innovation process

This subsection describes the applications of Social Web technologies in the Innovation process. It shows the current importance of Social Web platforms in the enterprise, some market solutions and efforts made by the community.

In 2010 several architectures of participation were analysed by Errasti et al.[?] including Facebook, Digg, Wikipedia, IBM Idea Factory, IdeaScale Innocentive, SalesFoce Idea Management, Hominex, Mindmeister, LaboraNova, IBM Lotus Connections 2.5, Microsoft Office Sharepoint, Brightidea, Imaginatik Idea Central, Jive SBS, Elgg, Drupal, Liferay, Joomla and Plone. Since the work presented by Errasti et al., the IMS market has presented a whole bunch of solutions. The number keeps changing with new introductions, failures, mergers, etc. Over 20 factors were considered as the comparison criteria for the different platforms. The most relevant criteria was; context gathering, type of license, ease of use, developing language, operative system, integration with social networks, integration with real time web, semantic web, blogs, wikis, RSS, email, etc. The main conclusions extracted from the analysis were:

- All analysed IMSs are idea centred (such as Idea Scale, Mindmeister or Ideatorrent), gathering little context information in the best cases.
- There are open source platforms with similar characteristics to proprietary software. The inclination to select open source is reinforced. Proprietary platforms are discarded.
- Any existent or new innovation support platform must consider the integration of its mechanism with most popular social web platforms, such as Facebook or Twitter in order to be successful by means of participation. Share ideas among those platforms guarantees reaching collaborators and in some cases open the process to new participants.

- Semantic Web technologies can support information management and consequently contribute into better idea generation tasks. Especially, Semantic Web can help in keeping all the information structured in electronic files accessible by anyone from anywhere. Moreover, intelligent agents can then deal with this structured data, to avoid direct human interaction, for instance when searching for new ideas. It can also strengthen ideas by complementing proposed ideas with content found automatically in other sites or repositories.

- Drupal and Liferay obtained the highest score, but the dimension of current and potential community of users and the effort made by the their users to integrate semantic technology or web 3.0 currently favours Drupal.

For the mentioned Jive SBS for example, the adding of idea management to its collaboration suite was announced in March 2010[28]. In the same data Acquia announced the intention to add idea management to the core platform in the Drupal Commons community[29].

After that a list of innovation and idea management software has been compiled by Ron Shulkin[30] and Lauchlan Mackinnon[31] in their respective blogs with over 40 references. The fact is that the market is so huge every vendor in the space can make enough sales to sustain themselves. The market for IMSs tools has grown into a noteworthy niche solution market. As is normal with any emerging technology market, even if the solutions are powerful enough, lack of user awareness creates a barrier to market growth and provider success.

Within a EU research perspective, there have been a few projects dealing with idea management in the last years:

- Disrupt-IT (IST-FP5-33372) is a project finished in 2004 that aimed at the specification of a dynamic management methodology to foster disruptive innovation in smart organisations. One of the outcomes of the project was an idea generation tool that provided a very basic interface for idea posting and evaluation. This initial platform developed by Atos has evolved into a new web site for idea collection, management and evaluation based on social Web technologies[32].

- Laboranova[33] (IST-FP6-035262) is a project that finished in 2010 aimed at supporting innovators, teams and companies within the development and management of innovative ideas and concepts in the early stages of the innovation process. The Labora-

[28] http://www.eweek.com/c/a/Messaging-and-Collaboration/Jive-Joins-Salesforcecom-Spigit-in-Idea-Management-Software-449863/
[29] http://commons.acquia.com/discussion/there-something-ideas-feature
[30] http://www.examiner.com/article/a-list-of-every-innovation-collaboration-cms-ideamanagement-tool
[31] http://www.ideamanagementsystems.com/2010/10/44-idea-management-software-solutions.html
[32] http://pgi2-en.atosorigin.es/node/236
[33] http://www.laboranova.com

nova consortium has developed and combined models and tools in three specific areas: ideation, connection and evaluation of ideas. There are a total number of 13 tools as a result of the project that provide support to different types of innovation and different stages of the innovation process. There are tools that support creativity, ideation, prediction markets or innovation jams. Most of the tools are based on Service-Oriented Architectures and base their data models in existing ontologies but no native support for semantic technologies is provided.

There are more European projects about innovation platforms, but they focus on specific application domains, such as:

- INNOVATION PLATFORM (2010-13): Innovation Management Platform for Aeronautics[34].
- BIO TIC (2012-2014): the Industrial Biotech Research and Innovation Platforms Centre towards Technological Innovation and solid foundations for a growing industrial Biotech sector in Europe[35].

### 2.4.2 Semantic Web applications for innovation process

The Semantic Web is a vision of the future where all information is machine processable and computers may interpret available content and its relationships to help humans in accessing, browsing and searching information. As it has been said, this subsection describes the applications of Semantic Web technologies in the Innovation process. It will show some ontologies that aim to describe the process and how can help making it more successful and the different efforts made in order to support the process.

The objective of this work is in the application or particular use of the Semantic Web rather than in the research of the technologies or foundations that constitute it. In this research project, the Semantic Web technologies are being used as a tool to introduce interoperability among heterogeneous systems and to model the information that is being stored and shared among those systems. Thus, the research concentrates on Semantic Web solutions such as ontology development, annotation with metadata, RDF storage/consume or linked data which enable interoperability and modelling. The specific domain of application has been the Innovation and Technology Watch context and although there are many experiences in other domains the work conducted in this particular one is incipient.

Semantic Web can feed previously mentioned Social Web applications with semantic annotation mechanisms such as Microformats, RDFa, OWL, etc. and enable an intelligent information process also allowing automatic correlation between all the relevant data within

[34] http://cordis.europa.eu/projects/rcn/97671_en.html
[35] http://cordis.europa.eu/projects/rcn/104298_en.html

organizations. The combination of Social Web and Semantic Web technologies enhances interoperability between related contents and already tagged information with meta-data. Especially, in order to align all the information covered into the innovation funnel process, there is a need to annotate all the relations between data. A proper description or model is needed for this purpose. That is an ontology is needed.

In the Innovation field and more specifically in the idea generation stage the research has focused in the modelling and application of meta-data for interlinking on Idea Management Systems (IMS). Two are the relevant innovation ontologies encountered in the literature; the innovation management ontology presented by Christopher Riedl[?] and the GI2MO ontology presented by Adam Westerski[?].

Riedl presents an ontology (Idea Ontology) that applies an approach where the effort concentrates on the integration of idea repositories and little impact is put on interlinking (i.e. relationships and dependencies between concepts).

Westerski[?] cover this aspect by proposing an ontology for IMS where the whole innovation process and its life cycle is taken into account. He presents a formalization of meta-data that can be used to describe ideas and associated information throughout the innovation process. As a result, concepts such as the idea meta-data changes in time and the role of various actors in the Innovation process influence the model in a significantly bigger degree than in the case of Idea Ontology. The ontology is proposed as a universal meta-data schema to be applied in any sort of IMS (see figure 2.9).

Aside those ontologies, there are other attempts to construct models for concepts related to different aspects of the innovation process. Bullinger[?] proposes the concept of Onto-Gate for idea assessment though usage of ontologies that model domain specific knowledge (e.g. product structure, market description, organization strategy etc.). The proposal of Bullinger compliments GI2MO Ontology as a tool that can be connected with existing Idea Management System meta-data to provide a new solution for idea assessment.

Stankovic et al.[?] propose an ontology related to innovation modelling that cover serialization of information system meta-data for integration mainly targeting Idea Marketplaces and as a result only focuses on modelling aspects related to challenges and competitions that are central for this group of systems.

Lorenzo et al. [?] propose an ontology for brainstorming systems that covers a large number of concepts related to idea modelling and communities. They focus modelling community collaborative processes and pays less attention to the management, assessment and metric measurement aspects that are central for Idea Management Systems.

GI2MO ontology has been adopted in this work as the most representative model for the innovation process and consequently it will be the starting point for the research to be carried out. Along with the ontology Adam Westerski has contributed in his thesis work

[36] http://www.gi2mo.org/ontology/

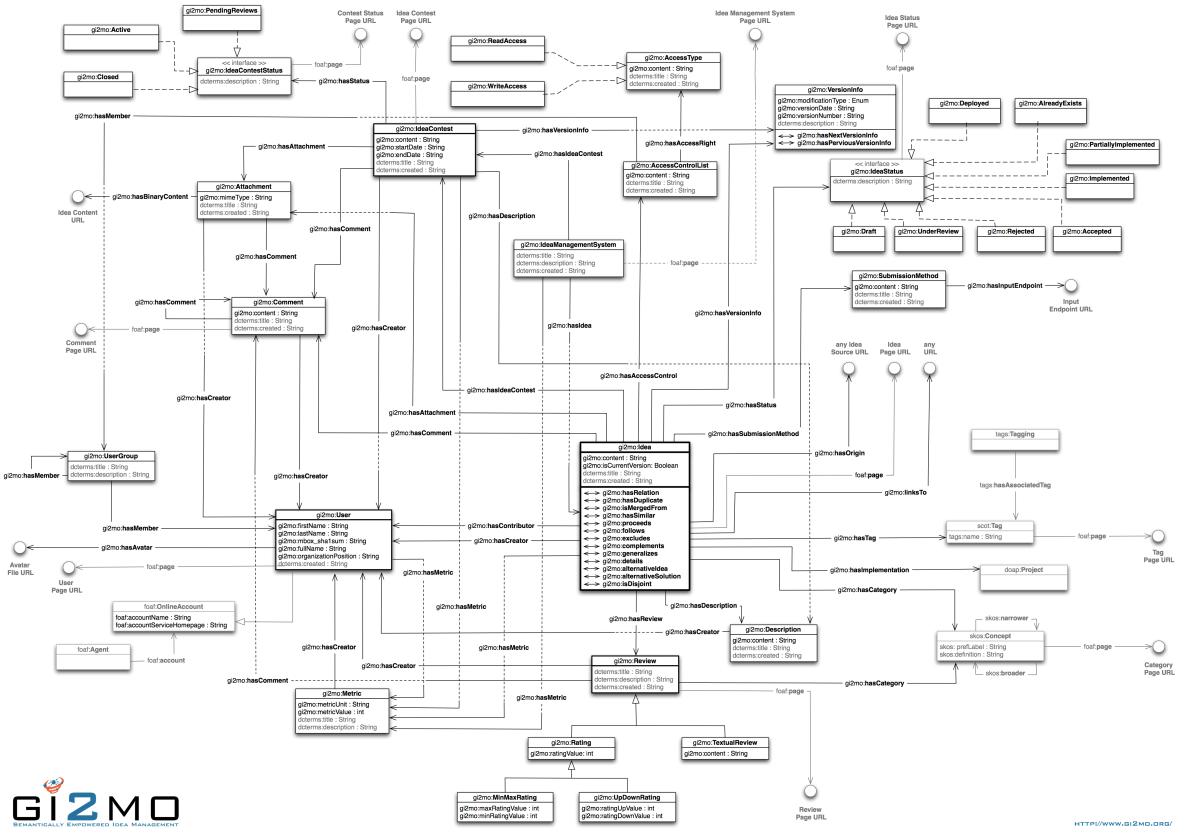

Figure 2.9: UML Class Diagram for the Gi2MO Ontology.[36]

with additional semantic solutions and experiments in the innovation context. The overall objective of the thesis is to obtain more meta-data that would allow better comparison and assessment of ideas. The contribution areas are summarise next:

- Community Opinion Model for Idea Management Systems -The objective at this stage is to use opinion mining techniques on community generated content in the form of idea comments to analyse this data and generate additional metrics for idea assessment. The results show that the utilized technology and solutions build on top of the IMS delivering new information for decision makers that have impact on the acceptance or rejection of ideas.

- Idea Characteristics Model for Idea Management Systems - In this field the thesis identifies the characteristics specific for ideas in the Idea Management Systems independently of the domain or market segment in which the system in deployed, proposing a new taxonomy. It also proposes a number of experiments where those characteristics are applied both manually and in an automatic manner using semantic technologies (machine learning approach). Finally, the thesis delivers a study of transforming annotations into metrics that identify information stored in the IMS. Although metrics derived from innovation models are equally relevant to identification of winning ideas as any other currently used community metrics (up/down ratings, comment count etc.), the proposed metrics allow to identify community behaviour to verify if submitted ideas follow the initially set goals of organizers.

- Idea Relationships Model for Idea Management Systems - The final contribution explores further the relationship types as well as identifies if the new meta-data obtained via previous contributions can be used to facilitate relationship identification (idea dependencies). The contributions in this area are: 1) a classification of the systems related to IMS; 2) a methodology for interlinking those systems by means of extending GI2MO ontology; 3) hierarchy of relationships between ideas and its use for clustering of ideas.

Among the many lines of future research identified in Westerski's thesis, it is worth mentioning the following because of their relation with the proposal outlined in this document:

- Further research on automatic idea annotation. The thesis experimented with automatic idea annotation using a single method: supervised machine learning approach based on k-NN algorithm with nearest neighbours detected using keyword similarity. Westerski proposes the evaluation of other methods and recommends automatic annotation for the full taxonomy.

- Clustering based on idea characteristics. The thesis delivered a number of studies on idea characteristics and quantitative analysis of those characteristics. The author proposes to test idea similarity considering more than a single characteristic. A potential direction for this kind of research could be the use of clustering algorithms and the treatment of idea characteristics as feature vectors. It remains to be seen if such methods could deliver distinctive clusters of ideas that could have a meaningful impact on analysis of idea datasets.

- Idea annotation with domain ontologies. The thesis studied automatic annotation in the context of domain independent taxonomy in order to deliver a tool for comparison of different IMS deployment. As an extension of this work, the thesis also investigated impact of those annotations on idea similarity. In terms of future work, the author proposes to evaluate the use of domain ontologies in the same way, i.e. automatic annotation of ideas with concepts related to that domain, development of metrics based on those annotations, and computation of idea similarity based on domain related annotations.

- Improved Enterprise Linked Data evaluation. Westerski's work had a major problem for the evaluation of idea relationships. The lack of sufficient data and large enterprise partners that would share their information for the needs of experiments. In terms of future work, he proposes the evaluation of the proposed solution in the environment of a large enterprise.

- Usage of the newly discovered idea relationships. The thesis presented results of clustering based on new relationships. The author points to the use of new relationships in other ways like idea recommendation or ranking.The future work in this area should verify if such a use would aid idea assessment and if ranking generated based on relationships would have an impact on ideas implemented.

- Automatic idea mash-ups. The research has shown that ideas are not only duplicates but are connected to each other in a variety of ways. The thesis experimented with downsizing the idea dataset via clustering but perhaps a viable future line of research could be allowing users to mash-up ideas together from the existing idea database. The room for novelties is quite broad there and could include research on: idea mash-up operators, idea similarity for automatic mash-up suggestions, metrics for ranking the mashed ideas vs. regular ones, and finally research on incentives and community take-up with relation to reusing ideas of other people.

As it will be shown later the research proposal of this work will be based on this contribution, exploring the relation among IMS and Technology Watch platforms or repositories.

Finally it must be addressed that some of the tools proposed in Westerski's work have been also adopted for this research. For example the developers of GI2MO ontology created a RDF meta-data publishing module for Drupal called RDFme that has been integrated with the innovation platform developed by our team (Innoweb).

It is expected that by applying Semantic Web features into the process, performance will improve. For example avoiding data duplicity, applications will be able to offer or suggest similar ideas. Moreover, semantic data will enable the possibility of grouping similar ideas and make it easier to users to handle all the information. Semantic searches are also another feature that can improve the potentiality of data usage into the idea management process. By adding semantic information to data, users will retrieve better solutions and more related information from all the content.

### 2.4.3 Semantic Technologies and Semantic Web applications for Technology Watch

Semantics not only can help the Innovation process but also the Technology Watch one. This subsection describes the applications of Semantic and Semantic Web technologies in the TW process. First, it introduces the functions where ICT can help TW process and a classification of the technologies. Next, some information sources will be outlined, the kind of formats that are commonly used in order to publish the information and tools in order to exploit that data will be presented. After that, existing Semantic Technology tools that can help on that data exploitation will be presented. Finally, Semantic Web applications that could assist TW will be described.

In relation to the role of the techniques and tools of *Technology Watch* (TW) and *Competitive Intelligence* (CI), many of them have found great support in ICT through different software tools, basically, because the treatment of a large amount of information without any computer aid could prevent convenient results. As well as collaborating with software tools in order to monitor both competitive and technological environment, ICT can perform an extraordinary work in the development of CI systems. Thus, the following functions can be performed through software tools:

- Identify information sources.
- Capture and organize information according to the needs of the organization.
- Support the work of the Competitive Intelligence Unit by assigning tasks and monitoring the treatment that has been given to a particular document.
- To facilitate the analysis task, distribution of information and the results obtained from the analysis.

ICT tools optimize each stage of the CI system. Thanks to a higher level of automation, systematization and personalization, they enable greater investment in higher value-added tasks from the professional CI, and also being able to cover more in less time. Moreover, ICT tools provide great advantages to all users receiving this information: key information reception and more easily analysable and, therefore, consumable. Figure 2.10 shows a classification of the different semantic technologies.

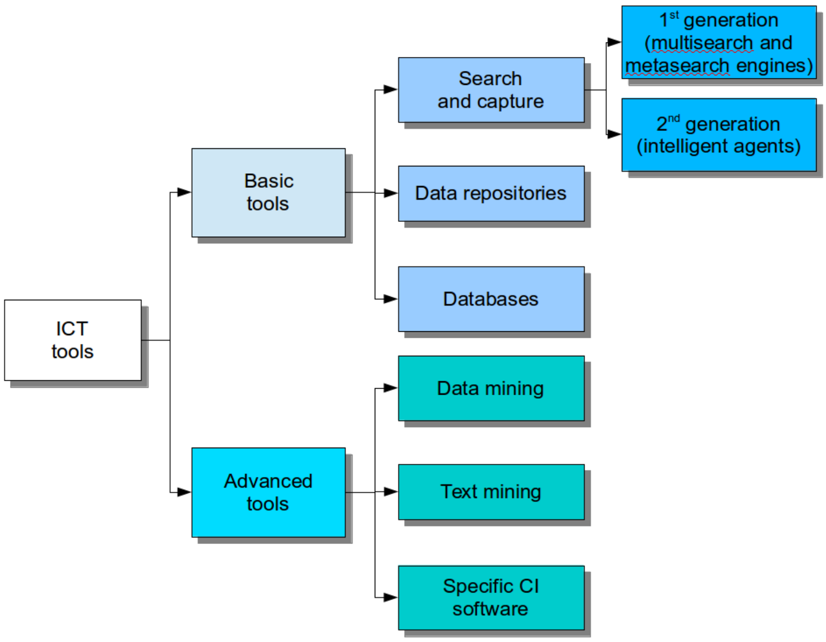


Figure 2.10: Semantic Technology classification.

Bibliometric software, for example, provides users with easy and useful tools for them to treat and recover large amounts of information, introducing the notion of evolution by technology watch and foresight. The continued progress in information technology has encouraged the development of useful statistical tools for Scientometrics, where France has been a pioneer in the development of these statistical programs designed exclusively for scientometric analysis. These are the main advantages of the analysis programs:

- Saving time by reducing the laborious analysis tasks.
- To analyse trends.
- Profiles of the activities that competitors patent.
- To identify trends in specific technology areas.

Table 2.1 lists the countries and their most recent and important developed analysis programs.

On the other hand, regarding to patents, it should be noted that the existing information sources are many and each one focuses on specific geographical areas. This makes the searches complex and the need to look something up in numerous sites. Some of this information sources and patent databases that provide the data for free are listed below:.

| SOFTWARE | CONTRIBUTION |
|---|---|
| | **FRANCE** |
| Matheo Patent | French Software specialized in the scientometric treatment of information, particularly aspects relating to Intellectual and Industrial Property. |
| Tetralogie | Complete solution for analysing large volumes of scientific information and patents |
| Dataview | Software tool for treatment of scientific and technological information experts. |
| | **JAPAN** |
| PAT-LIST | It offers two products for Industrial Property Management:<br>• PAT-LIST-WPI ver. 3.0E<br>• PAT-LIST-CLM (IFI) ver. 3.0E |
| | **UNITED STATES** |
| PatentLab II | Analysis software capable of visualizing the relationships between a wide range of patent data in graphs, tables and reports. Analyses data for the company to make important decisions. |
| Derwent Analytics | Data mining software and visualization of large volumes of patent information obtained from the databases of Thomson Derwent. The aim is to extract critical approaches to make business decisions. Its use is easy and intuitive. |
| VantagePoint | It enables to quickly analyse the search results of bibliographic databases and literature R&D. It is specifically designed in order to interpret search results from science and technology databases. |
| ClearForest Analytics | Analytical system specifically designed for text analysis, which adds value to the existing tools for Business Intelligence. |
| Aurigin-Aureka | It enables to annotate and organize data, to use advanced analysis tools, data mining and text, or to disseminate its results and provide visualizations. |
| Anacubis Desktop | New and sophisticated tool of analysis of relationships between entities. It's used to analyse data obtained from a wide range of sources and file types, both textual and quantitative. |
| | **SWEDEN** |
| Bibexcel | Program developed specifically for the handling and transformation of bibliographic records. |

Table 2.1: Analysis programs by Country.

- OEPM (Spanish Patent and Brand Office) under the Ministry of Industry, Tourism and Trade, offers through its website access to databases of different kinds, inventions and designs in Spanish (Invenes), inventions in Spain and Latin America (Latipat), brands location, international classification of products, services and patents, records condition, issues of law and access to other databases in other languages. Besides, it also offers other related services to technology watch such as creating Patent Technology reports, Technology Watch reports, Retrospective searches and Technology Watch bulletins.

- European Patent Office: it belongs to The European Patent Organisation, an intergovernmental european organization founded in 1977. It allows searching for patents in different languages (German, English and French) on its database (Espacenet[37]). It also offers other services such as searching of publications, alerts about changes in documents, patents reports and RSS service (Open Patent Services). Finally, it allows worldwide access to other data sources.

- United States Patent and Trademark Office: It is under the Department of Commerce of the United States and provides numerous utilities for patent and trademark search. Not only offers tools and services similar to those of the other offices but provides an extensive legal service in relation to patents.

Many countries offer free access to their patent collections. Below, some of those country databases are listed:

- Japan Patent Office (JPO)[38]. This site also offers access to automatic translations of Japanese patents .

- World Intellectual Property Organization (WIPO) provides the search service PATENTSCOPE ®[39], which contains a search engine of international published patents and automatic translations of some documents as a list of worldwide patent databases.

- Korean Intellectual Property Rights Information Service (KIPRIS)[40].

- Other International Intellectual Property Offices that offer search engines in databases are: Australia[41], Canada[42], Denmark [43], Finland [44], France [45], Great Britain [46], Ger-

[37] http://es.espacenet.com/
[38] http://www.jpo.go.jp/
[39] http://patentscope.wipo.int
[40] http://eng.kipris.or.kr/
[41] http://www.ipaustralia.gov.au/auspat/
[42] http://brevets-patents.ic.gc.ca
[43] http://www.dkpto.org/
[44] http://patent.prh.fi/
[45] http://www.inpi.fr/fr/services-et-prestations/bases-de-donnees-gratuites/base-statut-des-brevets.html
[46] http://www.ipo.gov.uk/

many [47], India [48], Israel [49], Sweden [50], Norway [51], Switzerland [52] and Taiwan [53].

Digital information sources which are relevant to Technology Watch are usually presented in different formats. With regard to blogs, forums, websites, newspaper articles, etc., usually the formats used are those of the web (HTML, XML or similar).

For publications, bulletins, patents... besides the above web formats other ones are often used, especially PDF or Postscript format. In many cases this type of source is stored in databases which do not always allow access.

Many organizations, especially the administration, put general data to everyone's disposal on what is commonly known as Open Data. It's becoming more likely that significant data about technology watch is offered to companies as Open Data. The formats used in this philosophy are very varied: DOC, XSL, ODF, PDF, CSV, ZIP or RDF-XML and they can also be offered through APIs or Web Services.

There are many forms to exploit these digital sources, using tools to gather all that data. Duran et al.[?] described some tools to obtain information from different databases. Fernandez et al.[?] also mentioned some tools to help in the stages of TW process. Table 2.2 shows those tools and some recently produced applications, that can be usefull for TW.

| **NAME** | **DESCRIPTION** |
|---|---|
| **Monitoring media** | |
| Iconoce[54] | Scans about 500 media, collecting 1,500 news/day.<br>Allows full text search.<br>Alert Service. |
| iMente [55] | Collects news headlines from 13,000 sources in 13 languages, 55,000 news/day.<br>Allows search in title and text.<br>Alert Service. |

Continued on next page

[47] http://www.dpma.de/
[48] http://ipindiaservices.gov.in/patentsearch/
[49] http://www.justice.gov.il/MOJEng/RashamHaptentim/
[50] http://www.prv.se/en/
[51] http://www.patentstyret.no/en/
[52] http://was.prv.se/spd/search?lang=en
[53] http://twpat.tipo.gov.tw/tipotwoc/tipotwekm
[54] http://www.iconoce.com
[55] http://www.imente.com

| | |
|---|---|
| Acceso[56] | Makes selective gathering from diferent type of:<br>Press releases.<br>Events or courses organized by a company.<br>Financial Releases.<br>Newswire. |
| My News[57] | Full text of more than 100 newspapers, mostly Spanish.<br>Alert Service.<br>Check online archive. |
| **Spanish and international company watch** | |
| Ardan[58] | Business reports.<br>Analysis on competition. |
| Informa[59] | Business data and reports.<br>Balance analysis. |
| e-Informa[60] | Bussiness data, business reports and grants. |
| Dun&Bradstreet[61] | Busines reports.<br>Credit reports.<br>Incidents notifications. |
| Dialog Company Profiles[62] | Gathering information about companies from other databases. |
| **Webpage watch** | |
| Karnak[63] | Sends an alert with the number of new results on a topic of interest. |
| Tracerlock[64] | Sends an alert with the new results on a topic of interest or the changes of a web page. |
| Northernlight[65] | Sends an alert about new results on a topic of interest. |
| **Web monitoring with crawlers agents** | |
| | Continued on next page |

[56] http://www.acceso.com
[57] http://www.mynews.es/
[58] http://www.ardan.es
[59] http://www.informa.es
[60] http://www.e-informa.com
[61] http://www.dnb.com
[62] http://login.profiles.dialog.com/dialog
[63] http://www.karnak.com
[64] http://www.tracerlock.com
[65] http://www.northernlight.com

| | |
|---|---|
| WebsiteWatcher[66] | Enables:<br>Checking an unlimited number of web sites for changes.<br>Record macros to reach websites whose address is ignored.<br>Use regular expressions to define the filtering mechanisms.<br>Check for updates automatically at a speed of over 100 web addresses per minute. |
| Copernic Agent-Pro[67] | Meta search engine able to filter results with keywords, sort by relevance, etc.<br>You can send email with new search results or changes in a page.<br>It integrates as IExplorer toolbar. |
| **Web browser pluggins (IExplorer.)** | |
| Vivisimo tool bar[68] | It is a meta engine that analyzes the text of the results, create categories according to most representative terms and groups the results in those categories. |
| Copernic-Meta[69] | Integrates the search window on the taskbar. It can be customized in order to include any search engine. |
| Scirus tool bar[70] | Permite lanzar búsquedas en varias secciones de Scirus o en el buscador Alltheweb. Allows you search in several sections in Scirus or Alltheweb search engines. |
| **Agregators** | |
| Bloglines [71] | It is a web based RSS feed agregator. Once registered in the web aplication, a user can subscrive to different feeds and recive the updates automatically. |
| Feedly[72] | It is a RSS feed agregator, similar to Bloglines. It has been one of the most benefited aplications of the Google Reader closure. |
| Google Reader[73] | It was a RSS feed agregator similar to Bloglines. |
| | Continued on next page |

[66] http://www.aignes.com
[67] http://www.copernic.com
[68] http://vivisimo.com/toolbar/toolbar-download.html
[69] http://www.copernic.com/en/products/meta
[70] http://www.scirus.com/srsapp/toolbar/
[71] http://www.bloglines.com
[72] http://www.feedly.com
[73] Discontinued service since July 2013

| | |
|---|---|
| Google Currents | It is a publication agregator. The main difference with the previous ones is that the user does not enter the feeds that they want to subscrive to, they just agregate some interesting topics and they recive related publication. This makes easier for the users to search an specific subject, but they lose control over the information sources. |
| Flipboard[74] | It is another publication agregator. |
| Del.ici.ous[75] | It is a social bookmark manager. It is not exactly an agregator, but any user can add gather bookmarks and group them using folcsonomies (tags). The system can also thell the users how many people has the same bookmark and suggest tags for it. |
| **Data collector tools** | |
| Zotero | It is a Firefox web browser plugin. It creates a database with user's bookmarks and enables exporting that database in different formats (reports, bibliography...). |
| Connotea[76] | It was a free online reference management and sharing tool, that could gather bibliography, keywords and tags, and share them. |
| **Bibliography reference management tools** | |
| Endnote | It is a web based bibliography management tool. |
| JabRef[77] | It is a bibliography management tool that works localy. Stores the bibliography in a *.bib* data base and works natively with writing tools like *LaTeX*. |
| RefWorks[78] | It is a web bibliography management tool. Stores many kind of bibliography formats and can export or show the bibliography in many styles. |

Table 2.2: Tools to help in the Technology Watch process stages.

The data gathered using this tools is usually extensive and hard to exploit, therefore, Semantic Technologies are used in order to automatize tasks and classify the data. The use of those technologies in the field of knowledge acquisition is extensive. Many are the publications that employ NLP, AI or Semantic Search to withdrawn additional information

[74] http://flipboard.com/
[75] http://delicious.com/
[76] Discontinued service since March 2013
[77] http://jabref.sourceforge.net/
[78] http://www.refworks.com/

from databases, web sites, documents or similar. The objective at this stage is to focus the study on the Technology Watch area. The scope of application of the technology is diverse but can concentrate in information search, information identification and extraction, and classification or categorization.

One of the techniques that has attracted much of the scientific interest is the application of NLP with the goal to turn text into data for analysis, document classification or domain identification. This technique is called *Text Mining* (TM) or text data mining and refers to the process of deriving high-quality information from text. Text Mining is now a wide area of research that provides useful techniques that can be used in the context of technology watch.

According to Jacquenet and Largeron[?], the term appears for the first time in 1995 (Feldman[?]) and was defined by Sebastiani[?] as the set of tasks designed to extract the potentially useful information, by analysis of large quantities of texts and detection of frequent patterns. Losiewicz et al. [?] for example show that clustering techniques, automatic summaries, information extraction can be of great help for business leaders. Zhu and Porter [?] show how bibliometrics can be used to detect technology opportunities from competitors information found in electronic documents. Another use of text mining techniques for technology watch has been proposed in Lent et al. [?] where the authors tried to find new trends from an IBM patent database using sequential pattern mining algorithms.

In all these works, text mining techniques have been mainly used to help managers dealing with large amount of data in order to find out frequent useful information or discover some related works linked with their main concerns.

Bolasco et al. [?] presents a study on the application of Text Mining in different scopes. They claim that there isn't a "ready for use" instrument available for users and usable to handle an entire TW process. Instead they identify concrete cases of application. More specifically in the analysis of patents, they identify the use of TM techniques to retrieve textual information in the short description of the patent, to complete the picture offered by the codified information. To achieve that purpose they propose the following TM techniques; text indexing and extraction, concept clustering and graphical display and navigation. Evaluation and analysis of relevant documents such as scientific literature and congress papers, newspapers, technical journals or Internet content is another dimension where TM technologies can be applied. They also identified the use in the KM field (document classification) and in Customer Resource Management (CRM) to collect or induce customer opinion (Opinion Mining). The article also describes the necessary steps to correctly implement a TM project. These are presented in table 2.3.

Finally they identify the following issues that need to be addressed when boarding a TM project: The intervention of experts at the time of annotation or term list (taxonomies) definition is a must. Relevant sources and documents must be identified and metrics to

| | |
|---|---|
| | **DOCUMENT PRE-PROCESSING** |
| 1 | Defínition of rules for extraction/collection of text (data selection and filtering) |
| 2 | Defínition and identification of document format. |
| 3 | Text normalisation (cleaning, recognition of dates and of currencies, ... ). |
| 4 | Reduction and transformation of text (elimination of stop words, identification of proper names). |
| | **LEXICAL PROCESSING** |
| 5 | Choice of unit analysis: words (tokens or lemmas) and multiword expressions, terms. |
| 6 | Definition of grammatical rules to solve text ambiguity. |
| 7 | Linguistic and lexical analysis (lemmatisation, key words detection, other tagging). |
| 8 | Definition of semantic categories to be searched for in the text (marking of terminology of interest), extraction of key words. |
| 9 | Classification according to concepts and/or other meta-data, information extraction. |
| | **TM PROCESSING** |
| 10 | Classification of texts. |
| 11 | Clusterisation of texts and summarisation. |
| 12 | Knowledge extraction (in some cases integrated with the aid of experts). |
| 13 | Visualisation techniques. |
| 14 | integration of TM results with data mining processes. |

Table 2.3: Necessary steps to correctly implement a Text Mining project.

measure the impact set. The importance of modelling or structuring information based on a specific domain is crucial. Content enrichment is limited by the definition of the domain. The established scope for the domain limits the learning capabilities.

Text Mining techniques have been employed not only to detect frequent useful information. One important goal of technology watch and more generally business intelligence is to detect new, rare, unexpected and hence generally infrequent information. In Jacquenet and Largeron text mining techniques are used to discover unexpected documents in large corpora of documents (patents, scientific papers, datasheets, . . . ).

Apart from Text Mining, other Semantic Technologies have been also employed in fields that can relate to TW. For example in the field of Topic Detection Tracking, Rajaraman and Tan [?] proposed to discover trends from a stream of text documents using neural networks.

Ibekwe-SanJuan propose Semantic Technologies to find relations among text for thesaurus construction and maintenance. In their methodology, they combine NLP with a clustering algorithm and an information visualization interface. The resulting system called TermWatch, extracts terms from a text collection, mines semantic relations between them using complementary linguistic approaches and clusters terms using these semantic relations. They point to the possible advantages of using external semantic resources to complement the other two types of resources they use; internal evidence (structure of the terms themselves) and contextual relation markers (conceptually related terms). This can be done via a domain thesaurus, ontology or a general language resource.

But Semantic Technologies are not the only ones that can be used in order to assist TW process. The use of the Semantic Web in the Technology Watch field has also been the target for this study. Although we have not found an ontology that represents the Technology Watch process as was the case with the Innovation process the work developed in the field is prolix. Mainly the ontologies encountered deal with the representation or modelling of information extracted during the Technology Watch process but do not address the process itself or the interoperability issues of this process with other enterprises systems or platforms. The classification identified in this area, groups proposals where general ontologies are presented and proposals where specific ontologies are developed to solve Technology Watch challenges.

General ontologies model how the information of publications, patents, feeds or similar should be structured semantically. With relation to publications and patents, the Korea Institute of Science and Technology Information (KISTI) provides a platform called OntoFrame[79]. OntoFrame is an information service platform that uses Semantic Web technologies. It includes OntoURI a semantic knowledge management tool that creates ontology schema and instances and identifies co-references between ontology individuals; OntoReasoner an inference engine that stores and infers ontology-based RDF triples and answers SPARQL queries; and Mariner, that provides search functionality. The ontology models research entities (e.g., persons and institutions), their accomplishments (e.g., articles), publications which indicate specific journal issues or proceedings, locations and topics. The ontology connects researchers to the affiliations they were members of at the time they had their accomplishment. It also connects institutions to their locations such as countries, states and cities. The ontology schema is composed of 16 classes and 89 properties, and supports inference (see figure 2.11). Currently, the semantic portal contains about 450,000 articles, which have been written by 1.35 million researchers, on 340,000 topics, who work for 90,000 institutions, spanning 410,000 locations. In total, the system has a total of about 300 million RDF triples.

The ontology schema model and the ontology instance model are subject to license although a non-commercial permit could be acquired.

Another general ontology for the modelling of information received by means of syndication is AtomOWL. AtomOWL is an ontology whose aim is to capture the semantics of rfc4287. RFC4287 is a format to syndicate online content, such as weblogs, podcasts, videocasts, etc. Syndication is a helpful way to alert interested readers to changes to a web site, be it to new content or changed content. *Really Simple Syndication* (RSS) is commonly used once sources of information are identified in the Technology Watch process.

As AtomOwl captures the semantics of rfc4287 it is easy to convert rfc4287 feeds to AtomOwl statements (see figure 2.12) and thus add them to a triple database, which can then be queried using a SPARQL endpoint. This should help looking for updated content by making powerful queries. AtomOWL being built on RDF is very easily extensible and it

[79] http://www.w3.org/2001/sw/sweo/public/UseCases/OntoFrame/

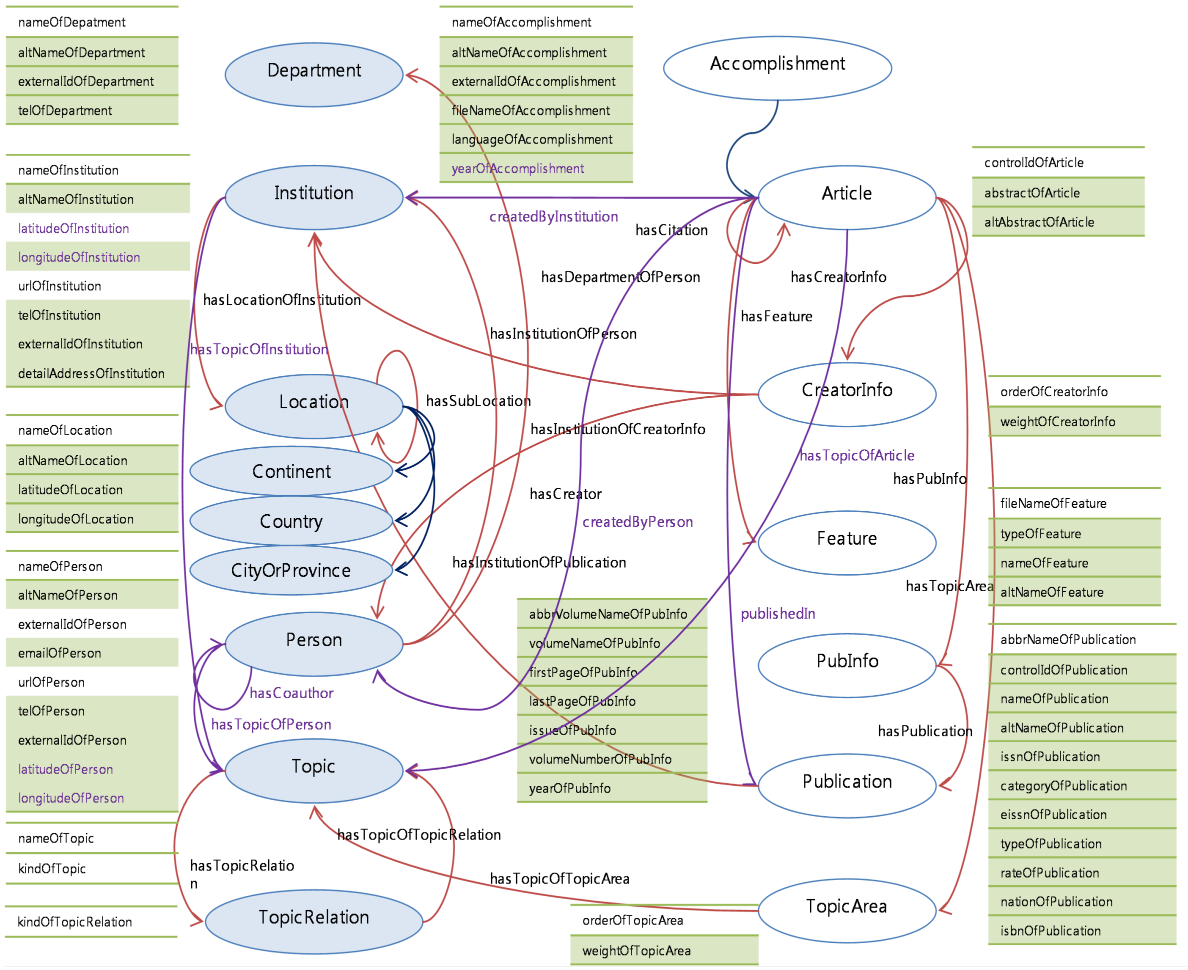


Figure 2.11: KISTI Reference & Academic Ontology for intellectual property and patents.

meshes very well with other Ontologies such as FOAF or SIOC.

Other research points to the importance of developing specific domain relevant ontologies. That is, they propose the use of ontologies that model the Technology Watch target domain.

Thus, Maynard et al.[?] present a knowledge Management platform that integrates a variety of technologies to observe resources in the internet. The method for Information Extraction (IE) and annotation proposes the use of specific ontologies (Employment and Chemist Industry domains).

They claim that "one of the important differences between traditional IE and Ontology-Based IE (OBIE) is the use of a formal ontology rather than a flat lexicon or gazetteer structure. The advantage of OBIE over traditional IE is that the output (semantic meta-data about the text) is linked to an ontology, so this enables us to extract much more meaningful information about the text, for example making use of relational information or performing reasoning. Another difference is that OBIE not only finds the (most specific) type of the extracted entity, but it also identifies it, by linking it to its semantic description in the ontology. This allows entities to be traced across documents and their descriptions

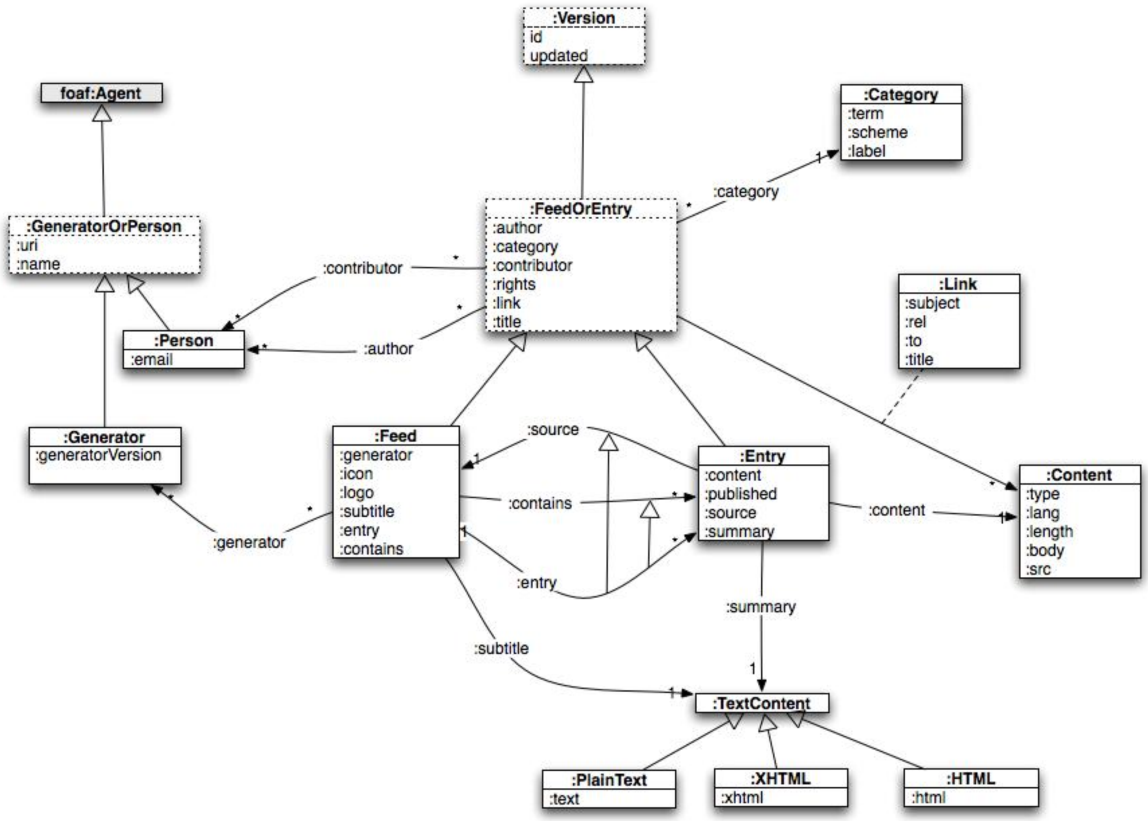


Figure 2.12: AtomOWL ontology for feeds (RSS).

to be enriched through the IE process". Thus the application of the platform (including the domain ontology) can be used for analysis and enhancing discovered information (advantages of identifying instances from the ontology in the text).

The tools proposed are:

- A tool/model for the development of ontologies, which can be used to describe concepts and trends in the user's domain of interest;

- A tool/model for the development of generic and targeted search agents which can use these ontologies to search for business intelligence from diverse web-based sources;

- A platform for integrating information from various sources and consolidating, analysing and publishing this information.

They also outline the 5 basic stages of their method:

- Web mining application to find relevant documents (or manual input of relevant documents).

- Selection of concepts in which the user is interested;

- Information Extraction (IE) and annotation (identifying instances from the ontology in the text).
- Visual presentation of results (annotation of instances) and statistical analysis.
- Ontology modification (an ontology editor is used to enrich the existing ontology from the results of the analysis).

The main inputs for the platform are the feeds gathered from Internet and the domain ontology. They also identify the need of domain specific lists to support annotation with the ontology. Here there is room for domain experts allowing manually annotation at the beginning to follow with automatic annotation once the taxonomy and the ontology are well defined.

As drawbacks they identify the following:

- Annotation complexity depends on the domain's complexity. The identification is easier in rule-based systems than in learning-based systems where training data is required. Part of the annotation is manual.
- Adaptation of the platform/method to new domains is not easy task for non domain/IE experts.

### 2.4.4 Technology applications in the enterprise

This section shows how Semantic Web and Semantic Technologies currently help the enterprises in order to integrate data from different sources, briefly describing some existing tools. In order to address data integration issues in the Enterprise, Semantic Web provide an adapted solution. Since it defines languages and design principles for data interoperability, it can be efficiently used to integrate data from several sources. Among the seminal work in the area, following the long tradition of middle-ware systems, Maedche et al. [?] defines enterprise knowledge management systems using ontologies through OMKS – Ontology-based Knowledge Management System. This focuses on integrating and aligning several internal data sources (databases, directories ...) thanks to a central middle-ware system. Recently, several use-cases and case-studies about Semantic Web technologies in the enterprise gathered by the W3C[80] also showcase the use of Semantic Web for make enterprise data more interoperable, such as biomedical information management at Eli Lilly[81]. A relevant approach in these use-cases is the work done by NASA for expert finding[82], Other approaches for semantic web in the enterprise focus on direct data exchanges between applications in the

[80] http://www.w3.org/2001/sw/sweo/

[81] http://www.w3.org/2001/sw/sweo/public/UseCases/Lilly/

[82] http://www.w3.org/2001/sw/sweo/public/UseCases/Nasa/

Enterprise Architecture Integration (EAI) realm, as proposed by Anicic et al.[?], using OWL ontologies and dedicated scripts to align XML inputs and outputs of several applications. However, these approaches generally do not take into account users and their social interactions, as they mainly focus on knowledge extracted from static knowledge-bases and do not convey a collaborative aspect, neither exploit fully knowledge from the Web to augment their internal information system.

Recently, similar efforts have also been tackled by various projects, including:

- LOD2[83], that provide in particular "Supporting both institutions as well as individuals with tutorials and best practices concerning Linked Data publication and consumption" which could be used for enterprises that wish to integrate data using Semantics.

- The Corporate Semantic Web project[84] that focuses on various facts of improving enterprise work within the enterprise.

In addition to the previous work on integrating legacy data with Semantic Web technologies, some research has extensively done in order to combine Social Web applications and Semantic Web technologies[?] in the enterprise. Applications such as OpenLink DataSpace offers a complete Web 2.0 suite including blogs, wikis and bookmarking systems, all relying on Semantic Web technologies. They are based on some work that will be used, such as the SIOC[85] and MOAT[86] ontologies to provide this integrated architecture. Another related system is Talis Engage[87], a collaborative platform based on different ontologies also including SIOC. In terms of project, relevant work include:

- European Organik project[88] also aims at extending the original Enterprise 2.0 vision[?]. The KiWi project[89] also focuses on enterprise social networking using semantics, and provides use-case in the enterprise for semantic publishing and search.

- Not directly focused to the enterprise the IP FP7 project IKS[90] aims an extending CMS with semantic Web technologies. By working with Drupal, a well-known CMS, and extending it with Semantic Web features, the IdeaWatch project could liaise with IKS to augment the EU-research on Semantic CMS.

- The IP ECOSPACE project also focused on integration of CSCW components using Semantic Web technologies, and used notably SIOC, that will serve as a base model in our context. That way, we can build upon the results of EcoSpace in IdeaWatch.

[83] http://lod2-project.eu/
[84] http://www.corporate-semantic-web.de
[85] http://sioc-project.org
[86] http://moat-project.org
[87] http://talis.com/engage
[88] http://www.organik-project.eu/
[89] http://kiwi-project.eu
[90] http://www.iks-project.eu/

Other relevant projects include ACTIVE[91], COIN[92] and DIGITAL.ME[93].

## 2.5 Conclusions

This section summarizes all the conclusions drawn from the state of the art and set the antecedents for this proposal. First a brief description on general issues related to the processes analysed in this work, Innovation and Technology Watch process, is given. Next, the importance of interoperability is addressed. The work continues with the identification of the ICTs applied in each of the stages of both processes. The level of implementation or usage of these technologies in industry solutions today is outlined next. Finally, areas of further research on the application of the Social Web, Semantic technologies and Semantic Web are outlined.

Innovation is essential for business survival. For innovation to be successful, collaboration among different agents and systems is essential. Among those sytems a well planned Technology Watch process is vital. Sometimes TW acts as the trigger of an innovation process. That is, as the result of a Technology Watch process new ideas arise and the initial phase of the innovation process starts. In other situations, idea validity needs to be assured by contrasting technology, market viability, possible anticipation of the competence (patents), risks and other Technology Watch issues. This contrast or synergy between both processes occurs more than once during the different innovation process stages. TW makes possible anticipating, reducing risks, progressing or cooperating. Both Innovation and Technology Watch processes are iterative. In the research it has been identified that the complexity in managing the innovation process is not trivial and management involvement is imperative.

Sharing data between processes and within the stages of a process is another important issue. The actors for the different stages rely on the quality of content (ideas, publications, market data, ...) to make the appropriate decisions. They need internal data from the own enterprise but also external information. Interoperability among systems, repositories and process becomes essential to assure sufficient information and the quality of it. Without interoperability data is isolated and underused. For example a company may have lots of data but if people proposing new ideas is not aware of it, they will not be able to exploit those data. Interoperability reduces the amount of time spend on swapping from a system to another providing additional relevant information to the user.

In the state of the art it has been proved the importance of using ICT in the Innovation and the Technology Watch processes. Innovation process stages (Idea generation, analysis, enrichment, selection, development and implementation) and Technology Watch

[91] http://www.active-project.eu/
[92] http://www.coin-ip.eu
[93] http://www.digitalme.co.uk/

process stages (planning, information gathering, data analysis, dissemination and feedback or evaluation process) have also been defined.

On the first stage of innovation process, *idea generation*, social collaboration platforms have been proven advantageous to achieve greater participation or collaboration. Sharing ideas among Social platforms guarantees reaching more collaborators. In some cases it opens the process to new participants which translates into greater amount of ideas. Interoperability can be achieved using Semantic Web and Linked Data by relating ideas with the data gathered in Technology Watch or any other information systems or repositories. This could support users work by adding automatically relevant data to ideas.

On the *Idea Analysis* and *Idea Selection* stages of the innovation process, Semantic Technologies could help experts in filtering ideas. NLP and classification tools can assist them finding similarities between ideas or reducing the amount of time spent analysing them. Social tools, such as comments or ratings, can help giving some clues to the experts when choosing which of them will pass the filter.

On the *Idea Enrichment* stage, information from other systems can be used to automatically add relevant data to ideas. Therefore, Semantic Web technologies can be used for interoperability with other information systems.

During the last to Innovation process phases, *Idea Implementation* and *Idea Development*, the interoperability provided by Semantic Web can be useful. It could maintain ideas linked with data collected from the Technology Watch process or available in other repositories (internal or external).

For example it could be linked with patent data, in order to see if during the implementation a new patent related to the idea appears. It can also be related to other ideas in the early stages and see if additional content can be attached to the idea in the implementation stage.

The same technologies can be applied to the TW process. Semantic Web standards will assure interoperability by linking data with content from other systems, platforms or repositories. In order to achieve interoperability, Semantic Web standards can help linking data with other systems. That way, enabling the access to other systems, they can help in the *information gathering* stage of the process. Besides those Semantic Web technologies can help TW process in the *dissemination* stage, helping the systems giving the correct information to the people that need it.

On the other hand, semantic technologies can be useful for the *data analysis* stage and even on the *dissemination* stage too, classifying data and helping experts to group information and give it to the correct people. Moreover, when companies have a great amount of data flow, there is a risk of info-intoxication. The information gathered through TW is very diverse and is extracted from multiple sources. Therefore, there is a need for filtering, classifying and distributing data without saturating the experts, optimizing productive time.

Having described the current state of the technology and its technical application in the Innovation and Technology Watch fields, now is time to describe which is the level of implementation or usage of these technologies in industry today. From the study we extract the following conclusions:

- There is a growing market for tools and platforms to support Innovation and Technology watch.
- Most of the solutions consist of proprietary software, although there are open source platforms with similar characteristics.
- The level of implementation and use of these tools is uneven among companies and depends on management implication.
- Innovation and TW tools and platforms are mainly implemented in large companies. Implementation of these tools in medium and small enterprises is incipient.
- Most of the time those solutions are scattered and concentrate on specific issues of the processes they do not have a holistic vision of the problem.
- TW tools in particular concentrate on specific sources of information and provide solutions for only a specific aspect. Thus, the number of existent different tools is large and there is room for platforms that integrate those solutions.
- All analysed IMSs are idea centred gathering little context information about the process itself.
- Most IMSs understand the importance of collaboration among different agents in order to be successful. Nevertheless, not many of them share ideas with other platforms and in some cases they not even open the process to new participants. Any existent or new innovation support platform must consider the integration of its mechanism with other popular social web platforms.
- There is not a consensus among developers on providing standards that will support integration between platforms. This means that most systems are not interoperable with other systems o solutions.
- Although most of the references in the study understand the relation between the Innovation and the TW processes, there is not a model, tool or platform that supports that integrated vision of both processes.
- In the last couple of years there is a trend to apply IMS solutions to specific domains. That is build solutions to gather ideas oriented to very particular areas (i.e. aeronautics or biotechnology).

- Neither innovation platforms or TW commercial tools employ Semantic Web annotation of content to encourage interoperability.

The state of the art on recent research and previous work from our research team provide answers to some of these issues although some of them remain unanswered and new questions arise. The work presented in this proposal continues previous research on the application of Social Web, Semantic Technologies and Semantic Web to the Innovation process. As a result of this research, issues such as participation and collaboration among agents have been addressed by means of constructing a platform based in Social Web technologies. Details of this platform and can be found in chapter 5 (Work and preliminary results). The selection of open source software to build the testing platforms opened the room for the integration of the Semantic Web to boost interoperability. As has been shown throughout this chapter there are models to semantically represent the innovation process and provide interoperability. There is not a need for creating a new model. GI2MO ontology has been adopted in this work as the most representative model for the innovation process and consequently it will be the starting point for the research to be carried out. Along with the ontology Adam Westerski has contributed in his thesis work with additional semantic solutions and experiments in the innovation context but also has identified room for further experimentation. Among the lines opened for investigation it is worth mentioning the following:

- Usage of the newly discovered idea relationships for idea assessment. The integration with other systems or platforms and the relations obtained as a result could be an area of research.

- Idea annotation with domain ontologies. That is analyze the application of specific domain ontologies over ideas. Ideas will be linked to additional content and consequently richer ideas will be available.

- Further research on automatic idea annotation by means of applying semantic technologies (NLP or AI).

- Improved Enterprise Linked Data evaluation by testing in real industrial scenarios.

Taking into acount prevously mentioned drowback identified by Maynard et al. (*annotation complexity depending on the domain's complexity* and *the need of experts for the construction of specific domain ontologies and taxonomies*), the first one needs to be addressed by means of semantic technologies such as NLP or AI. The second one must be addressed by enterprise experts in association with semantic web specialists. Both cases open the gate to further research.

As a summary it can be stated that there is not a study that analyses the application of the social and semantic technologies in the integration of the Innovation process and

Technology Watch process with a holistic or systemic view. Although there are models that represent the innovation process and specific areas of the TW process, there is not a model that integrates both. The application of the Semantic Web in the Innovation field is incipient and there is still room for further research and improvement in the areas of idea enrichment, idea relations with other ideas and automatic annotation. Semantically linking specific content for a particular domain obtained during the TW process to ideas has not been research yet. Gather context information on the implementation of both processes is important to evaluate successful campaigns. The ontologies identified in both processes gather only little information about the context in which processes occur. Finally, it can be concluded that not only ICTs are needed in order to achieve successful innovation, management implication is much more important. The systematic application of an innovative culture in the company is one of the key factors for success. Proper application of the tools will only be valid if the company establishes the conditions (resources, methodologies...) necessary to encourage that innovation culture.

# 3. Objectives

The goal of this chapter is to describe the objectives and hypothesis of the thesis. Firstly, the objectives will be shown grouped by characteristics to fulfil, management objectives and areas of impact.

## 3.1 Objectives of the Thesis

The main objective of the Thesis is to define a holistic model that integrates collaborative Innovation and TW processes and is supported by the following technologies; Social Web, Semantic Technologies and Semantic Web. The model will be defined, implemented and validated in real scenarios with the collaboration of experts in both processes. A software platform will be developed for the testing these scenarios. The model and the platform have to fulfil the following characteristics:

- They have to be flexible enough to accommodate different types of innovation campaign or waves simultaneously; firm-centric innovation process or crowd sourcing contest with hundreds of participants.
- They have to be flexible to include content from specific domains gathered during the Technology Watch process. The content could be stored in local or remote/external repositories. Automatic classification of content will be desired.
- They have to accommodate to different sources of information.
- They have to guarantee interoperability among processes, stages and data. For example by semantically linking ideas to specific content for a particular domain obtained during the TW process.
- They have to be able to collect context information about the innovation process in its different campaigns. Idea context gathering is essential in order to reproduce successful idea contests.
- The platform has to provide a common space to represent and gather all the information related to innovation processes. This common platform will be used to collect ideas,

identify experts, introduce comments and follow idea progress or search for similar ideas.

- The platform has to offer a collaborative space where employees could develop contacts, share knowledge, improve communication between experts, gather interest in new projects or ideas and enrich ideas.
- The baseline platform has to be deployed on open source software.

The model will enable the measurement of the different factors influencing the process considering the application of the proposed technologies. Thus the objective of the Thesis concentrate in two:

- Analyse the influence factors behind the deployment of the proposed technologies within the innovation and technology watch process.
- Establish the evaluation metrics that will enable a better understanding of the issues related with both.

The expected areas of impact to be analysed with the proposed model are the following:

- Increase collaboration among agents in the process.
- Increase the amount and quality of ideas related with strategic objectives that become projects.
- Reduce management time and time to market of ideas.
- Increase the amount of patents.
- Identify success factors of innovation campaign.

## 3.2 Hypotheses

### Increase collaboration among agents in the process

One of the outcomes of the Thesis achieved by applying the proposed technologies is to increase the number of ideas created by users, encouraging participation. Therefore the following hypotheses have been made:

1. The content dissemination and reporting of real-time changes will increase the number of ideas created by users: this hypothesis is based on the real-time dissemination of changes in the innovation process can attract the attention of users and encourage them to create new ideas.

2. Comments on existing ideas will increase the number of ideas created by users: the different comments on the ideas can trigger new thoughts on users helping them creating new ideas.

### Increase the amount and quality of ideas related with strategic objectives that become projects

Another expected outcome is to obtain an increase in the number of ideas that pass the filters of the innovation process (successful ideas), that is, ideas that turn into projects. These are the hypotheses:

3. The comments and context of the ideas can make more ideas end up becoming into projects: the comments on the ideas can add relevant information and make better an misconceived or poor idea, so it could go more easily through the filters and achieve better rating on selection criteria.

4. Linking data with internal/external repositories and legacy systems, such as Technology Watch process, will help generating richer ideas, automatically showing additional relevant information to the ideas. This way more ideas will become into projects.

### Reduce management time and time to market

The management time of the innovation process is often very large, and especially when there are many ideas. It is therefore important to reduce management time for experts to focus on more productive tasks. The hypotheses are:

5. Gather context information about the processes will allow campaign managers repeating successful waves, avoiding wasting time when defining new campaigns.

6. Identifying mentioned entities will help finding similarities between ideas. This will reduce the amount of repeated ideas and will help merging similar ideas, reducing the administrative time and generating richer ideas.

7. User ratings will help managers identify the best ideas, reducing the time wasted on such identification.

### Increase the amount of patents

Next expected outcome for the thesis is the increasing of the amount of generated patents by the innovation processes. In order to perform this task, two hypothesis have been defined:

8. The automatic content gathering will bring an increase on the amount of new patents: the automatic information gathering will add relevant information to the ideas, generating more innovative and patentable ideas.

9. Technology Watch will reduce time spent on patent analysis: it will help to discover the already existing patents and ideas, in order not to waste time with non-patentable ideas. Therefore, more time will spend working with other ideas, augmenting their chance to be patented.

## Identify success factors of innovation campaign

The last expected outcome or result for the thesis is to identify success factors that can make an Innovation Process thriving. Thus, this hypothesis has been defined:

10. Gathering context data will help Innovation process managing in order to identify success factors. This way, new idea contests can be launched repeating these factors that made the contest successful.

# 4. Methodology and Planning

This chapter summarizes how the thesis and research project are planned to be implemented. Firstly, the methodology approach will be described. Secondly, the proposed metrics in order to identify innovation success factors are enumerated. Next, the work schedule for the research project and thesis is shown. Finally, the tools used for the research development are described.

## 4.1 Methodology

The methodology approach followed in this research project is based on an incremental development cycle, where requirements guide implementation. The experience collected in a cycle will help to improve the next one. The results and conclusions obtained with the developed prototypes can generate new requirements. The methodology life cycle can be seen in figure 4.1.

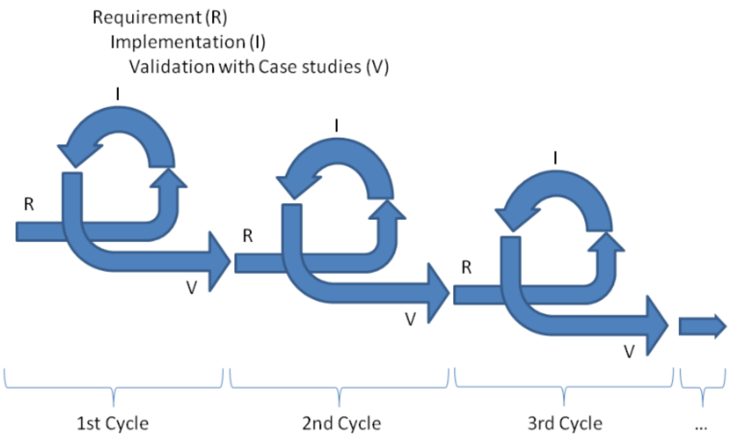


Figure 4.1: Evolutionary life cycle methodology.

Next, the phases that summarize each cycle in the development are presented.

1. **Requirements:** a field study is performed through a set of interviews to different representatives of the involved organizations in order to assess the use of Social Web

technologies. This input, together with the state of the practice research, is used to depict the case studies.

2. **Implementation:** the necessary prototypes and the methodology are developed. The methodology will provide a stepwise approach for the adoption of the prototypes within an organization.

3. **Validation:** a set of piloting activities are carried out within real production scenarios and are based on the depicted case studies. A set of indicators are set up in order to evaluate the result and the success of the contest.

## 4.2 Metrics

In order to identify innovation success factors, metrics have to be defined. Environment conditions or context information is an important issue that needs to be addressed. Data and metrics about the innovation process, innovation campaigns (*waves*), the activity and outcome has been identified and classified according to the following criteria.

### General Wave Characteristics

Enterprises usually launch innovation contests (or *waves*) in their innovation processes. Wave information shows the framework environment on which ideas are generated.

- *Innovation type:* stores the level of innovation of the wave (radical, incremental...).
- *Stages:* stages the wave will follow on the Idea Management life cycle.
- *Status:* describes the current stage.
- *Target:* indicates the objective or target aimed by a wave.
- *Topic:* how the wave has been classified.
- *Contest type:* indicates the type of innovation searched by the wave (*offer, platform, solution, client, client experience, value capture, process, organization, supply chain, presence, network* or *brand*).
- *Fields of the idea:* fields the user will have to fill in order to submit an idea.
- *Duration:* time the wave will last.
- *Situation:* environment in which the wave is created (relaxed, time or condition pressure...).
- *Selection criteria:* indicates the criteria used for idea ratings (set by experts).

### Structural Metrics

These metrics measure the structural properties of the contest and their impact.

- *People:* groups, dedication, number of participants, active users, roles...
- *Enterprise:* time and resources assigned to R+D+i by companies.
- *Resources:* locations (e.g. meeting rooms), amount of resources spent on awards and prices...

### Activity Metrics

These metrics measure the activity in the platform.

- *Traffic measures:* number of views, unique visitants, average time spent, repeat visitors...

### Stimulation Metrics

These metrics register actions or stimuli provided in the wave to boost participation and improve quality of ideas.

- *Events:* the number of events, participants, type, duration, location...
- *Awards:* the number of awards and the amount of money earned in each...

### Outcome Metrics

The outcome metrics measure the result of the contests.

- *Wave level:* number of ideas that fit the target, become a product, create a spin off...
- *Idea level:* innovative ideas, innovation level, number of innovations introduced, generated sales...

### Quality metrics

Quality metrics measure the quality of the ideas using the next criteria:

- *Innovation:* How different the idea is from the ones existing in the market, covers the needs of that market and strategically fits within the company.
- *Degree of convergence:* Degree of convergence and leverage business proposal with technical capabilities, technical expertise and business management.
- *Economical Benefits:* How much benefit can be obtained with the idea.

- *Cost-effort:* the amount of cost and effort required to develop the idea.
- *Technical and business feasibility:* how feasible it is to develop the idea, taking in to account the technical level of the enterprise and how easy it is to introduce the idea in the market.

#### Performance metrics

This metrics measure the amount of time saved as a consequence of using developed tools:

- *Users time saved:* The amount of reduced time for the users, comparing the same task using the platform and without it.
- *Time to market reduction:* The amount of reduced for an idea since it is created till it is developed, comparing the same task using the platform and without it.

## 4.3 Work Schedule

This section makes a brief description of the work schedule for the four years of the thesis. Firstly, the work schedule for the past two years will be shown (2012 schedule in table 4.1 and 2013 schedule in table 4.2) and next future two years planned work schedule (2014 schedule in table 4.3 and 2015 schedule in table 4.4).

### 4.3.1 Description of the tasks

This subsection describes the tasks of the thesis enumerated in the work schedule tables, grouped by years.

#### 2012 (table 4.1)

1. *Analysis of Innovation and Technology Watch process:* this task's objective was to analyse the state of the art for Innovation and Technology Watch processes, and what were the existing technologies and applications that could help them. As a consequence, chapter 2 was developed.
2. *Development of the first prototype:* this task was to develop a first prototype platform for testing the social web impact on the first states of the innovation process. Therefore, a platform named *Innoweb* was developed using an open source CMS called *Drupal*.
3. *Testing of the prototype:* this task was to test the platform created in the previous task. For this purpose, the team tested the platform in different idea generation scenarios within the university and collaborator enterprises.

4. *Result dissemination:* This task was to disseminate the results to the community. As a result, 2 papers were published (see section 5.3.1 and section 5.3.2).

| 2012 | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.- Analysis of Innovation and Technology Watch processes. | | | | | | | | | | | | |
| 2.- Development of the first prototype | | | | | | | | | | | | |
| 3.- Testing of the prototipe. | | | | | | | | | | | | |
| 4.- Result dissemination. | | | | | | | | | | | | |

Table 4.1: Time-line of 2012. First year of the thesis.

### 2013 (table 4.2)

5. *Development of the second prototype:* this task was to develop a second prototype platform for testing semantic web capabilities for the first stages of the innovation process. *Innoweb* platform was taken as a base.

6. *Internship on abroad university:* in order to contact with experts on Semantic Web technology an internship on abroad university was planned. Thus, three months were spent collaborating on DERI (see section 5.2.1) in order to analyse semantic capabilities.

7. *Result dissemination:* This task was to disseminate the results obtained in the previous ones to the community. As a result, a paper was published (see section 5.3.3).

8. *Composition of the Researching Project document:* this task was to develop a document in order to present the Research Project progress and future plans and to obtain the Master of Advanced Studies. Thus, this document was written.

| 2013 | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5.- Development of the second prototype. | | | | | | | | | | | | |
| 6.- Internship on abroad university. | | | | | | | | | | | | |
| 7.- Result dissemination. | | | | | | | | | | | | |
| 8.- Composition of the Researching Project document. | | | | | | | | | | | | |

Table 4.2: Time-line of 2013. Second year of the thesis.

**2014 (table 4.3)**

9. *Development of the third prototype:* this task aims to develop a third prototype, based on *Innoweb* that integrates data from different systems and repositories, linking both Innovation and Technology Watch processes.

10. *Testing of the prototype:* This task aims to test the third prototype on internal and external innovation processes (within the university and collaborator companies) in order to check the hypothesis of the thesis.

11. *Result dissemination:* this task's objective is to share the result of the third prototype with the community, trying to publish them in a conference or a journal.

| 2014 | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 9.- Development of the third prototype | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | | | | |
| 10.- Testing of the prototype. | | | | | | ■ | ■ | | ■ | | | |
| 11.- Result dissemination. | | | | | | | | | | ■ | ■ | ■ |

Table 4.3: Time-line of 2014. Third year of the thesis.

**2015 (table 4.4)**

12. *Elaboration of the PhD dissertation:* This task aims to elaborate the document to obtain the PhD. That document should collect all the work done within the four years of the research and afterwards it should be presented.

13. *Result dissemination:* this task's objective is to share the result of the whole research project with the community, trying to publish them in a conference or a journal.

| 2015 | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 12.- Elaboration of the PhD dissertation. | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | ■ |
| 13.- Result dissemination. | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | ■ |

Table 4.4: Time-line of 2015. Fourth year of the thesis.

## 4.4 Used tools.

This section aims to summarize the tools used so far in order to implement the thesis. Drupal CMS has been use in order to exploit Social Web functionalities and as a base for the platform.

There are two tools to implement semantic technologies, Freeling that uses NLP in order to identify entities and nouns; and GNU Octave, that is used for implementing a SVM in order to classify documents from the Technology Watch process automatically.

Two tools have been used to exploit Semantic Web capabilities. Protégé has been use in order to analyse the existing ontologies and build GI2MOWave; and Virtuoso to create a local dataspace in order to store data in semantic format and analyse the ability of linking.

Below, those tools and their functionalities are shown.

### Drupal[1]

Drupal is an open source Content Management System. It offers the possibility of extending functionalities using existing community modules or developing custom ones. It is build on *PHP* and uses relational DataBases in order to store information. Many community efforts are made for adding Semantic Web capabilities. Drupal 7 for example, comes with native RDFa support in core, and there are some contributed projects in order to add more of those capabilities:

- *SPARQL*[2]*:* This is a module that enables the use of SPARQL queries with the RDF API.
- *SPARQL Views*[3]*:* This allows using Views (a Drupal functionality to create pages for displaying data) to access remote and local SPARQL endpoints. In the near future, it will also work for RDFa on Web pages and hopefully microdata.
- *RDFme*[4]: RDFme is a Drupal extension that allows to publish RDF metadata attached to regular Drupal HTML pages.

### Freeling[5]

Freeling is an open source language analysis tool suite. It is a library that provides language analysis services, but it can also be used as a program itself. Currently supported languages are Spanish, Catalan, French, Galician, Italian, English, Russian, Portuguese, Welsh and Asturian. Czech and Slovenian have partial support.

Main services offered by FreeLing are:

- Text tokenization.
- Sentence splitting.

[1] https://drupal.org/
[2] https://drupal.org/project/sparql
[3] https://drupal.org/project/sparql_views
[4] http://www.gi2mo.org/apps/drupal-rdfme-plugin/
[5] http://nlp.lsi.upc.edu/freeling/

- Morphological analysis.
- Suffix treatment, retokenization of clitic pronouns.
- Flexible multiword recognition.
- Contraction splitting.
- Probabilistic prediction of unkown word categories.
- Phonetic encoding.
- SED-based search for similar words in dictionary.
- Named entity detection.
- Recognition of dates, numbers, ratios, currency, and physical magnitudes (speed, weight, temperature, density, etc.).
- PoS tagging.
- Chart-based shallow parsing.
- Named entity classification.
- WordNet-based sense annotation and disambiguation.
- Rule-based dependency parsing.
- Nominal correference resolution.

FreeLing has been used for extracting word roots from documents in order to classify them using a SVM tool. It has been also used to identify nouns from the ideas and find their URIs on internal and external dataspaces.

Drupal has been used as a base for the Innovation process supporting platform (*Innoweb*).

### GNU Octave[6]

GNU Octave is an open source high-level interpreted language, primarily intended for numerical computations. It provides capabilities for the numerical solution of linear and nonlinear problems, and for performing other numerical experiments. It also provides graphical tools for data visualization and manipulation. The Octave language is quite similar to Matlab so that most programs are easily portable.

GNU Octave has been used in order to create document classification programs using SVMs.

[6] http://www.gnu.org/software/octave/

### Protégé[7]

Protégé is a free, open source ontology editor and knowledge-base framework.

The Protégé platform supports modeling ontologies via a web client or a desktop client. Ontologies can be developed in a variety of formats including OWL, RDF(S), and XML Schema.

Protégé is based on Java, is extensible, and provides a plug-and-play environment that makes it a flexible base for rapid prototyping and application development.

This tool has been used in order to analyse the existing ontologies and to add more fields to GI2MO ontology creating GI2MOWave.

### OpenLink Virtuoso[8]

Virtuoso is an open source multi-model data server. It's hybrid server architecture enables offering different type of functionalities, such as:

- Relational Data Management.
- RDF Data Management.
- XML Data Management.
- Free Text Content Management & Full Text Indexing.
- Document Web Server.
- Linked Data Server.
- Web Application Server.
- Web Services Deployment (SOAP or REST).

Virtuoso has been used in order to work as a repository and link it with the Innoweb platform. The stored information has been liked with the platform using the URIs identified by the *IdeaMentions* module developed over Drupal.

[7] http://protege.stanford.edu/
[8] http://virtuoso.openlinksw.com/

# 5. Work and preliminary results

This chapter collects the work done to date of writing this document, and preliminary results which may serve to indicate the validity and feasibility of the doctoral thesis. It will describe the development of the first three cycles (see methodology on chapter 4) of the platform, the collaboration with universities and companies, scientific publications, projects involved during the research for founding and the findings in performed case studies.

## 5.1 Platform

Following the previously mentioned methodology (see section 4.1) a prototype platform, called *Innoweb*, has been developed for testing Innovation and Technology Watch success factors. That platform uses *Drupal* CMS as base. Functionality can be added to *Drupal* using existing community modules or building new ones. Some modules have been developed in order to build *Innoweb* platform. The objective of the platform is to create an holistic model that integrates Innovation and Technology Watch processes and enables interoperability between both internal and external information systems (see figure 5.1).

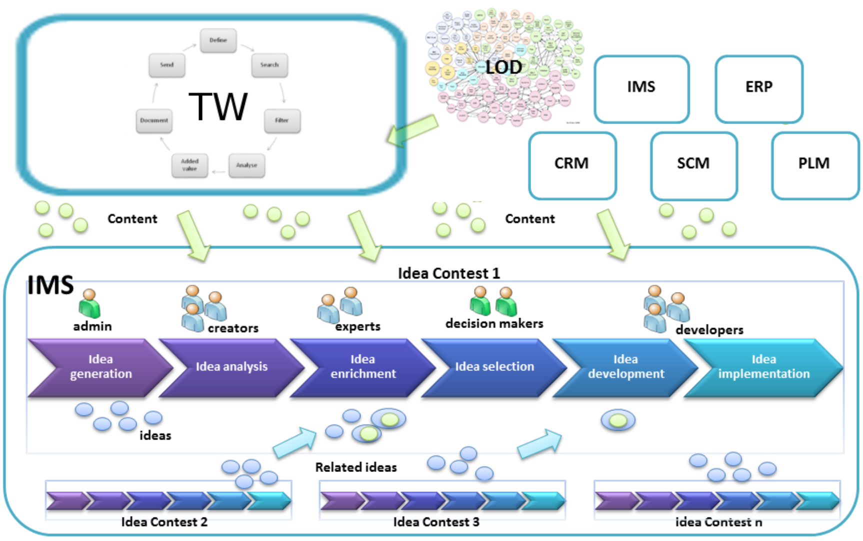


Figure 5.1: Holistic model for Innovation and Technology Watch processes and system interoperability.

Below, those modules are described grouped by the cycle where they were developed.

## First prototype cycle

The first cycle of the prototype was centred on developing a platform to cover the first stages of the innovation process (see figure 5.2) with social functionalities. The objective of this cycle was to enhance collaboration between Innovation process agents. Below, the modules developed in this life-cycle are described, and next, the community modules used in order to create the platform.


Figure 5.2: Early stages of the innovation process

- *Idea:* This module enables the generation of ideas in a blog format (see figure 5.3). Depending on the parameters chosen on the wave that owns the ideas, some fields are asked to fill to the users. It also provides idea visualization in a list (see figure 5.4). That way, it enables the first stage of the innovation process, the *Idea Generation*.

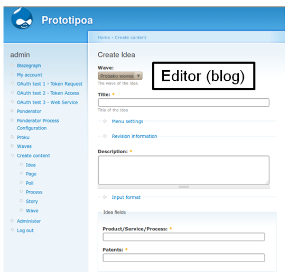


Figure 5.3: Idea generation in blog format.

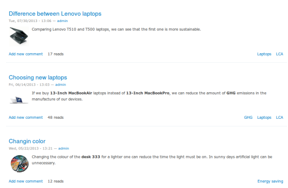

Figure 5.4: Idea list.

- *Idea2wiki:* This module performs the second stage of the innovation process, the *Idea Analysis*. Here, the first filter is done. From all the ideas generated in the first stage, some experts select the ones that are going to be enriched (see figure 5.5) and the other ones are discarded but stored, in case they can be exploited in the future.

- *Ideawiki:* This module enables editing the ideas in wiki format so expert can enrich them (see figure 5.6). This module performs the third stage of the innovation process, the *Idea Enrichment* state.

- *Ponderator:* This module enables the possibility of rating ideas following some criteria. From those criteria and different rating methods the ideas are rated by experts and the ones that are going to become into project are selected (see figure 5.7).

Figure 5.5: Idea analysis implementation with idea2wiki module.

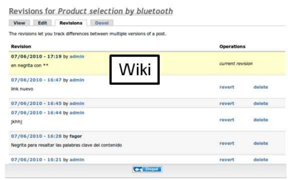


Figure 5.6: Wiki format for Idea enrichment.

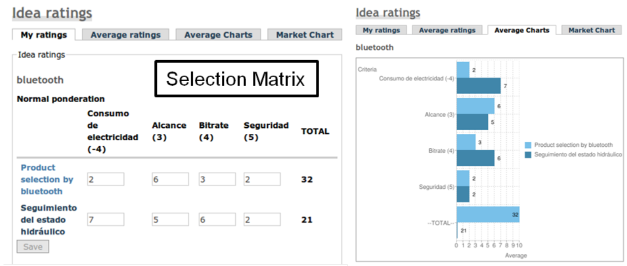


Figure 5.7: Ponderator module for Idea Selection.

- *Ponderator Chart:* This module uses Chart API[1] in order to visualize the expert ratings of the ideas from *Ponderator* module (see figure 5.8).

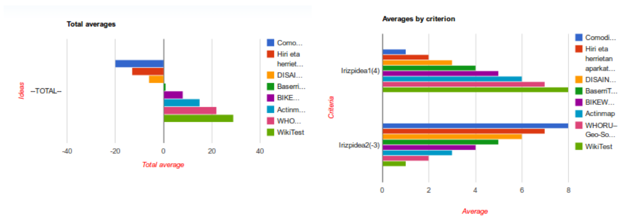


Figure 5.8: Idea Selection rating visualization example from *Ponderator Chart* module.

There are some community modules that have been used in order to add functionalities to the platform. Those modules were not developed by the team, but are configured in order to work along with developed modules. Below those modules are described:

- *VotingAPI:* This module enables functionalities to perform ratings in nodes and comments.
- *Vote Up Down:* VoteUpDown adds a rating interface in order to use VotingAPI. This way, the users will have some widgets to perform the ratings.
- *CCK:* This module adds different functionalities to add fields to any node inside Drupal.
- *Workflow:* Workflow module allows Drupal nodes to have different life-cycles. This way, nodes can have some states that change depending on the situation.

[1] https://drupal.org/project/chart

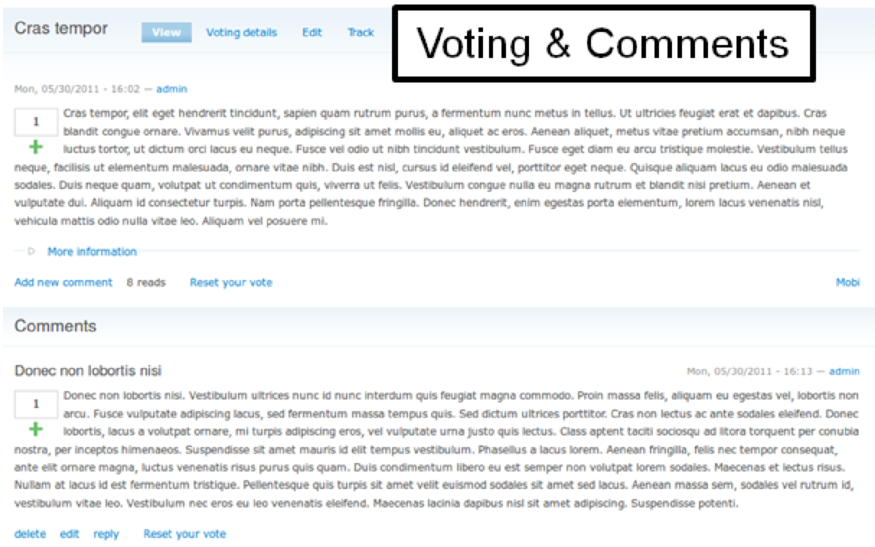


Figure 5.9: Voting widgets in the ideas.

For example, it is very useful in order to manage the ideas life-cycle through the innovation process.

## Second prototype cycle

This second cycle of the prototype was centred on enabling the possibility of creating multiple innovation processes at once in the same platform (see figure 5.10). Those processes were called *Waves*, and each of them has an independent workflow management and can gather ideas with different fields. Below, the modules developed for this life-cycles are described:

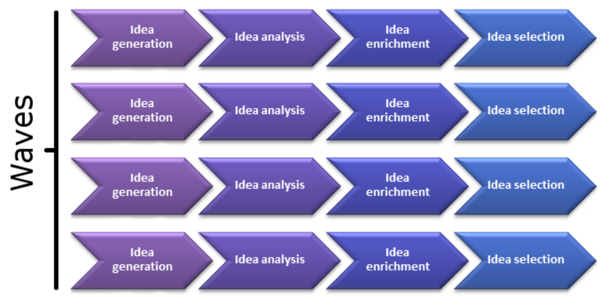


Figure 5.10: Multiple innovation processes (*Waves*) at once.

- *Blazegraph API:* Blazegraph[2] is a dynamic graph layout engine implemented in Flash and ActionScript. Developed Blazegraph API module enables using this tool for data visualization (see figure 5.11).

[2] http://blazegraph.sourceforge.net/


Figure 5.11: Blazegraph visualization tool.

- *JSchart API:* JSChart API module enables using dinamic charts from Google Charts[3], such as Scratter Charts or Bar Charts.

- *Wave:* This module enables generating different Idea contests (called waves) and gathering their context information. The module enables the possibility of controlling each of the contest independently. It allows controlling the stages the contest will have, the permissions of the users in each of the stages, the possibility of choosing what fields will the idea have, duplicating an existing wave...

- *Wave Event:* Wave Event module stores events happened within a Wave, so their impact can be analysed. A visual example can be seen on figure 5.12.

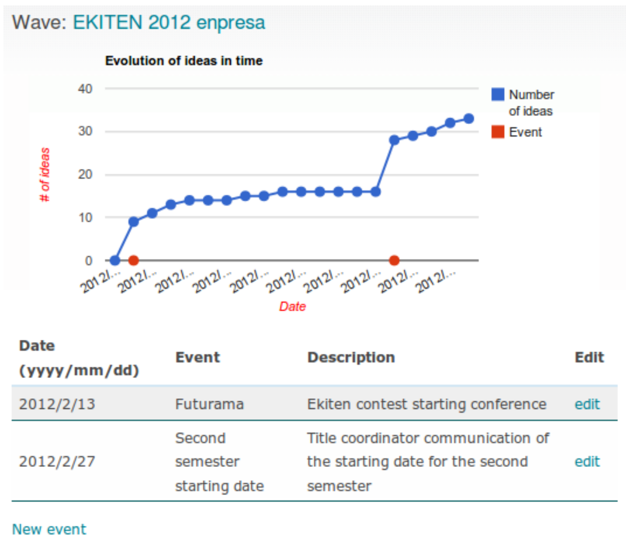


Figure 5.12: Wave Event visualization example.

- *WaveRT:* WaveRT enables the possibility of real time notifications (see figure 5.13) of

[3] https://developers.google.com/chart/

different actions. For example, it can notify to a user that one of his ideas has been promoted.

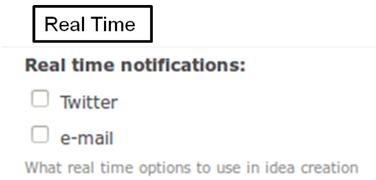


Figure 5.13: Real time notification selection.

- *WaveBG:* WaveBG uses Blazegraph API to show relations between ideas (see figure 5.11), making easier for the administrators the task of analysing ideas.
- *WaveChart:* This module uses JSChart API to show Wave data visualizations. A graphical example can be found on figure 5.14. This makes easier the wave administration tasks.

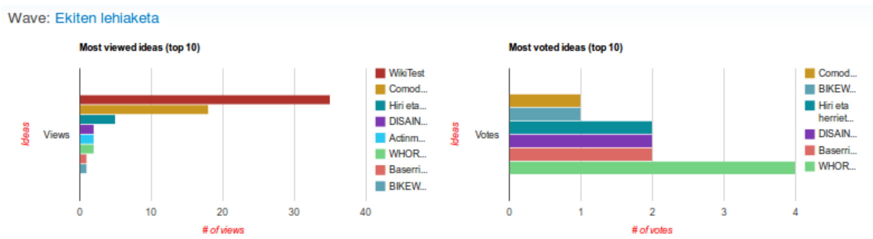


Figure 5.14: Wave Chart visualization example, showing idea related data.

- *Ponderator Chart 2.0:* This module uses JSChart API in order to visualize the expert ratings of the ideas from *Ponderator* module. This evolution enables interaction with the chart in order to visualize the data in a better way.

## Third prototype cycle

This third cycle of the prototype was centred on enabling the interoperability of the Innovation process platform with external and internal repositories (see figure 5.15). This would help exploiting the data from other systems, such as TW, adding it to the Innovation process.

Below, the modules developed for this life-cycles are described:

- *FreeLing API:* FreeLing API module makes possible integrating FreeLing NLP libraries with Drupal. This allows other modules to use those NLP services.

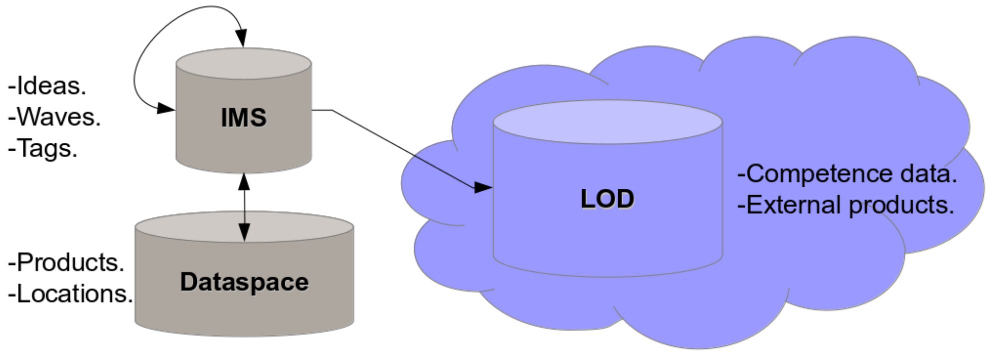


Figure 5.15: Linking IMS with external and internal repositories.

- *IdeaMentions:* This module enables identifying the URIs of the elements mentioned in the ideas. That way, data can be automatically added from external (DBpedia[4] for example) or internal repositories (using dataspaces such as Virtuoso). It also helps identifying similar ideas, comparing the mentioned elements each of them have. An example of this can be found on figure 5.16, where an idea can be found. A table can be seen on the left with idea similarity recognition and another one on the right with automatically added information (showing the amount of different materials for each mentioned element). In order to find those mentioned elements FreeLink API module is used, extracting the proper nouns found on the text and searching them on other repositories.

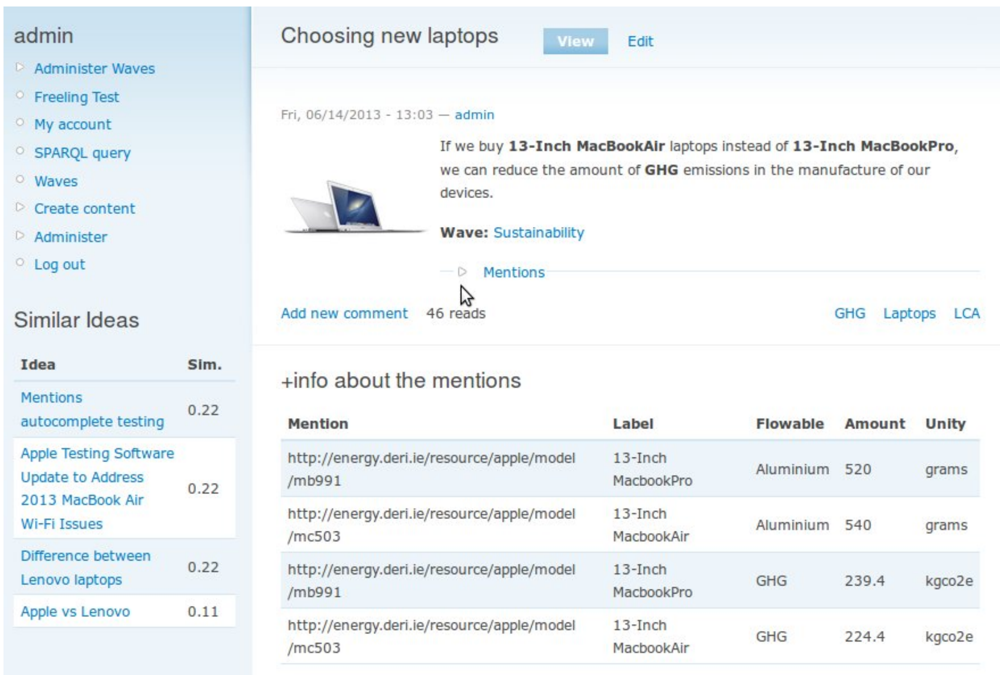


Figure 5.16: Example of automatically added information from IdeaMentions module.

[4] http://dbpedia.org/

- *IMS2RDF:* This module enables sharing in RDF format different resources, such as waves, ideas, events, tags... An example idea in RDF format can be seen on figure 5.17. The resources are described using GI2MOWave ontology created in collaboration with UPM (see subsection 5.2.2).

```
        <gi2mo:Idea rdf:about='http://localhost/innowebd6/data/idea/14'>
                <foaf:page rdf:resource='http://localhost/innowebd6/node/14'/>
                <gi2mo:hasCreator rdf:resource='http://localhost/innowebd6/data/user/1'/>
                <gi2mo:hasIdeaContest rdf:resource='http://localhost/innowebd6/data/ideacontest/10'/>
                <gi2mo:hasCategory rdf:resource='http://localhost/innowebd6/data/category/1'/>
                <gi2mo:hasTag rdf:resource='http://localhost/innowebd6/data/tag/8'/>
                <gi2mo:hasTag rdf:resource='http://localhost/innowebd6/data/tag/7'/>
                <gi2mo:hasTag rdf:resource='http://localhost/innowebd6/data/tag/6'/>
                <gi2mowave:mentions rdf:resource='http://dbpedia.org/resource/GHG'/>
                <gi2mowave:mentions rdf:resource='http://dbpedia.org/resource/MacBook_Air'/>
                <gi2mowave:mentions rdf:resource='http://dbpedia.org/resource/MacBook_Pro'/>
                <gi2mowave:mentions rdf:resource='http://energy.deri.ie/resource/apple/lca/footprint/ghg'/>
                <gi2mowave:mentions rdf:resource='http://energy.deri.ie/resource/apple/model/mb991'/>
                <gi2mowave:mentions rdf:resource='http://energy.deri.ie/resource/apple/model/mc503'/>
                <gi2mo:created>2013-06-14T14:03:38Z</gi2mo:created>
                <gi2mo:title>Choosing new laptops</gi2mo:title>
                <gi2mo:content>If we buy &lt;b&gt;13-Inch MacBookAir&lt;/b&gt; laptops instead of &lt;b&gt;
b&gt;, we can reduce the amount of &lt;b&gt;GHG&lt;/b&gt; emissions in the manufacture of our devices.
</gi2mo:content>
        </gi2mo:Idea>
```

Figure 5.17: An example idea on RDF format.

Some community modules have also been used in this cycle, in order to give the main Semantic Web capabilities to the platform. All three modules have been used in order to offer an endpoint along with the previously mentioned *IMS2RDF* module.

- *RDF:* this API makes use of the ARC2 library if available, and will integrate RDF capabilities with other modules.

- *SPARQL:* This is a module that enables the use of SPARQL queries with the RDF API.

- *RDFme*[5]*:* This is a Drupal extension that allows to publish RDF metadata attached to regular Drupal HTML pages. It also enables the possibility of publishing an SPARQL endpoint with any RDF data.

## 5.2 Cross university and company collaboration

During the investigation some collaborations with universities and companies have been made. In the following subsections those collaborations will be described and the works done in each of them. First, collaboration with some universities will be presented: DERI, from National University of Ireland; Universidad Politécnica de Madrid. Next, collaboration with some enterprises will be presented: ISEA and Koniker. Finally, the collaboration within the different parts of Mondragon Unibertsitatea itself will be presented.

[5] http://www.gi2mo.org/apps/drupal-rdfme-plugin/

### 5.2.1 Collaboration with Digital Enterprise Research Institute (DERI) of the National University of Ireland (NUI)

DERI[6] is a Centre for Science, Engineering and Technology (CSET) established in 2003 with funding from the Science Foundation Ireland. It has become an internationally recognised institute in semantic web research, education and technology transfer. Therefore, and as DERI's researches are focused on the Semantic Web, it was choose as a good option in order to make a collaboration.

Tree months were spend in a collaborating project that linked some structured data spaces about sustainability with the innovation process. Thus, *IdeaMentions* module for *Innoweb* and some use cases were developed. As a result, a paper was published with the work done in DERI (see publication in subsection 5.3.3).

### 5.2.2 Collaboration with Universidad Politécnica de Madrid (UPM)

Universidad Politécnica de Madrid (UPM[7]) is a University founded in 1971 as the result of merging different Technical Schools of Engineering and Architecture, originated mainly in the 18th century. In the state of the art of this project, some works were found on Innovation process Ontologies involving UPM researching teams. Therefore, some contacts were done to collaborate in this researching areas.

As a result a new version of GI2MO ontology was proposed, GI2MOWave[8] (see Figure 5.18), more focused on the context of the ideas. The ontology has been tested and will be tested in future campaigns to collect context information about innovation processes.

### 5.2.3 ISEA

Innovation in Advanced Business Services - ISEA S. COOP.[9] is a private and non-profit innovation and entrepreneurship Centre, specialized in Business Services Sector, promoted by the Division of Engineering, and Business Services of MONDRAGON Corporation.

ISEA S.COOP. is a Science and Technology Agent, integrated into the Basque Science, Technology and Innovation Network. Today is part of the Basque Innovation Agency - INNOBASQUE. Additionally, ISEA is a Certified Agent of the Basque Entrepreneurial Service of the Society for Competitive Transformation - SPRI.

In line with its corporate purpose, ISEA S. COOP. has promoted a Business Acceleration Center (BAC), a specialized structure designed to boost the process of launching new business initiatives in the Advanced Services Sector. In this context, ISEA used the developed

[6] http://www.deri.ie/
[7] http://www.upm.es/internacional
[8] http://www.gi2mo.org/wave/0.1/ns.html
[9] http://www.iseamcc.net/isea

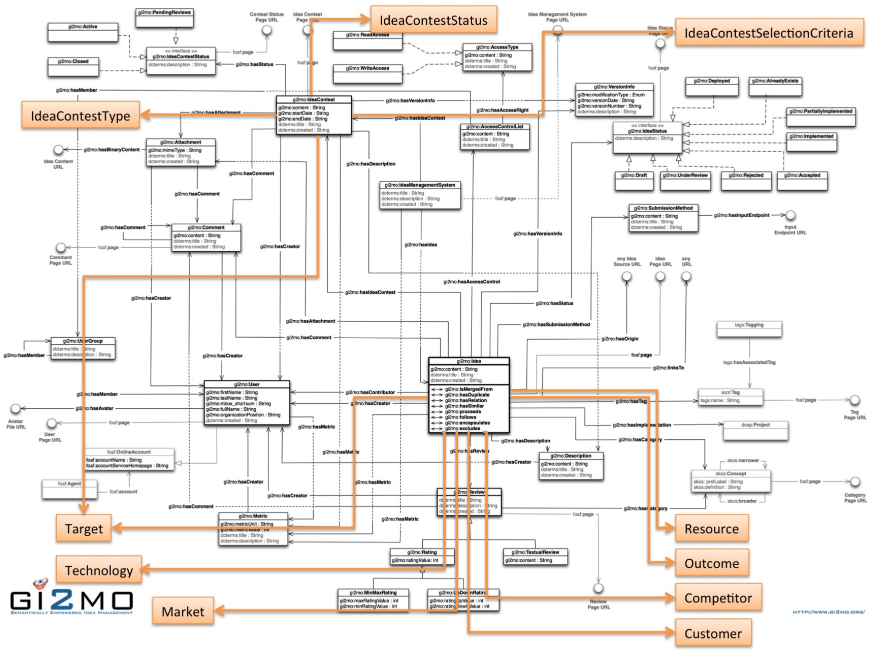


Figure 5.18: UML Class diagram for GI2MO Wave Ontology (bsed on Gi2MO Ontology diagram)

platform to launch campaigns and collect ideas with an open innovation philosophy. The project is called *Elkarbide*[10].

### 5.2.4 Koniker

Koniker S.COOP.[11] is a non-profit Technology Centre of public utility, specialising in the research and development of new technologies related to forming and assembly processes.

Koniker provides services for the following companies: Fagor Arrasate, Batz, Mondragon Assembly, Onapres, Aurrenak, Matrici, Loramendi and Mondragon Corporation.

It was established in 2002 with the aim of giving a coordinated response to the R&D projects of all of these companies and of making the most of synergies present in the forming, assembly, machinery and tool areas. They have adopted Innoweb in a platform called *Ideak*, in order to collect ideas in the companies they work for. Using the findings of this case study two publications were made (see publications on section 5.3.1 and 5.3.2).

One of the main services provided by Koniker is consultancy on Technology Watch issues. Koniker possesses an important catalog or repository with data gather by means of TW actions. They have identified their TW workflow and the amount of time spent on each task (see figure 5.19). They consider that many of the tasks are unproductive and could be automatize by using ICTs. They consider that up to 55% of that time could be reduced.

[10] http://www.elkarbide.com/
[11] http://www.koniker.coop/en

Figure 5.19: Current Technology Watch workflow and time spent in the steps.

### 5.2.5 Mondragon Unibertsitatea

Innoweb has also been used inside of the University, creating an idea gathering platform called *Ekiten*. The main objectives of EKITEN are two:

1. The development of entrepreneurial spirit and ability of students.

2. The launch of new business initiatives.

In Ekiten, multidisciplinary teams have been made. Those teams are made of students from different faculties of the University with different profiles working together and contributing with their own specific job profile. Also, all the equipment and therefore all projects will have a tutor who will guide the students and, moreover, will also feature the expertise of University faculty in the areas involved in the project .

This project is a collaboration between Mondragon Unibertsitatea and the Center for Business Innovation Saiolan, and has the support of the Department of Innovation and Information Society of the Provincial Council of Gipuzkoa and the Department of Industry of the Basque Government.

## 5.3 Scientific Publications

Those collaborations have been performed, creating some tools and achieving some results for the research. Those tools and results have been gathered and three papers have been published so far. This section will describe those scientific publications accepted by the community. Firstly, a publication of the base platform to support an innovation project is presented. Next, a publication of the prototype adding semantic capabilities and context information gathering is described. Finally, the publication where the work done in the abroad university internship is presented.

### 5.3.1 A case study on the use of community platforms for inter-enterprise innovation (ICE2011)

This paper is enshrined in a larger research project which main goal is to determine key factors in the design, implementation and use of collaborative environments for the management of inter-enterprise innovation processes based on practical experiences. The paper presents an innovative approach to address the challenge of collaboration and participation in the submission of new ideas in the front end of the innovation process. The paper describes a case study on the adoption of an open source community platform based on Drupal in the context of a set of cooperative companies within Mondragon Corporation. The platform leverages social computing, real-time Web and semantic technologies to support collaboration, basic technology watch and idea management in the early stages of the innovation process. Preliminary field results show that the platform provides a powerful collaborative tool that eases administration work and enhances collaboration among participants.

### 5.3.2 INNOWEB: Gathering the context information of innovation processes with a collaborative social network platform (ICE2013)

This paper describes the development of a collaborative social platform to support innovation process management. The Drupal based platform accommodates different types of innovation processes (also called waves or idea contests), enhances collaboration and eases management. The main contribution lies on the gathering of context parameters which helps enterprises on the detection of critical success factors, enabling later reproductions. The development leverages open source social computing, real-time web and semantic web technologies adding new functionality in a modular way. Blogs, wikis, graphical tools and voting systems support collaboration and idea management in the early stages of the innovation process. A workflow module launches customized innovation campaigns where topics, participants, stages, selection criteria and communication methods are optimized to enterprise needs.

### 5.3.3 The Role of Linked Data and Semantic-Technologies for Sustainability Idea Management (MoKMaSD2013)

Idea Management Systems (IMS) manage the innovation lifecycle from the moment of invention until ideas are implemented in the market. During the lifecycle the IMS supports collaboration, allows idea enrichment with comments, contextual data, or connected to other relevant ideas. Semantic technologies can improve the knowledge management capabilities of IMSs allowing relevant information to be easily linked to ideas. Many Enterprises concerned with sustainability encourage employee's participation as a means to boost creative innovation within their Sustainability Initiatives. However little work has examined the role of an IMS within Sustainability. In this paper we analyse the impact of a semantic-enabled IMS within a sustainability innovation process. In particular, how ideas can be enriched with contextual Linked Open Data (LOD), especially Life-Cycle Assessment (LCA) data, to improve the understanding, implication and value of the idea from the sustainability perspective.

## 5.4 Projects

The team has been involved in many projects during the research for founding. Those project were and are being developed collaborating with some universities and enterprises described in section 5.2. Some collaboration have not been in direct contact, but with some of previously mentioned collaborators as intermediary. Below, those project are listed.

- **SELENE**: *Social and Semantic Web Environment for Network Innovation*, National Applied Research Program (Madrid) along with LKS, ISEA and Deusto Unibersity (2009-2011).

- **COLABORANOVA**: *Collaborative Innovation System Based on Participation*, Provincial Council of Gipuzkoa (2010).

- **ELKARWEB**: *Social and Semantic Web Platform to Support Collaborative Innovation*, GAITEK program from Industry Department of the Basque Government with KONIKER and DANOBAT (2011).

- **PATENTAWARE**: *Support System for Strategic Decision and Patentability in the Field of Home Energy*, GAITEK program from Industry Department of the Basque Government with LKS INGENIERÍA (2012-2014).

- **INNES (Strategic Innovation)**: *Smart Platform for Strategic Innovation in the Field of Health*, GAITEK program from Industry Department of the Basque Government with AURRENAK, LORAMENDI and PROSPEKTIKER (2013-2015).

- **K-INNES**: *Smart Platform for Strategic Innovation in the Field of Health*, Servicing for KONIKER.

- **ACCELERATE**: *A Platform for the Acceleration of GO-TO Market in the ICT Industry.* AEESD (Acción Estratégica Economía y Sociedad Digital) program with PLANET MEDIA STUDIO, SIVSA SOLUCIONES INFORMATICAS, S.A. and FAGOR ARRASATE (2013-2015).

## 5.5 Findings

This section describes the findings of the research project so far taking as a base previously mentioned collaborations, publications and projects.

Preliminary field results obtained with the conducted case studies show that the process responses to the requirements identified in the front-end of the innovation process (idea generation and management) and that the platform provides a powerful collaborative tool

| Topic | Needs | Expressions of interest | Evaluations | Direct invitations |
|---|---|---|---|---|
| *Mobility and information technologies and the internet of the future.* | 12 | 30 | 1 | 15 |
| *Health, sport, sanitary services and aging.* | 7 | 52 | 13 | 37 |
| *Nutrition* | 3 | 12 | 1 | 1 |
| *Energy, sustainable development and eco-efficiency* | 14 | 42 | 7 | 9 |
| *Automotive, transport and sustainable mobility.* | 4 | 21 | 4 | 2 |
| *Urbanism, Architecture, Building and Infrastructure* | 2 | 3 | 0 | 1 |
| *Capital goods* | 1 | 6 | 0 | 0 |
| *Durable consumer goods.* | 1 | 1 | 0 | 0 |
| *Mechatronic components* | 3 | 5 | 1 | 2 |
| *Totals* | 47 | 167(5) | 27 | 67 |

Table 5.1: Data collected from Elkarbide case study.

that eases administration work and enhances collaboration among participants. The platform provides better knowledge about existent ideas, higher quantity and better quality of ideas.

In addition, a module to personalise the innovation process parameters has been created so the platform can accommodate different types of processes. The module allows the adaptation of the generic platform to different case studies situations and workflow conditions. Thus, the module enables for example the definition of different scopes and the agents involve in each stage (few experts, administrators, enterprise users or many participants), the conditions to pass from one stage to the next (time, number of ideas...) and the tag dictionary to be used in a specific innovation process.

The data collected from the case study performed with ISEA, named *Elkarbide*, can be found in table 5.1:

In the case study performed with Koniker, named *Ideiak*, the following data was collected:

- 27 users have been registered into the platform. From those two have the role of community managers, 5 users have the role to evaluate (a role created to assess ideas), and the rest are standard users that can introduce, comment and vote on ideas.

- Ideas from previous contests have been including to the platform adding to a total of 108. The aim pursuit by this measure is to provide a base of previous ideas to use as reference.

- Among the members, 22% is considered active user because of their contributions in creating ideas, voting on the ideas or making comments about the ideas.

- About 50% of new ideas submitted have received more than one positive vote, no negative feedback for the moment.

- About 40% of new ideas submitted have been commented by other members.
- The percentage of returning visitors is 32%.

In the case study carried out in Mondragon Corporation[12], which is divided into four main areas; Finance, Industry, Retail and Knowledge, and is today the top Basque business group and the seventh biggest in Spain. Mondragon Corporation has a total of 256 companies and bodies, of which approximately half are co-operatives. The average number of employees at Mondragon Corporation is 83.859 and approximately 9000 students course their studies at Mondragon University.

The case study, named *Ekiten*, is an idea contest driven by the Engineering, Business and Humanities faculties of Mondragon Unibertsitatea (MU) and the sponsorship of MONDRAGON Corporation itself, SAIOLAN entrepreneurship development centre, Debagoiena commercial development centre, Gazteempresa foundation and Athlon enterprise. The objective of the contest is to promote entrepreneurship among students by collecting their ideas on the creation of new enterprises or business models. Information about Ekiten has been collected using Innoweb from 2010. Wave module has been used since 2011. Each year three main topics were withdrawn; rural development, youth-leisure-sports and innovation enterprise. A wave was launched for each main topic.

Every wave had a sponsor from outside the university as the owner or manager for that wave. An external selection committee was appointed by each sponsor company. Each committee was formed by five external experts that established the selection criteria for each wave and made the actual selection of ideas. The criteria considered in most of the waves was related to the level of innovation, definition and maturity of the idea, the technical and economic feasibility, the level of alignment of the idea with the strategy and the priorities set for the topics dealt with in each wave, and finally, the confluence and leverage of the proposal with the capacities and competences available in Mondragon Corporation's companies.

Wave administrators set the general parameters for the wave and configured participants, stages, permissions, vocabularies and timelines prior to opening Innoweb idea management tools to users. Students were allowed to introduce their proposals including their description and the title of the idea, the outcome expected (become a new product, process, service or spin-off), the issues addressed with the proposal, the type of innovation and the objective market or customer. Experts used Innoselect to rate and select best ideas. Events were registered in the platform during the whole process. The events registered were of 4 types: success stories, workshops, information bulletins and coaching sessions. The amount of participants, events, experts, etc. can be found on table 5.2. The outcome values, such as amount of ideas, average or promoted ideas, can be found on table 5.3.

As an example of the type of information that can be gathered with the platform, an

[12] http://www.mcc.es

| Context parameter | 2010 | 2011 | 2012 |
|---|---|---|---|
| *Groups* | 10 | 27 | 46 |
| *Multidisciplinary groups* | 0 | 0 | 2 |
| *Participants* | 40 | 92 | 155 |
| *Experts* | 10 | 10 | 10 |
| *Evaluators* | 9 | 9 | 9 |
| *Companies* | 6 | 6 | 6 |
| *Ideas* | 10 | 27 | 49 |
| *Promoted ideas* | 3 | 2 | 3 |
| *Spin-offs* | 0 | 1 | 1 |
| *Events (success stories)* | 2 | 5 | 8 |
| *Event (workshops)* | 1 | 3 | 5 |
| *Events systematic for each group* | 8 | 8 | 8 |
| *Event (boletines)* | 0 | 0 | 3 |
| *Events total* | 11 | 16 | 24 |

Table 5.2: Inputs

| Outcome | Ekiten 2010 | Ekiten 2011 | Ekiten 2012 |
|---|---|---|---|
| *Ideas* | 10 | 27 | 49 |
| *Promoted ideas* | 3 | 2 | 3 |
| *Spin-offs* | 0 | 1 | 1 |

Table 5.3: Outcomes

extract for the three waves hold in 2012 is presented on table 5.4. The influence of events in the quality of ideas can be observed. Further tracking of those events can be done in the platform. Thus, while 9 events were common to all campaigns, 4 were specific to the Youth-Leisure-Sports wave that obtained better grades and where all ideas were online with the contest objectives.

Taking into account the last use cases presented on 2013 (see section 5.3.3), IMS ideas would be enriched automatically with relevant-specific data provided by LOD/SW repositories. Taking the sustainability example, it's impact will be included among that content. This could help users and administrators to see how important an idea could be. Secondly,

| Context parameter | Rural Development | Innovation Enterprise | Youth-Leisure-Sports |
|---|---|---|---|
| *Ideas* | 4 | 33 | 12 |
| *Average grade of ideas* | 3,17 | 2,78 | 6,68 |
| *Ideas online contest* | 2 | 15 | 12 |
| *Promoted ideas* | 0 | 1 | 2 |
| *Spin-offs* | 0 | 0 | 1 |
| *Events (success stories)* | 6 | 6 | 8 |
| *Event (workshops)* | 3 | 3 | 5 |

Table 5.4: 2012 wave comparison

the tools should help users in order to understand the context of the ideas. With additional data and related information users could understand some issues that the idea itself does not explain explicitly. And finally, the tools are expected to help administrator to perform their task faster and in an easier way. On one hand showing them the relations of the ideas can help them identifying in what are the users concerned and see if there are repeated ideas. On the other hand, measuring their possible impact could help them recognizing the most important ideas and select the best ones, helping even more if there is a big amount of data.

# 6. Conclusions

The basic requirements established for the first prototype have been implemented. Thus a baseline solution to cover the different stages of the idea management process has been deployed. Innoweb platform allows the creation, share, analysis, enrichment and selection of ideas.

The platform also provides semantic format for ideas according to GI2MOWave ontology, a new version of GI2MO ontology that can gather context information(see section 5.2.2). This enables machine automatic search, exploitation of semantic meaning to improve ideas content and the possibility to incorporate ideas to the Linked Data.

Drupal social Web and real time Web modules have been analysed with the objective to select those that will enhance participation and collaboration in the innovation process. The resultant list of modules will be employed in future case studies. The results obtained in the new case studies will be compared to the ones obtained in this case study determining which techniques and tools have a bigger impact on the process. It can be concluded that user activity is satisfactory and higher than in previous idea contests.

Innoweb and the case studies where it has been tested, some conclusions have been drawn:

- According to the requirements a platform to support the innovation process has been built. The platform gathers context data that can be further analysed to determine the influence in the outcome and detect success factors.

- The platform offers data in semantic format. This enables interoperability, machine automatic search, exploitation of semantic meaning and the possibility to incorporate ideas to the Linked Data.

- Although the volume of the waves analysed is small it can be concluded that the platform presents not only activity or traffic metrics, but also quality measures such as grades or relation between ideas collected and ideas on target among other.

- The impact of management decisions can also be measured. That is, campaigns can determine if a workshop or brainstorming session translates into more (quantity) and better (quality) ideas.

- The platform provides a better campaign control. Campaigns can be easily stopped, paused, shorten, expanded or re-launched depending on activity or environment conditions.

- Previous examples and experience recorded in the platform will allow a better design of new campaigns.

- The platform permits the identification of active participants and allows co-creation traceability. This is, if multiple users collaborate in the creation of an idea, their inputs can be traced in the platform.

The aim now is to enhance the platform with new functionality taking into consideration the incremental development cycle approach in order to create a model to integrate Innovation and Technology Watch processes. Thus next cases studies will be conducted at Mondragon Corporation making different platforms interoperable between them.

The next steps on the research are the following:

- Enhance the platform to contemplate other aspects of the innovation process; Technology Watch or outcome traceability.

- Exploit the semantic possibilities already available.

- Keep on collecting ideas and measuring the performance of the platform.

- Keep on analysing the success factors for the innovation process in case studies.

Finally, we can conclude that applying Semantic Web technologies can add information automatically and create an interoperable system that joins different platforms or repositories.

# 7. Acknowledgements

I would like to express my gratitude to all those who supported me during the development of this research work.

Firstly, I would like to thank my supervisor, Felix Larrinaga for giving me the chance of doing this research and helping me on the development of the work.

Secondly, I would like to thank my co-supervisor, Juan Ignacio Igartua for giving the Innovation perspective to the project.

Furthermore, I would like to thank Osane Lizarralde for all the advices and the help given since I started with the research.

Additionally, I would like to thank the people from the Security and Systems lab, Enaitz Ezpeleta, Aitor Lizeaga, Aitor Arrieta, Maitane Barrenetxea, Lorena Martínez, Iñaki Garitano, Aritz Legarda, Maite Beamurgia, Idoia Jimenez, Mikel Iturbe and Lorea Belategi, for the good moments we have shared in all this time.

Moreover, I would like to thank Nuria Fernandez for her support and for putting in all the bad and good times.

Last but not the least, I would like to thank my family for supporting me on every decision I make, for their advices and for helping me in all possible ways.