ABSTRACT
Recent advances in networking and telecommunication technologies combined with the vast load of readily available scientific data, urge towards the implementation of software systems that are able to collect, combine and present information from distributed repositories. At the same time, scientists, researchers, academics and experts demand flexible infrastructures that provide just-in-time information, with reliable, peer-reviewed data, through ubiquitous interfaces. Towards this direction, several international organizations including the European Commission, promote the development of software platforms that will facilitate universal access to information aggregated from open access repositories. In this paper we discuss the requirements engineering process and the key functionality for a system that supports knowledge extraction and dissemination from distributed repositories, while at the same time enabling and integrating the social interaction of its users.

Categories and Subject Descriptors
D.2.1 [Software Engineering]: Requirements/Specifications – Elicitation methods, Methodologies

General Terms
Documentation, Design, Human Factors

Keywords
Information Systems, requirements elicitation.

1. INTRODUCTION
Knowledge is considered as one of the largest capitals for organizations and its effective management is of prime importance. Scientific research produces a vast amount of knowledge, leading thus to the creation of repositories of excessive size.

Meanwhile, there is an ongoing demand for content that is available in the form of open access, as it facilitates knowledge reuse without restrictions or royalties and fees. As many organizations create separate, standalone content repositories it is important to provide ways that enable utilization and exploitation of the ever growing amount of available knowledge.

The advent of the mobile networking paradigm as well as the expansion of research repositories make access to content feasible from anywhere, at any time, using wired or wireless devices, enabling the ubiquitous access to information paradigm; however, exploitation of this vast volume of scientific information is hindered due to mainly two reasons: first, because a part of it is not accessible to all the members of the scientific community and second because the systems’ user interfaces are not adequate and intelligent enough in order to enable fast and easy identification of the most relevant and valuable scientific resources.

In order to circumvent these obstacles and develop software systems that maximize benefit for all users, enabling and fostering information management, it is essential during the requirements engineering process as well as during the design and specification phase to pay attention in specific issues, namely usability, user interface design and interoperability. Our work has focused on employing methodologies and tools for the effective utilization of accurate data, making it available through intelligent interfaces that enable efficient filtering and provide access to the most valuable and relevant pieces of information.

In this paper we present a system that enables access to different open access research repositories, providing intelligent interfaces and content classification based on the use of ontologies; we describe the techniques, methodologies and tools that were used for the requirements engineering process with the active engagement of end-users of different technological and cultural backgrounds, originating from different countries. Furthermore we indicate the value of using mockups in order to describe upfront the system functionality and better elaborate and agree on the intended use cases with all stakeholders.

The remainder of the paper is organized as follows: Section 2 presents related work in context. Section 3 describes the requirements elicitation and verification methodology, with tools and techniques used, while Section 4 concludes the paper.

2. RELATED WORK
During the last years there is an increasing interest to develop effective interfaces and systems that facilitate access to diverse knowledge repositories. As the international academic community moves towards open access and seeks also ways to improve the provision of accurate information through effective interfaces, it is essential to build systems that provide access to scientific information using appropriate infrastructures.
One of the important challenges is to be able to utilize Open-access repositories. Lately, a number of systems have focused on the creation of tools that provide quick access to people seeking information. MERLOT is a free and open online community of resources designed primarily for faculty, staff and students of higher education from around the world to share their learning materials and pedagogy [1]. Through MERLOT’s features, users are able to access peer-reviewed online learning materials and, as registered members, to contribute material that will be later peer reviewed and catalogued by groups of users with different skills and privileges within the framework of MERLOT communities.

The Organic.Edunet platform [2] facilitates access and exploitation of educational content in respect to the use of digital content. The platform provides a multi-lingual environment for searching, retrieving and accessing content from different users.

Mendeley [3] is another system that combines information management functionality with social networking features. It allows automatic extraction of metadata from PDF papers and allows back-up and synchronization across multiple computers and with a private online account. It also supports social networking features (comments, profile pages), importing of documents from various external sources (PubMed, Google scholar, ArXiv), creation of both private and public groups to collaborate and annotate research resources.

2.1 Integrating social networking features to information management

Lately, social networking attracts a lot of attention and has become part of everyday life to researchers, practitioners, students and information seekers in general. A lot of portals and systems integrate social networking facilities to their functionality. Most notable ones are:

The VIVO [4] system, an interdisciplinary networking portal for researchers that enables the discovery of research and scholarship across multiple institutions. Content in any local VIVO installation may be maintained manually or brought into the database in automated ways from local systems of record providing the best available human resources, grants, courses, and faculty activity information. In addition to centralizing information from a variety of authoritative sources, VIVO also makes information available for consumption by other web sites.

The Oxford Blue Pages [5], a pilot system to gather, store and share publicly available research activity data (i.e. data about researchers and research at Oxford) for the benefit of all members of the University of Oxford and others. It develops an infrastructure that harvests and archives research activity Data and to disseminate and reuse this kind of data by using a lightweight solution based on semantic web technologies. Data held in the registry can be re-used for new purposes enabling innovative links and new services to be created.

CiteULike [8], a platform aimed to promote sharing of scientific references between researchers. In fact it combines scientific research features with social networking services.

Zotero [13] is a tool that allows researchers to gather, organize and analyze sources (citations, full texts, web pages, images, and other objects), and also allows research information collaboration and sharing. Zotero mainly interacts with external systems to trace the actual link.

2.2 Repository management software

An important role in systems that manage information is played by the underlying management platform that handles the resources. Most notable is the DSpace [6] platform that enables organizations to:

- capture and describe digital material using a submission workflow module, or a variety of programmatic ingest options
- distribute an organization’s digital assets over the web through a search and retrieval system
- preserve digital assets over the long term

Another notable platform is Europeana [7], which provides access to different types of digital content providing thus support to promotion of cultural heritage. It supports different type of content, such as paintings, books, films and archives.

3. REQUIREMENTS ENGINEERING

A requirement in software engineering is a property that the system must have to provide value to the stakeholder [9]. Requirements engineering is a demanding task, which requires skills, experience and planning as well as the use of appropriate methodological tools in order to complete successfully its purpose.

The requirements engineering process can be seen as a structured set of activities which are followed to derive, validate, and maintain a systems requirements document [10]. The basic stages in the requirements engineering process includes requirements elicitation, requirements analysis and requirements validation. Each phase has a distinctive role in the software engineering process:

- Eliciting requirements: the task of communicating with stakeholders in order to determine their requirements.
- Analyzing requirements: determining whether the stated requirements are unclear, incomplete, ambiguous, or contradictory, and then resolving these issues.
- Recording requirements: Requirements might be documented in various forms, such as natural-language documents, use cases, user stories, or process specifications.

A number of criteria such as personal preferences, non technical requirements (such as technical limitations, available configurations) will direct the development process towards the selection of the desirable system features. The major focus of these phases is to identify the key stakeholders and to try to make them participate in the software engineering process, bringing up their real needs.

It is essential to point out that there is variation to the requirements engineering processes that would fit an organization, since each organization has different characteristics. A number of challenges are associated with the management of the process, in order to adjust to the various demands and tasks according to the needs of the organization. In our approach we had to follow a hybrid approach that allowed interaction with the main stakeholders from the early phases of development in order to ensure that the system under development would be fitted to their personal needs.

3.1 Requirements elicitation

The complete list of essential requirements was identified after processing the results from different requirements gathering
techniques such as: face to face interviews, and workshops held between different stakeholder groups. These results together with results from an online questionnaire created to support the requirements engineering process helped to record the most essential demands from the users perspective. During the workshops often several users provided arguments in favor of or against a specific requirement. When there was substantial debate and the decision was not apparent, the workshop organizers used dotmocracy sheets to verify the opinion towards whether the stakeholders were basically leaning \[11\]. This is a simple approach were a statement to be verified is recorded in a document in which the users have the chance to rate whether they are against or in favor of a specific requirement. In the end by the number of dots (votes) in each rating category the workshop organizer can easily extract the results and estimate the average user’s opinion. Our approach included the organization of several workshops organized in different countries each, and interviews with an adequate number people involved in different organizations or universities. The main groups engaged in these included academics, researchers, Open Access / Metadata / Semantics Experts, practitioners, students and library personnel. The proposed system as recorded by the user demands, aims in supporting the scientific collaboration through which synergies may additionally multiply the research impact. The concept of socially supported research also and the need to identify the impact of scientific contributions attracted the attention for most of the stakeholders group. Another notable issue was the need to determine the presence of the stakeholders to keep the system alive through the interaction and promotion of research material. The recorded needs and the features that the system aims to provide are depicted in Figure 1.

3.2 Requirements verification

After the completion of the workshops, the developer’s team worked in groups in order to decode and integrate them in the elicitation documents. At first the fundamental requirements were identified to which there was almost consensus between the different stakeholders. Next, after a study and comparison with other systems that provided similar services and by estimating requirements that were practically missing from the system, an online questionnaire was compiled so that the average opinion of stakeholders could be reported in a flexible to processing way. There was special concern while creating the questionnaire so that the answers were on a number scale in order to make easier the processing and to avoid recording vague or difficult to be processed answers from the users.

The online questionnaire facilitated the participation of several researchers from different countries. More than 100 answers from people with diverse backgrounds and cultural influences were gathered from more than 15 countries.

Accordingly, the key requirements identified from the previous phase were recorded in the form of use cases. In order to be able to verify the requirements, a list of use cases was created describing the different system functionality. A number of indicative mockups were created in order to facilitate further the verification process. Finally the extracted requirements were recorded using the UML modeling language in the form of Use Case Diagrams. Figure 2 depicts the main use cases as identified during the requirements elicitation and verification process. The main outcome of the system’s requirements engineering was that users demand efficient ways that enable both uploading of content as well as identification and downloading of reliable information. The need to link the system under development with open-access repositories was identified by all stakeholders as an essential requirement. Typical search facilities and advanced search capabilities, interfaces that allow retrieval through browsing as well as through semantically enhanced search were all considered essential by the stakeholders. In addition, most of the users commented on the necessity to integrate mechanisms for the provision of additional information associated with the resources as well as a quality control mechanism that ensures the quality of the information stored within the system repositories.

3.3 Managing the process

One of the main challenges in our approach as stated previously was associated with the fact that different parties in the development process spanned between different countries and also had diverse backgrounds both knowledge and cultural. In order to overcome the difficulties associated with these issues we had to use different tools while the management dimension proved to be of prime importance.

Therefore we utilized tolls that facilitated collaboration and joined contribution such as wikis and forums that enabled the different parties to cooperate, mailing lists to circulate changes and to ensure timely cooperation between different parties, online questionnaires and at the different sites of the requirements gathering teams we also used traditional requirements gathering techniques such as face to face interviews and workshops.

From a technical perspective, except of the wikis the developers used a web accessible requirements management tool in order to be also able to manage changes in the requirements and each party to be able to access it independently of location.

4. CONCLUSIONS

As research content is substantially increasing on a daily basis and as advances in networking technologies enable ubiquitous access to information, intelligent interfaces are necessary to provide access to just-in-time information and to identify the content of both the highest relevance and impact. In order to create a system
that would provide value to the stakeholders, their participation is vital from the early stages of system development, these of requirements elicitation and verification. In our case several difficulties had to be tackled: on one hand the developers’ team was distributed, originating from different countries and also the stakeholders were also originating from different countries and also had different technical backgrounds.

We described the approach that we followed in order to successfully overcome those issues and complete the requirements engineering process, securing the upfront active engagement of all different stakeholders’ parties, especially through the use of mockups, live documents in the form of wiki pages and interactive discussion for consensus building. Additionally, we provided insight on the techniques and tools that enabled the successful collaboration between the distributed members of the software development team.

The recorded requirements are part of an ongoing development phase for a project that aims to support the integration of scientific resources from different open access repositories in order to promote scholar research.

The described methodology and the tools aim to provide a basis for future distributed projects while also drafting guidelines for successful requirements elicitation in similar context. We plan to integrate a framework that allows for timely determination of preferences, issues and suggestions and also handles conflicts between different stakeholders.

5. ACKNOWLEDGMENTS

This work has been carried out within the VOA3R project (Virtual Open Access Agriculture & Aquaculture Repository – http://voa3r.eu), which receives funding from the European Commission’s ICT Policy Support Programme (ICT PSP), Theme 4 – Open access to scientific information.

6. REFERENCES


