A Case Analysis of Semantic Technologies for R&D Intermediation Information Management

Ricardo Colomo-Palacios [a], Ángel García-Crespo [a], Pedro Soto-Acosta [b,*], Marcos Ruano-Mayoral [c], Diego Jiménez-López [c]

[a] Computer Science Department, Universidad Carlos III de Madrid, Av. Universidad, 30, 28911 Leganés (Madrid), SPAIN
[b] Department of Management & Finance, Universidad de Murcia, Campus de Espinardo, 30100 Espinardo, Murcia, SPAIN
[c] EgeoIT, Av. Brasil, 17, 28020, Madrid, SPAIN

Abstract:

It is evident that companies and research centres can find a myriad of instruments to fund their R&D (Research & Development) activities. Due to the dynamics of the information management in these circumstances there is a need for intermediary businesses intervention. Consequently, intermediary companies are thus involved in a complex process that needs to be managed and controlled. Within this scenario an Information Technology (IT) platform, named RDi-Advise, was developed by EgeoIT using semantic technologies to support R&D process decisions. This case analysis describes the development and implementation of the platform including the lessons learned during the activity period.

Keywords:

Research and development; Semantic Technologies; Information Management; R&D Intermediation

1. Introduction

Organizations must continuously innovate in terms of product, process, market and business model to remain sustainable (O'Sullivan & Dooley, 2010). Public support programmes are a way for every type of company to secure funding for their R&D initiatives. These programmes, designed to encourage private R&D effort, have been implemented in the US, Japan and the European Union (Busom & Fernánadez-Rivas,

* Corresponding author. Tel.: +34-868887805; fax: +34-868887537

E-mail addresses: ricardo.colomo@uc3m.es (R. Colomo-Palacios), angel.garcia@uc3m.es (A. García-Crespo), psoto@um.es (P. Soto-Acosta), marcos.ruano@egeoit.com (M. Ruano-Mayoral), diego.jimenez@egeoit.com (D. Jiménez-López).

2008). As a part of the European Union, the Spanish government has seen R&D as a strategic issue. In Spain, R&D policy-making is centralised at a high political level in the Ministry for Science and Innovation. This Ministry and the Interministerial Commission for Science and Technology (CICYT) are in charge of developing and controlling the National R&D Plan. This plan known as the National Scientific Research, Development and Technological Innovation Plan (National R&D&I plan) includes twenty-five National Programmes grouped in nine broad areas or fields. The Information Society Technologies Area receives 20% of the support given to the program. Apart from this, the growing importance of R&D for policy-making in Spain is also reflected in the reorientation of the European Structural Funds, as the percentage of these funds devoted to R&D will rise from 12% in the period 2000–2006 to 35% in the period 2007–2013. Moreover, Spain is one of the main beneficiaries of the European Technology Fund created by the European Commission to compensate partially for Spain's loss of funds related to the Regional Cohesion Funds due to the entry of new countries in the Union (Heijs, 2010).

Given the opportunities of funding, many organizations in Spain are pursuing these funds to support their R&D&I plans. Due to the intrinsic difficulty of the process, many organizations hire a private intermediary to help them in the process. According to Howels (2006), an intermediary can be defined as an organization that acts as an agent or broker in any aspect of the innovation process between two or more parties. Such intermediary activities include: helping to provide information about potential collaborators; brokering a transaction between two or more parties; acting as a mediator, or go-between, for bodies or organizations that are already collaborating; and helping find advice, funding and support for the innovation outcomes of such collaborations. The benefits of this kind of agents are well known. An intermediary can help companies to maximize their chances of innovation and increase their likelihood of success in developing new products and services (Lee et al., 2010). However, the growth in the number and range of these actors within the system creates confusion for their clients and to the innovation system overall (Howells, 2006).

This intermediate managing process is complex and may be assisted by technology through the application of a decision support system (DSS). The main objective of this case analysis is to present a system that enables the management of R&D intermediary

firms based on semantic technologies and open standards that have been developed and put into practice by an IT consulted company located in Madrid.

2. Company Background

EgeoIT is a consulting company created in 2002 and initially focused on technological areas related to Geographic Information Systems / Location Based Systems (GIS / LBS) and Customer Relationship Management (CRM). During the last years the company has developed projects for different clients (other consulting companies and public and private end-clients) in the aforementioned areas. These projects comprise all the phases of the cycle of life of information systems, although consulting, analysis and design activities have been the most relevant ones.

Given the expertise gained by the professionals in the company in the different national and international R&D programmes, in November 2008, EgeoIT was named (by CDTI Centre for the Development of Industrial Technology) as a certified mediator of SPAIN under the *Plan de Bonos Tecnológicos* to increase the presence of Spanish firms in the 7th Framework Programme and to facilitate the internationalization of their R&D activities. Regarding this role, the company should focus its effort related on R&D management to the vigilance of calls with an eye to the configuration of valid and sound consortia capable of successfully developing the different projects planned. Therefore, two kinds of activities can be found in the call management process. In the first place, the analysis of the different calls regarding the technological and functional areas they cover, their funding schemes and their eligibility criteria. In the second place, the analysis of the client portfolio of the company to suggest their incorporation in the *consortia* on the basis of their past experience in R&D projects and their interests regarding their technological position. Once the call and proposal management phase has concluded and led to a R&D project, the management of this project is then another area of EgeoIT competence. During this phase, the company puts into practice project management methodologies, techniques and activities to ensure the correct development of the project in terms of costs and scheduling.

3. The project

3.1. Project Genesis

After several years of activity, at the end of 2008, EgeoIT had a large portfolio of clients, most of them small- and medium-sized enterprises (SMEs), participating in regional, national and supra-national competitive calls. The success of the proposed projects could be considered as moderate and was directly influenced by the nature of the call, with regional and national ones being more successful than international ones. Once EgeoIT had been certified as a mediator by CDTI, the number of targeted calls and the set of suitable candidates to consortia grew up significantly. Therefore, the traditional schema used for the management of the calls has turned out to be ineffective since the information to be processed involves a higher number of companies, projects and human resources. Additionally, the client portfolio was fed with medium-sized corporations in which the number of resources and projects required the use of integrated project management platforms. This last fact was crucial in the decision of EgeoIT managers to adopt a new modus operandi. In the prior schema, human resources were pre-assigned to the different projects according to their competences. EgeoIT's consultants were aware of resources' competences; either they were formalized in talent or competence management solutions. The new management platform should allow the integration of national and international consortia that is sharply focused towards the specific targeted call and the organizational competence of the partners and the personal competence of their human resources. Additionally, a 360º feedback process should be established aimed at the evaluation of the partners participating in each project and at the improvement of the partner selection process.

3.2. The tool

The RDi-Advise project presented three relevant characteristics when the managers at EgeoIT decided to put it into operation. Firstly, the support of the company to the project since its very initial phase. Secondly, the decision to apply semantic technologies in its development. It is agreed that semantic enrichment of resources would lead to better search results (Scheir, Lindstaedt & Ghidini, 2008). Ontologies (Fensel, 2002) are the technological cornerstones of the Semantic technologies, because they provide structured vocabularies that describe a formal specification of a shared conceptualization. Ontologies provide a common vocabulary for a domain and define, with different levels of formality, the meaning of the terms and the relations between

them. Ontologies provide information systems with a semantically rich knowledge base for the interpretation of unstructured content (Mikroyannidis & Theodoulidis, 2010). Knowledge in ontologies is mainly formalized using five kinds of components: classes, relations, functions, axioms and instances (Gruber, 1993). The semantic issues of technologies are concerned with ensuring that the precise meaning of exchanged information is understandable by the applications. Semantic technology is one of the most promising areas to be developed by the company and this project could aggregate different initiatives developed using this technology but, with no relations between them. Thirdly, the key requirement of the project would be to foster the adoption of commercial tools instead of new developments. Therefore, the analysis of the user requirements showed that several aspects related to project management could be accomplished by means of conventional tools, which should be open to customization and capable of establishing communication interfaces with additional solutions. Furthermore, most of the activities related to personnel evaluation are already implemented in commercial management solutions.

Thus, the core of the project development can be found in a software component capable of performing the following tasks:

- Analysis of the documentation of each call to determine their requirements using Natural Language Processing (NLP).
- Matching between the required competences and skills and those available in the organization.
- Interface with Human Resources Management (HRM) tools at the clients. This interface will allow the communication of information related to the evaluation of the resources.
- Interface with project management solutions. This interface is aimed at providing these solutions with rich management elements such as task competence decomposition.

Bearing in mind the specified tasks, the architecture of RDi-Advise is based on component groups that interact among themselves to offer an automatic solution to the problem. The conjunction of these systems allows the correct operation of the whole set of components, and the necessary data to achieve the desired outcome. Components might be related to the behaviour as specified in the collaboration between those

elements, turning those structural and behavioural elements into progressively larger subsystems and the architectural style that guides this organization. The main architecture of the system is shown in Figure 1:

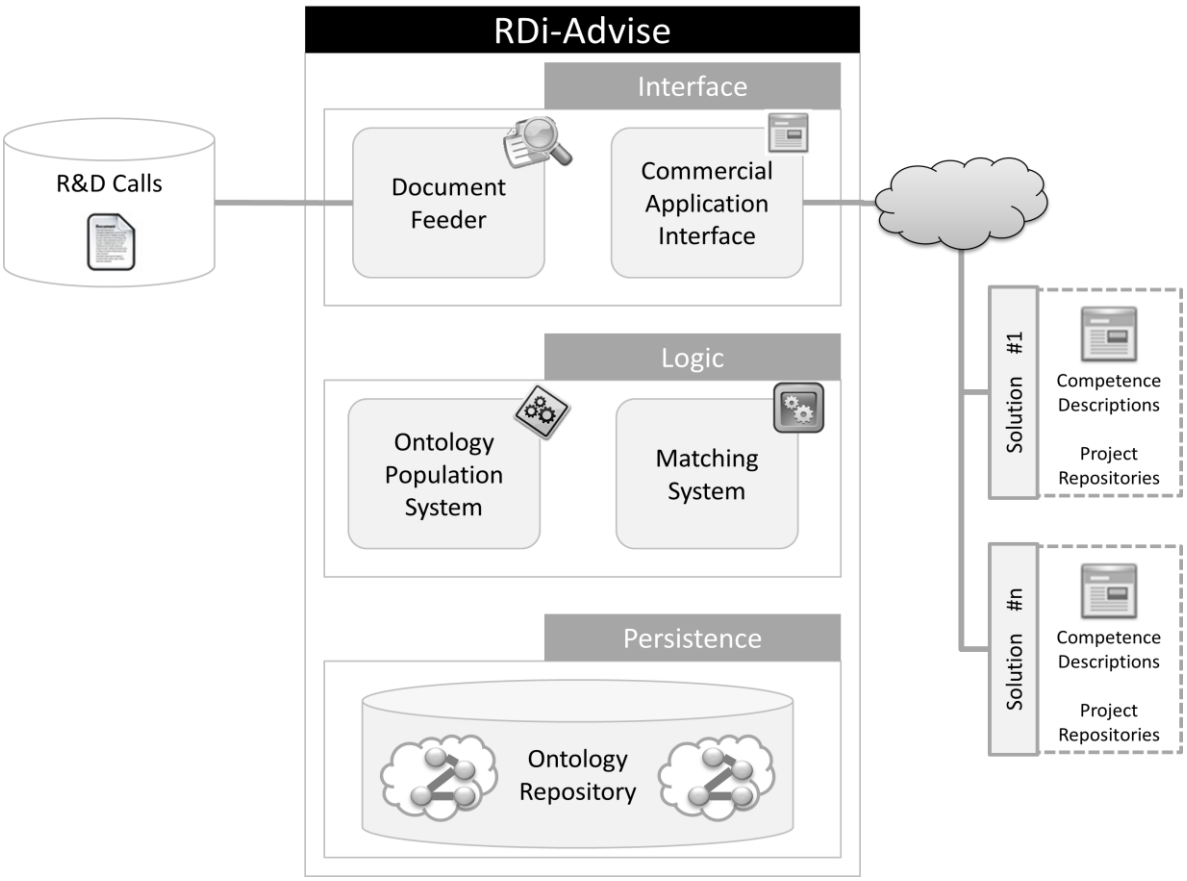

Figure 1. RDi-Advise Architecture

The architecture comprises three operating layers or subsystems. In the following, several of the core components in each layer are detailed.

- INTERFACE. The interface layer is composed of a number of interfaces through which end systems can interact with the system. This layer presents two components.
    - Commercial Application Interface. This component allows the communication with external HRM solutions and Project Management Tools using a web service.
    - Document Feeder. Given a document repository, this component uploads and formats documents and sends them to the Ontology Population System.

- LOGIC. This layer encompasses the reasoning, inference and business logic management functionalities through two components:
    - Ontology population system. This component will take as an input the public documents that rule every call to analyze them and populate the domain ontologies related to R&D calls. This component will include implementations using GATE for name annotation and JAPE for element extraction. Subsequently, these elements are categorized regarding the taxonomy defined in the ontology and sent to the persistence layer to be stored. The ultimate goal of our approach is to populate the ontology with all the relevant information identified. The populated ontology will then serve as the keystone component for the matching system.
    - Matching System. The outcome of this component is a target consortium and, if possible, the human resources taking part in it. The information that is used as an input is the domain ontology populated with the data about the different target calls and the competential information acquired via the Commercial Application Interface. This information is exchanged by the conversion of the data to the HR-XML standard.
- PERSISTENCE. Finally, the persistence layer stores the knowledge about the competence needs and calls requirements. Ontologies are defined using the Ontology Web Language (OWL) (Bechhofer et al., 2004) while the storage and ontology reasoning has been developed based on the Jena framework. Jena is a framework for building Semantic Web applications that provides a programmatic environment for RDF, RDFS and OWL, SPARQL and includes a rule-based inference engine. Both ontology schemas and its populated instances are stored in the KAON2 ontology repository, an infrastructure for managing OWL-DL ontologies.

3.3. Output

By the end of 2009, the first operative versión of RDi-Advise was available to be put into operation in the management environment at EgeoIT. From the point of view of its operation, RDi-Advise is highly dependent on HRM systems. Therefore, specific interfaces for several open-source HRM tools were developed by leveraging the extensibility capacity of HR-XML standard, and thus, the interoperability with additional HRM solutions that support this standard can be also achieved.

The output of the system is a proposal for a consortium of organizations to attend a certain call. In those cases in which competence information about employees is not available, the suggested composition of the team is made by selecting those resources that show the competence references that best fit the requirements identified.

4. Lessons Learned

The lessons learned during the different phases of the RDi-Advise project can be classified into two categories. Firstly, according to the tool and the technology it involves and in the secondly, on the basis of its implementation and deployment.

4.1. Technology

Regarding the development of the RDi-Advise and the technology it involves, the first lesson learned is related to the construction of the domain ontology. In this case, the ontology was built on the domain of Information Technology research. Such ontology should have a controlled vocabulary, which is structured to annotate and detail the information related to the different aspects of the domain. However, ontology construction is not a simple task. This domain ontology has been defined using the Ontology Web Language (OWL) (Bechhofer et al., 2004). The OWL language presents three variants: OWL-Lite, OWL-DL and OWL-Full. OWL-Lite provides a small set of features, while OWL-DL is more expressive than OWL-Lite, providing decidability based on description logics. OWL-Full allows full expressivity but decidability is not warranted. For this reason, OWL-DL was used for the ontology definition. However, even though from a technical point of view the architecture solved the problems satisfactorily, the ontology was not sound, requiring several concepts and relations additions, even after having released the first stable version of the platform. Bearing in mind that project partners had previous experience in ontology development, a software engineering approach to ontology building was chosen according to the works by De Nicola, Missikoff and Navigli (2009), who present an iterative method for ontology building. Conversely, although the construction process was closely managed, the results from the ontology point of view where not satisfactory. The analysis of the origins of this discrepancy discovered that it was focused in one of the areas of the domain and was caused by the lack of information provided by one of the domain experts external to the company. Thus, the revision of the project documentation

revealed that the aforementioned expert had not provided the necessary information for the development. Therefore, the requirements were elaborated from contributions to the state-of-the-art and the research performed by the development team, a circumstance that was neither detected by the project manager nor reported by the team nor traced from the sources of the requirements. Once this element was identified, the missing information was completed and the ontology versioning stabilized around suitable values according to adaptive maintenance. Nonetheless, even after ontology completion and when additions were less focused, concepts addition was constant in the ontology, because of the volatility of the technology and its inherent innovation dynamics.

The second of the relevant aspects is the operation of JAPE and GAPE. These technical components, while presenting a stability supported by their history, are rather inflexible and require a really precise grammar definition. This precision can be achieved if a domain ontology is available but it has not, in this project, guaranteed a reliability of over 90% when identifying concepts in any of the analyzed processes. The third of the elements is the disparity of the results obtained after the analysis of HTML pages and PDF documents. During the population of the ontologies it could be observed that HTML pages returned a higher number of results and those results where more accurate since they were manually corrected less often. Therefore, these two aspects (second and third) evidence that human supervision is required for ontology population as NLP recognition does not output satisfactory results.

The last of the elements that should be highlighted is related to the behavior of the system and its evolution over time. RDi-Advise has been designed to provide consortia configurations based on the requirements of the calls, and the personal and organizational competence of a set of companies. However it can be stated that the system provided anomalous consortium configurations and, furthermore, those results were reproduced under certain conditions. This circumstance was caused by the absence of reliable data from the companies regarding their competence areas. Therefore, it is crucial to state that having a DSS without an automatic learning component perpetuates errors and makes the system seem rather static when these systems seek to be as flexible as possible.

4.2. Implementation and adoption

The first lesson learned regarding the adoption and deployment of RDi-Advise is the scarce penetration of HRM tools in Spanish SMEs, contrary to the benefits described in the literature (e.g. De Grip & Sieben, 2009; Ordoñez de Pablos & Lytras, 2008). Only one of the company's clients, SME, was already using an HRM solution. Therefore, the deployment process in this scenario entailed the implantation of an HRM solution. This circumstance has deterred the development of the tool for a double reason, sometimes foiling its use. Managers considered that using HRM tools is a "loss of productive time". Additionally, the client using a HRM solution was not fully exploiting its functionalities (competence evaluation was not performed for every employee) preventing fully harnessing their potential. Therefore, the statement of the problem is that the adoption of a solution that requires the adoption of a second one for its implantation is conditioned by the adoption of the latter. This has been the case of RDi-Advise. Technology adoption issues have been profusely studied in the literature (e.g. Liao, Palvia & Chen, 2009), even in the scope of SMEs (e.g. López-Nicolás & Soto-Acosta, 2010) and their effects are highly relevant for the results of the projects.

The second of the aspects was related to the recommendation for the implantation of an HRM tool. A large number of consortium candidates not using HRM solutions (about 60% of the total), chose to implant a solution of this kind. All of them chose an open source market solution. The authors consider that the adoption of this kind of tools, generally distributed using free software licenses (GPL, AGPL, BSD, …), is a key factor for the success of RDi-Advise. Bearing in mind that RDi-Advise was costless for EgeoIT clients, counting on a tool complementary to RDi-Advise by means of a controlled expenditure has been a critical factor for the success in those companies really committed to the project. Therefore, free software tools should be considered as a technology adoption enhancer.

The third of the considerations is the mistrust shown by the companies to share information regarding their personnel performance and competences. The system traditionally used by EgeoIT for intermediation activities required an intensive exchange of information about resources and competences. In this scenario, clients did not have any qualm to circulate their information. Nevertheless, qualms began to appear when companies started to use the tool. In a large number of the cases (40%), direct connection to the system could not be achieved, which forced the managers to request

reports in HR-XML format that were transmitted via e-mail and subsequently uploaded onto the platform. Mistrust of this nature has been identified in the literature (e.g. Coakes, Coakes & Rosenberg, 2008). Therefore, companies are mistrustful when it comes to granting access to their information systems and, furthermore, when the information is related to the intellectual capital of the company.

The last of the aspects that the authors consider relevant refers to the use and functioning of the solution within a productive environment. In those cases in which all the requirements for the implantation of the platform were fulfilled, two problems arose that deterred its full exploitation. First, it was necessary to have historical data about projects and employees' assessments. This process, intrinsically onerous, was not performed in almost any case, and compelled the minimum set of data required for the platform to be exclusively loaded in order to operate in a relatively effective way. Second, managers found that resource planning and assignation in the client companies was a long way from following quality criteria. Even though the initial commitment showed the availability of a given resource at a certain time, awaress of the available resources was hampered. This circumstance meant that the schedule accomplishment for the overall solution was cast in doubt.

5. Conclusion

This case has illustrated the implantation of an interorganizational information system that uses semantic technology. It presents several intrinsic aspects related to information systems implantation in a number of organizations, contributing to the debate about the contribution of the different stakeholders in distributed environments and the previously known issues relating distrust to revealing strategic information. From the technical point of view, despite the adoption of specific methodologies, the construction of the ontology has not been fully satisfactory, due to different issues in the management of the ontology engineering process. Additionally, several flaws in the use of NLP tools have been detected, forcing us to carry out reviews on the outputs of the analysis of the call contents. Lastly, given the DSS characteristics of the system recommends the implementation of learning mechanisms to tune the recommendations.

Firstly, it will be necessary to develop learning mechanisms so that the tool can take advantage of the feedback generated during its operation. This characteristic will allow

the elimination of recurrent errors in partner selection, so allowing the continuous improvement in its results. Secondly, the addition of fuzzy mechanisms for consortium construction using approximate reasoning is suggested as proposed by Carlsson et al. (2007). Thirdly, the scope of project portfolio selection is intended to be broadened. In the current scenario it is supposed that all of the companies can participate in the analyzed projects and that they present a similar interest in developing them.

References

Bechhofer, S., Van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D.L., Patel-Schneider, P.F., & Stein, L.A. (2004). OWLWeb Ontology Language Reference. http://www.w3.org/TR/owl-ref/

Busom, I., & Fernández-Rivas, A. (2008). The impact of firm participation in R&D programmes on R&D partnerships. Research Policy, 37(2), 240–257.

Carlsson, C., Fullér, R., Heikkilä, M., & Majlender, P. (2007). A fuzzy approach to R&D project portfolio selection. International Journal of Approximate Reasoning, 44(2), 93-105.

Coakes, E.W., Coakes, J.M., & Rosenberg, D. (2008). Co-operative work practices and knowledge sharing issues: A comparison of viewpoints. International Journal of Information Management, 28(1), 12-25.

De Grip, A. & Sieben, I. (2009). The effectiveness of more advanced human resource systems in small firms. The International Journal of Human Resource Management, 20 (9), 1914–1928.

Fensel, D. (2002). Ontologies: A silver bullet for knowledge management and electronic commerce. Berlin: Springer.

Gruber, T.R. (1993). A translation approach to portable ontology specifications. Knowledge Acquisition, 5(2), 199-220.

Gutjahr, W.J., Katzensteiner, S., Reiter, P., Stummer, C., & Denk, M. (2008). Competence-driven project portfolio selection, scheduling and staff assignment. Central European Journal of Operations Research, 16(3), 281-306.

Heijs, J. (2010). Erawatch National Profiles. Spain. The Relative Importance of Research Policy. Available at http://cordis.europa.eu/erawatch/index.cfm?fuseaction=ri.content&topicID=612&parentID=12&countryCode=ES.

Howels, J. (2006). Intermediation and the role of intermediaries in innovation. Research Policy, 35 (5), 715–728.

Lee, S., Park, G., Yoon, B., & Park, J. (2010). Open innovation in SMEs—An intermediated network model. Research Policy, 39 (2), 290-300.

Liao, C., Palvia, P., & Chen, J.L. (2009). Information technology adoption behavior life cycle: Toward a Technology Continuance Theory (TCT). International Journal of Information Management, 29 (4), 309-320.

López-Nicolás, C., & Soto-Acosta, P. (2010). Analyzing ICT adoption and use effects on knowledge creation: An empirical investigation in SMEs. International Journal of Information Management, in press.

Mikroyannidis, A. & Theodoulidis, B. (2010). Ontology management and evolution for business intelligence. International Journal of Information Management, in press.

O'Sullivan, D., & Dooley, L. (2010). Collaborative innovation for the management of information technology resources. International Journal of Human Capital and Information Technology Professionals, 1(1), 16-30.

Ordóñez de Pablos, P., & Lytras, M.D. (2008). Competencies and human resource management: implications for organizational competitive advantage. Journal of Knowledge Management, 12 (6), 48-55.

Scheir, P., Lindstaedt S.N., & Ghidini, C. (2008). A Network Model Approach to Retrieval in the Semantic Web. International Journal on Semantic Web and Information Systems, 4(4), 56-84.