PROJECT FINAL REPORT

KnowHow

Demonstrating a Cloud-based Knowledge Management Platform for Social Business

Grant Agreement number: 606571, FP7-SME-2013

Project acronym: KNOWHOW

Project title: Demonstrating a Cloud-based Knowledge Management Platform for Social Business

Funding Scheme: Research for the Benefit of SMEs, Demonstration Activity

Period covered: 1st of July 2013 to 30th of June 2015

Name of the scientific representative of the project's co-ordinator, Title and Organisation:

Jiasheng Wang

Manager Innovation and Business Design

CAS Software AG

D-76131 Karlsruhe, Germany

Tel: +49 721 96380

Fax: +49 721 9638299

E-mail: innovation@cas.de

Project website address: www.knowhow-project.eu

TABLE OF CONTENT

Table of Figures

Final publishable summary report

1.1 Introduction

In line with the growing adoption of social business, a new social model of knowledge management is emerging, which reflects the mind-set of small and medium sized companies. The challenge is to place the massive amount of socially produced information into context, to enable knowledge-generating interaction, and to address the specific needs of SMEs such as flexibility, flat hierarchy, agile collaboration, and open innovation. The basis for a solution that can facilitate this new social model of knowledge management has been laid with the OrganiK project, which was dedicated to developing a knowledge management system for the specific needs of knowledge-intensive SMEs. The main outcome of the OrganiK project was a KM platform providing key KM functionalities with technology for intelligent information processing powered by semantic and statistical analysis of content and user behaviour, recommendation algorithms, and automated taxonomy learning. The results of the OrganiK project represent assets with high exploitation potential in the context of social business. Having recognized this, most of the SMEs who have participated in the OrganiK research came together to explore how they can leverage these results to develop innovative services and solutions for their social business markets.

As the successor of the OrganiK project, **KnowHow** aims to demonstrate the scale-up of the OrganiK KM platform from the research prototype stage into an industrial-grade Knowledge Management Platform as a Service (KM-PaaS) that supports social business applications. More explicitly, the KnowHow project intended to (a) carry out targeted improvements and optimization of specific OrganiK platform components,

(b) expose key OrganiK platform components as services, and

(c) re-package and re-configure the OrganiK platform distribution so as to enable deployment on a virtualized cloud infrastructure;

Finally, the central goal of the project was to demonstrate the scaled-up KnowHow KM-PaaS with four social business trials in the domains of (1) Collaborative Innovation, (2) Social Enterprise, (3) Supply Chain Networking, and (4) Knowledge-Based Trade Fair Services in order to validate its applicability. Through the proposed demonstration activities, the KnowHow consortium seeks to bridge the gap between the research results produced by the OrganiK project and the requirements for future commercialization, thus preparing the ground and assisting the partners to achieve a positive impact on the market.

To fulfil this goal, the consortium of the KnowHow project needs intended to achieve the following objectives:

I. Analyse the SME-specific social business market environment, assess competition and the commercialization opportunities for the KnowHow KM-PaaS, and decide on precise scope and objectives for the demonstration.

II. Analyse and plan scale-up activities in specific areas where enhancement of OrganiK KM platform components is critical to achieve industrial-grade strength, as well as to analyse the requirements for the trials and plan for iterative software development and releases.

III. Carry out the actual platform scale-up activities as planned, as well as implement four new technological solutions/trials on top of the KM-PaaS.

IV. Validate the KnowHow KM PaaS through the four trials in the domains of collaborative innovation, social enterprise, supply chain networking, and knowledge-based trade fair services.

V. Develop a business plan and marketing strategy that will support the participating SMEs to commercialize the KnowHow KM-PaaS and/or the results of the trials.

VI. Disseminate the project results and maximize awareness among European SMEs and beyond, taking into consideration the protection of the participants' IPRs.

VII. Provide the means for allowing technical and managerial staff to be trained on using and marketing the KnowHow KM-PaaS.

VIII. Ensure project results of high quality, appropriate handling of IPRs, and smooth execution of the entire project through coordination, quality assurance planning and risk management.

The KnowHow consortium is composed of five SMEs who have either been directly or indirectly involved in the OrganiK project. The project coordinator **CAS** who was also the coordinator of the OrganiK project is in charge of the collaborative innovation trial. **LTC**, as one of the leading language technology providers was also a partner of the OrganiK project and is in charge of the social enterprise trial. **Syria**, an Italian eCommerce provider who also participated in the OrganiK project is responsible for the supply chain networking trial. **ASN** is a small Greek eCommerce and social web service provider who was collaborating with the OrganiK partner QMS and is in charge of the knowledge-based trade fair services trial. Finally, **IDEAS**, a start-up founded by the core research team members involved in OrganiK is in charge of scaling up the KnowHow KM-PaaS.

KnowHow began in July 2013 and had duration of 24 months.

1.2 Project context

The environment in which modern organizations operate is undergoing a major paradigm shift fuelled by the pervasiveness of ICT and the increasing popularity of social computing. There is an ever-growing number of social media users — over 1.5 billion users between YouTube, Facebook, Twitter and LinkedIn now. At the same time, an ever-growing number of organizations are adopting a model of social business[1], by making active use of social media to communicate and engage with external audiences such as customers, prospective employees, suppliers, and partners.

In this context, Knowledge management is being reshaped, but nevertheless retaining its importance for organizations of all sizes. The role of knowledge management is in many cases ever-more important, considering the unprecedented volumes at which users of social media produce information which still needs to be placed into context before it can become actionable knowledge.

This area of intersection between knowledge management and social business was precisely the focus of the preceding research project OrganiK, which run from September 2008 to August 2010 and was co-funded by the European Commission's FP7 Research for the Benefit of SMEs programme (Grant Agreement number: 222225).

1.2.1 The OrganiK research project

The partners of OrganiK (An Organic Knowledge Management System for Small European Knowledge-intensive Companies) were motivated to setup the project in 2008 based on a common observation: the knowledge management approaches and tools adopted by large organisations tend to be expensive, relatively inflexible and mostly unsuitable for small knowledge-intensive companies. This is because traditional knowledge management approaches, as initially formulated during the 1990s and refined over the next decade, usually focus on top-down, rigid and tightly controlled knowledge repositories. This may certainly be appropriate in particular environments within larger organizations, but it does not meet the increasing need of small knowledge-intensive companies for loose/flat structures, agile collaboration, ad-hoc capture and sharing of knowledge, and open innovation. Identifying this need led the OrganiK project consortium to put forward the concept of creating a different kind of knowledge management system; one that places more emphasis on social interaction and provides advanced tools for filtering, pushing, and pulling information.

The consortium of the project, which was coordinated by CAS Software AG, consisted of five SMEs active in different knowledge-intensive business domains and four research organizations. The participating SMEs, in addition to CAS Software (Germany), were Language Technology Centre (UK), LeserAuskunft (Germany), Quality Maritime Services (Greece), and Syria Informatica (Italy). The research organizations contributing to the project were the German Research Centre for Artificial Intelligence, the Institute of Communication and Computer Systems (Greece), the South-East European Research Centre of the University of Sheffield (Greece), and the Information School of the University of Sheffield (UK).

The main outcome of the OrganiK project was a KM platform providing functionality for (1) knowledge discovery – search and retrieval of knowledge resources but also subject experts with the help of advanced information retrieval, visualization, and navigation tools); (2) knowledge publication – supporting the authoring, structuring, contextualization and release of knowledge resources; and (3) knowledge-based collaboration – enabling the joint creation, sharing and application of knowledge by knowledge providers and knowledge seekers with the help of the communication, coordination, and community management services.

The above functionality was delivered on the basis of combining key Web 2.0 technologies such as wikis, blogs, micro-blogs and social bookmarking, with technology for intelligent information processing powered by

[1] Dion Hinchcliffe and Peter Kim (2012) Social Business by Design: Transformative Social Media Strategies for the Connected Company.

semantic and statistical analysis of content and user behaviour, recommendation algorithms, and automated taxonomy learning.

The first fully working release of the OrganiK prototype platform was made available to the SMEs participating in the project in March 2010. During the final phase of the project the SMEs developed and evaluated four customized OrganiK solutions motivated by their respective business domains. By the time of the project's completion, OrganiK had produced an innovative KM platform, a well-tested and documented methodology for its application in different domains, and a set of five reference applications in use by the SMEs participating in the project.

Key findings of the OrganiK project:

- There is a strong demand among SMEs active in knowledge-intensive business domains for more flexible KM tools which reflect their loose structure and dynamic workflows and can effectively support their everyday operations and need for continuous innovation. KM initiatives are increasingly becoming more social and place more emphasis on the relationships and interactions between teams and individuals in every day work.
- The usage of the OrganiK KM platform within knowledge-intensive SMEs can significantly improve the effectiveness and efficiency of their processes by supporting better knowledge discovery, knowledge publication, and knowledge-based collaboration.
- The intelligent information processing capabilities of the OrganiK KM platform are highly practical. The participating SMEs reported particular satisfaction with the platform's information push functionalities (such as intelligent tag recommendation, content recommendation, user recommendation, automatic generation of taxonomies from unstructured text sources, notifications). These functionalities were considered a welcomed addition to the social networking functionalities of the platform (like wikis, blogs, micro-blogs and social bookmarking).

The project's final review report (issued 17.02.2011 by the CEC) attests to the excellent results produced by the OrganiK project, and certifies that the project has successfully met all of its objectives. As stated in the report, the flexibility and diversity of the OrganiK KM system has been proven, and the results represent a definitive contribution and advancement beyond the state of the art.

1.2.2 Enhancements needed in advance of commercialization

The results of the OrganiK project represent assets with high exploitation potential in the context of social business. Having recognized this, most of the SMEs that have participated in the original OrganiK project were keen to explore how these results can be leveraged to develop innovative services and solutions for the social business market. Having acquired the relevant experience by participating in OrganiK represents a major opportunity for these SMEs — and certainly an advantage over their competition which may need to develop their own expertise and social business software codebase from scratch. Therefore, the KnowHow consortium has been set-up to prepare the ground for commercialization through the support of FP7.

A key precondition for commercialization is to bridge the gap from the research prototype state in which OrganiK finished, to a full blown industrial scale software product. During the OrganiK project, the OrganiK KM platform was developed and evaluated on a small scale in order to verify its applicability and value within the SMEs participating in the project. Despite the technical improvements which were identified and carried out during the partner-specific case studies, there were specific enhancements required before being able to bring the OrganiK concept, methodology, and framework to the European and global market. In particular, these enhancements draw on the final validation results of the OrganiK project, and concern the following issues:

1. **Carrying out targeted improvements and optimizations in specific OrganiK platform components:** The social networking and intelligent information processing capabilities of the OrganiK platform have received consistently positive feedbacks from the participating SMEs. However, there are certain

features in relation to these capabilities which need to be improved, by way of extensions to functionality, or improvements to performance. For example, it would be highly desirable to decrease the effort required for the initial set-up of the platform, which currently involves some rather specialized manual work for bootstrapping the taxonomy learning component with initial data. Specifically, it would be preferable if administrators could simply provide the text sources on which the taxonomy learning component is to be statistically trained, and let the system handle the process of producing and registering the underlying SKOS taxonomy as appropriate. Similarly, it would be highly desirable to improve the performance of the content analyser component by employing some third-party open source libraries for Natural Language Processing which are more optimized than the ones currently employed.

2. **Exposing certain key OrganiK platform components as services (web APIs):** The participating SMEs in the OrganiK project see great potential for the OrganiK KM platform as a software system which can support both (i) the development of vertical social business applications as extensions layered on top of this system; (ii) the enhancement of external social business applications or other external information systems with OrganiK capabilities for social networking and intelligent information processing. Both usage scenarios call for scaling-up the OrganiK KM platform to a true software development platform in the true software engineering meaning of the term (i.e. a piece of software that third parties can build on by extending it, adding functionality, or integrating it to other software systems). This means that certain modules of OrganiK which are presently tightly-coupled with the platform's core (i.e. the content management system) will need to be decoupled from it and exposed as reusable services through web APIs. This will allow the respective capabilities to be accessible to software systems that are either layered on top of the platform (extensions) or situated outside the platform (integration).

3. **Re-packaging and re-configuring the OrganiK platform distribution so as to enable deployment on a virtualised cloud infrastructure:** The KnowHow consortium sees great value in leveraging the advantages that the cloud computing model has to offer today, and aims to "cloudify" the OrganiK KM platform such that it creates a Knowledge Management Platform as a Service (KM-PaaS) system. The key objective of the KnowHow consortium is to transform the OrganiK platform into a cloud-based knowledge management platform for social business. The KM-PaaS approach offers many advantages for knowledge-intensive SMEs. Besides technological capabilities, such as multi-tenancy, platform independence, scalability, and elasticity, cloud computing facilitates reduced initial costs that lower the entry barrier for SMEs to deploy an extensive knowledge management solution. To achieve this, we have re-packaged the existing distribution of the OrganiK platform in a form that is “cloud-ready”. More specifically, creating appropriate Virtual Machine Images to allow the platform to be hosted on a third-party Infrastructure as a Service (IaaS) provider like Amazon Elastic Cloud Compute or Rackspace Cloud Servers, with elastic scalability and automatic load balancing capabilities. This requires reconfiguring the platform such that the storage back-end can be deployed across a number of different cloud storage service providers, and likewise, all the exposed services can be deployed to appropriate containers within a Virtual Machine.

By carrying out **four trials** of our cloud-based knowledge management platform for social business, in the fields of **collaborative innovation, social enterprise, supply chain networking, and social media marketing**, the KnowHow consortium has validate and demonstrated that the OrganiK KM system can be extended to the broad market and employed by a variety of SMEs from different knowledge-intensive business areas.

1.2.3 Technological and business objectives of the project

KnowHow aims to demonstrate the scale-up of the OrganiK KM platform from a research prototype into an industrial-grade Knowledge Management Platform as a Service (KM-PaaS) that supports social business applications. Through the performed demonstration activities the KnowHow consortium aimed to bridge the gap between the research results produced by the OrganiK project and the requirements for future commercialization, thus preparing the ground and assisting the partners to achieve a positive impact on the market.

To reach this aim, the consortium of the KnowHow project has achieved the following objectives:

1. To analyse the SME-specific social business market environment, assess competition and the commercialization opportunities for the OrganiK KM-PaaS, and decide on a precise scope and objectives for the demonstrator.

2. To analyse and plan for scale-up activities in specific areas where enhancement of OrganiK KM platform components is critical to achieve industrial-grade strength, as well as to analyse the requirements for the trials and plan for iterative software development and releases.

3. To carry out the actual platform scale-up activities as planned, as well as implement four new technological solutions/trials on top of the KM-PaaS.

4. To validate the KnowHow KM PaaS through the four trials in the domains of collaborative innovation, social enterprise, supply chain networking, and social media marketing.

5. To develop a business plan and marketing strategy that will support the participating SMEs to commercialize the KnowHow KM-PaaS and/or the results of the trials.

6. To disseminate the project results and maximize awareness among European SMEs and beyond, taking into consideration the protection of the participants' IPRs.

7. To provide the means for allowing technical and managerial staff to be trained on using and marketing the KnowHow KM-PaaS.

8. To ensure project results of high quality, appropriate handling of IPRs, and smooth execution of the entire project through coordination, quality assurance planning and risk management.

1.2.4 Innovation from the viewpoint of cloud computing

An important aspect of innovation in the route to the market is in how the KnowHow KM-PaaS brings cloud computing into consideration, and specifically the cloud service model of Platform as a Service (PaaS). Cloud computing is an emerging paradigm that will have a transformational effect on the IT industry[2]. According to the US National Institute of Standards and Technology (NIST), cloud computing is a "model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction"[3].

[2] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R.H. Katz, A. Konwinski, G. Lee, D.A. Patterson, A. Rabkin, I. Stoica and Matei Zaharia, "Above the Clouds: A Berkeley View of Cloud Computing", UC Berkeley Reliable Adaptive Distributed Systems Laboratory, 2009.

[3] P. Mell and T. Grance, "The NIST definition of cloud computing", National Institute of Standards and Technology, Gaithersburg, 2011.

Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) represent three different service models in the wider scope of cloud computing[4]. IaaS concerns the provisioning of basic computing resources like storage, processing and networking. In PaaS, users are developers who are given access to development tools, Application Programming Interfaces (APIs) and configuration tools in order to develop and deploy their applications on the platform. Finally, SaaS refers to the concept of provisioning complete and ready-to-be-used software applications.

In all its different forms, cloud computing offers many advantages for SMEs. Besides technological capabilities such as scalability and elasticity, cloud computing facilitates reduced initial costs that lower the entry barrier for SMEs. The cost advantages associated with cloud computing are based on the fact that cloud computing users do not need to own any computing resources (server, storage, etc.) to support the capabilities they require. In addition, being able to avail oneself of installing, maintaining and managing an entire computing infrastructure, platform or service means that the entry barrier to adoption is drastically lowered.

Some academics and practitioners point out that cloud computing is more expensive over time than hosting an own infrastructure, despite the flexible cost models. This assertion is particularly valid when considering the "OpEX vs. CapEx" dilemma, which assigns cloud computing costs to the operational expenditures of a business[5]. However, this perspective on cloud computing expenses ignores the reality of SMEs who only have very limited resources and need maximum liquidity.

From the perspective of KnowHow, the Platform as a Service (PaaS) model is of particular interest. PaaS offers a combination of computing infrastructure that is accessible over the internet with tools and services allowing developers to create and deploy new applications on top of that infrastructure, or to integrate external applications. The market around the PaaS service model is growing and changing very fast[6]. The concept of PaaS is quite broad and there are in fact several different subtypes of PaaS in existence, such as application PaaS, integration PaaS, or business process PaaS, each of them serving different objectives[7]. Interestingly, however, at the time of this writing there exists no PaaS specifically addressed to the domain of knowledge management or to that of social business software.

Existing cloud-based knowledge management systems such as KMCloud, Kana and Nuxeo are not delivered as PaaS services. They are either software products that are only possible to deploy on an organisation's premises (i.e. following the traditional model of installable software products), or made available through a restrictive Software as a Service model. In other words, these products are not software platforms in the proper sense of the term, i.e. software systems that third parties can build on by extending them, adding functionality to them, or integrating them with other software systems[8].

As SMEs increasingly operate in collaborative alliances they demand cross-organizational knowledge management systems that allow them to share knowledge in a controlled fashion and integrate systems with low effort. Fusing the concepts of knowledge management platform and Platform as a Service yields a service model that supports a much easier and cost-effective way for knowledge-intensive organizations to collaborate with each other. A cross-organizational process of co-innovation will significantly benefit from an approach that not only allows stakeholders to flexibly access a KM platform from anywhere, but also allows them to easily extend the KM platform with custom applications and integrate it with external systems. With a KM-PaaS model, knowledge-intensive SMEs can push virtual collaborative alliances to the next level.

---

[4] Fang Liu, Jin Tong, Jian Mao, Robert Bohn, John Messina, Lee Badger and Dawn Leaf, "NIST Cloud Computing Reference Architecture", National Institute of Standards and Technology, Gaithersburg, 2011.

[5] Bernard Golden (2012), How to Break Down the OpEx vs. CapEx Cloud Computing Debate, CIO Magazine.

[6] Gartner Research (2011). PaaS Road Map: A Continent Emerging.

[7] Forrester Research (2011) The Forrester Wave: Platform-as-a-Service for App Dev and Delivery Professionals.

[8] David G. Messerschmitt and Clemens Szyperski. Software Ecosystem: Understanding an Indispensable Technology and Industry, The MIT Press, 2005.

1.2.5 Innovation through the trials of KnowHow demonstrator

The innovation contribution of the project is not only in relation to the proposed development of a cloud-based knowledge management platform for social business, but also, in relation to the use of this system for developing vertical social business solutions in four application domains: collaborative innovation, social enterprise, supply chain networking, and social media marketing. Through these four trials, the consortium has validated and demonstrated that the OrganiK KM system can be extended to the broad market and employed by a variety of SMEs from different knowledge-intensive business areas.

This demonstrator project has scaled-up the OrganiK platform beyond the state of the art via a four-phase-approach.

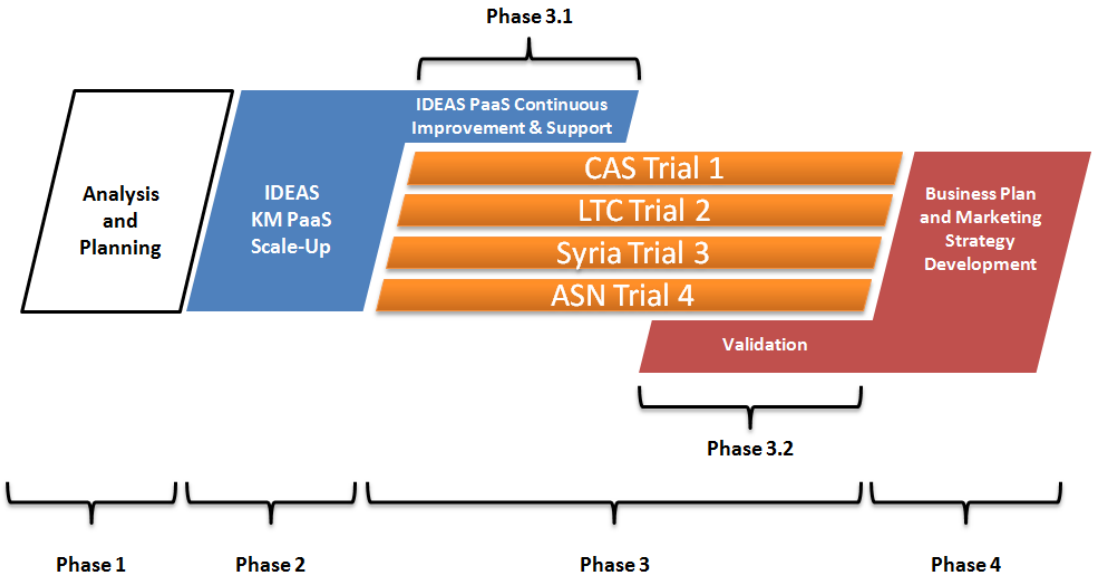


Figure 0.1 The Know How approach

During **phase 1**, all partners have analysed the business as well as technological environment in order to understand the potential and impact of the OrganiK KM-PaaS. Simultaneously and building on the results from the analysis, detailed plans have been defined for the following phases.

In **phase 2**, the OrganiK KM platform have been scaled up to an industrial-grade KM-PaaS by partner IDEAS who plays a horizontal role throughout the demonstration.

The following **phase 3** focuses on implementing, running, improving, and validating the OrganiK KM-PaaS through four SME trials in the fields of: (1) collaborative innovation, (2) social enterprise, (3) supply chain networking, and (4) social media marketing. This phase comprises two sub-phases (**phase 3.1** and **phase 3.2**). **Phase 3.1** is part of the continuous activities and horizontal support services of the PaaS scale-up, and runs parallel to the trials until these have reached a mature state. After the trials are up and running with all defined features, the validation phase kicked in, i.e. **phase 3.2**., which evaluates the KM-PaaS-based trials both quantitatively and qualitatively.

The results of the validation phase directly flew into **phase 4**, which focuses on the in-depth market preparation activities, including product definition, marketing strategy, business plan, and financial analysis. This activity will thus confer economic advantage to the participating SMEs and owners of the developed IPR.

We provide details of these achievements in the following section.

1.3 Main S&T results

The technological results achieved comprised the scale up of the OrganiK KM, by exposing the OrganiK functionality as web services and provide them on a cloud infrastructure (KM-PaaS), as well as the technological developments in the frame of the execution of four trials from different application domains. These results are detailed in the next sub-chapters. The technology and demonstration roadmap shown in the figure below shows a conceptual view of the scale up and trial activities, classifying their sub-tasks in the categories research, technology, product and market.

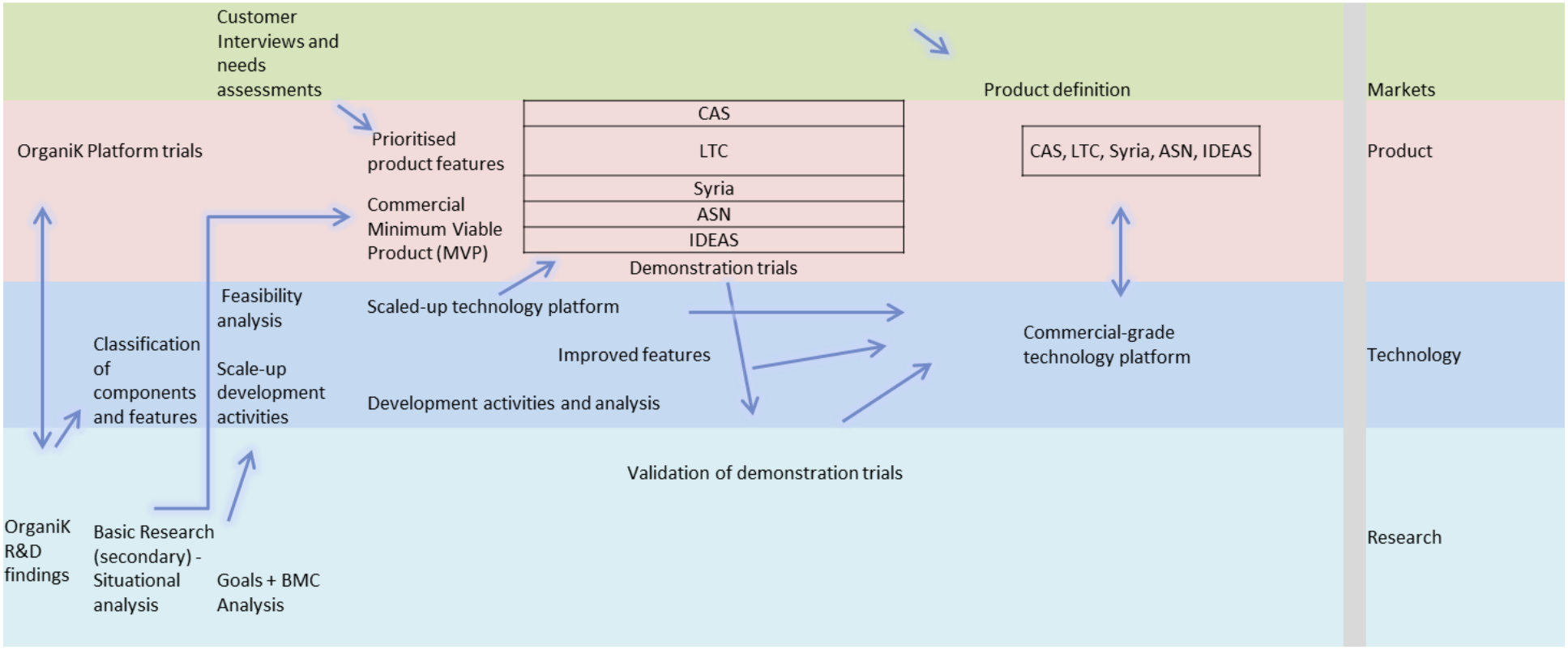


Figure 0.2 KnowHow Knowledge Management Platform-as-a-Service technology roadmap

1.3.1 KM-PaaS: scale up of OrganiK KM

The OrganiK KM system comprises several core modules and add-on modules. In the scope of the KnowHow project, we focus on six major components, which include:

- Social components – facilitate knowledge generation and sharing among users;
- Recommender component – helps with knowledge discovery by suggesting content to users;
- Content analyser component – extracts semantic metadata from content;
- User behaviour analyser component – extracts useful information from users' activity (history) which is then used for recommending relevant content;
- Taxonomy learning component – bootstraps a content classification taxonomy given a set of documents;
- Semantic search component – helps users find content based on the semantics (meaning) of the content, beyond term matching.

The next sections describe each component in more detail.

a) Social components

OrganiK comprises social components such as a wiki component, a blog component and a few others. The wiki component provides a web-based collaborative authoring tool, which enables users to collaboratively create, edit and share content (e.g., documents, pictures, binary files). It supports editing of wiki pages with a WYSIWYG (What You See Is What You Get) editor, or with Wiki syntax. Furthermore, it supports discussions for each wiki page, and is able to maintain multiple revisions (versions) of a wiki page so that one can easily revert to a previous version.

The blog component offers a personal blogging space for users. In a blog, the various blog posts are published in descending chronological order. This makes the blog component quite useful for recording the progress of a knowledge-intensive project over time, for instance. Users can comment on a blog post and interact with the content creator in a very straightforward way.

OrganiK includes a few other minor social components like social bookmarking, annotation, micro blogging and notification components, which facilitate the creation and sharing of micro content within OrganiK.

b) Content analyser component

The content analyser component employs natural language processing (NLP) techniques for extracting the most important words from a piece of text, in order to be used for annotating/tagging that content. The content analyser component uses the OpenNLP[9] framework for performing NLP processing. It mainly supports the English language and has experimental support for the German language, but with poor performance.

The content analyser component works as follows. Given a piece of text, the component performs language detection first to determine the language of the content, in order to load the appropriate language models for performing NLP. Next, it performs sentence detection, splits the text into sentences, and performs tokenization of each sentence. Next, it performs parts of speech tagging to identify what part of speech each word in a sentence is. At this point, the content analyser component uses a custom ranking algorithm to assign a rank to each word. Optionally, it can use weights in the ranking process based on an Inverse Document Frequency (IDF) measure, in order to improve the ranking. Finally, the component returns the list of ranked terms.

The content analyser component is mainly a back-end component and is used in conjunction with the recommender component for recommending tags and with the taxonomy-learning component for extracting taxonomy terms from textual content.

c) User behaviour analyser component

[9] http://opennlp.apache.org/

The user behaviour analyser component tracks and analyses the behaviour and actions of users while interacting with content that resides in the OrganiK KM system. The results of the analysis are interaction patterns, which are leveraged by the recommender component for providing personalized recommendations to users.

Each time that a user visits a page in OrganiK, the component records the date and time of the last interaction of a user with that page. These records constitute a trace of the browsing history of a user and can reveal the pages of interest to a user, along with the recency of their visits. More fine-grained interactions between users and content can be tracked too. User actions such as viewing or editing content items, viewing or posting comments can be utilized to compute a value for the importance of content items to users.

The user behaviour analyser component is a back-end component. The information produced by this component is mainly leveraged by the recommender component for personalizing the content recommendations.

d) Recommender component

The recommender component utilizes past information recorded by OrganiK about users, in order to provide personalized content suggestions to them. The aim is to help users of the OrganiK KM system to discover content such as blog posts or wiki pages that are relevant to their interest, as the volume of content increases in the system.

The recommender component has two main elements: a generic recommendation framework, which implements various recommendation algorithms, and specific recommender system implementations which use the algorithms of the framework to produce recommendations. The framework can be used to create new recommender system implementations.

The recommendation framework implements several recommendation algorithms, which work with statistics and probabilities, in order to compute similarities between users. The similarities can be then used to predict and recommend content items that could be of interest to a user. The recommendation framework only supports collaborative filtering[10] algorithms, which take into account user-item preference information and makes recommendations accordingly. The framework has been built using the Recommender API[11] of Drupal.

There are two specific recommender system implementations in the recommender component of OrganiK: i) recommendation of content based on user behaviour analysis, and ii) recommendation of tags that users can choose to annotate content, based on content analysis. For the recommender system, the information recorded by the user behaviour analyser component is used to provide accurate and personalized recommendations. Such a recommendation takes into consideration the browsing history of users and utilizes the pair “user – content item” and the recency of the user's interaction with the item.

For tags recommendation, the component combines several statistical methods to produce a common set of tags which are presented to the user, in order to help them to annotate some content. Two factors determine the suggestion of a tag, the frequency of a word in the specific document and the number of documents that have this word. Statistically rare and important words are promoted over commonly used ones.

The recommender component is one of the most useful components of the OrganiK KM system and it can be extended to implement specific types of recommendations in different contexts.

e) Taxonomy learning component

The taxonomy learning component is able to automatically bootstrap an initial taxonomy of terms based on a given set of text documents. The taxonomy is used in OrganiK to organize content into a hierarchical structure of terms. This component is very useful if there is a significant amount of content that needs to be inserted and categorized in OrganiK.

[10] S. Alag. Collective Intelligence in Action, Manning Publications, 2008

[11] https://drupal.org/project/recommender

The component analyses text documents using natural language processing and knowledge extraction techniques to create the initial taxonomy. It identifies noun phrases from sentences in the documents and for each noun phrase, it attempts to identify synonyms and concepts in external knowledge databases, such as DBpedia. The identified concepts are positioned in a hierarchy based on the narrower and broader relations in the external knowledge database. The component can follow this process only once to build a taxonomy and cannot evolve a previously built taxonomy with new terms.

The taxonomy learning component has been implemented as a standalone command-line utility, which takes a set of text files as input and generates a new file that contains the taxonomy in SKOS format[12]. The file has to be manually imported in the OrganiK KM system.

f) Semantic search component

The semantic search component provides semantic search functionality, by leveraging the relationships among terms in a predefined taxonomy. It supports browsing, searching, retrieving and displaying textual content. The component acts as an enterprise text search engine for the OrganiK KM system and supports searching and indexing in the English language.

The semantic search component has been implemented on top of Drupal's search API and the Apache Solr enterprise search platform. The semantic search function has been implemented as an extension to the regular searching functionality by leveraging the taxonomy of terms maintained in OrganiK. The user can enter a taxonomy term and the search will include results that contain or have been annotated with broader and narrower terms of the given term. Therefore, semantically related content is being included in the search results.

The semantic search component supports faceted search that allows a user to easily navigate and filter the search results. In addition, alternative spellings are supported, in order to expand a user's query based on the given term.

State of the OrganiK KM System as a research prototype

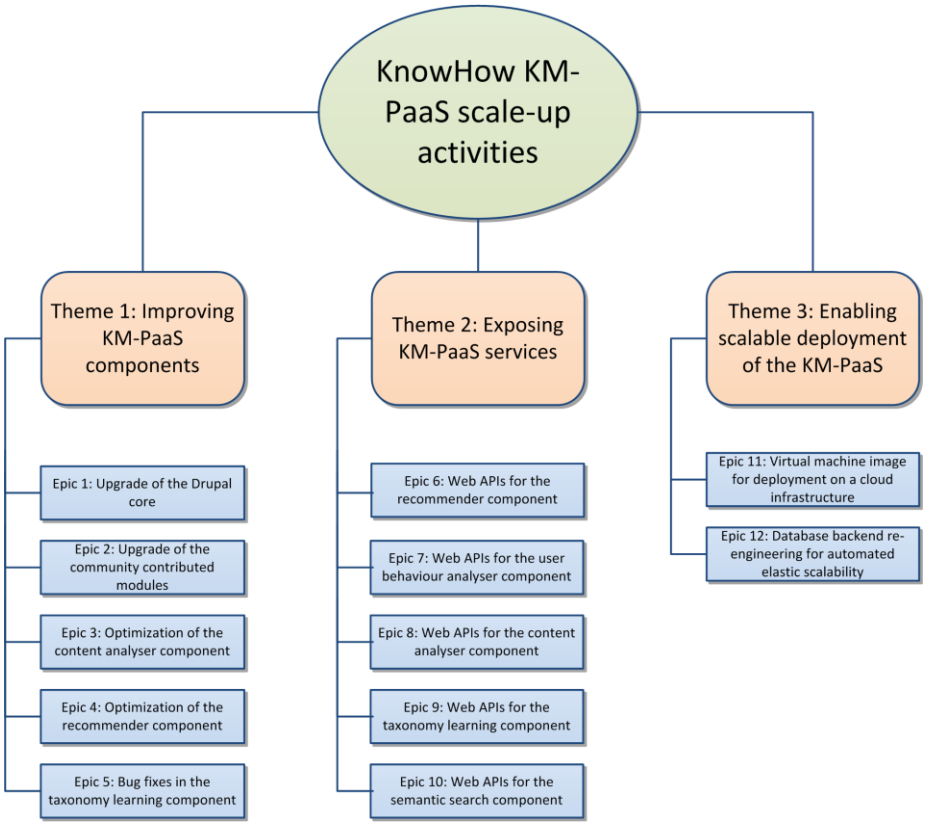
The capabilities of the OrganiK KM System make it an asset with significant exploitation potential in the context of social business. However, the system was still in the state of a research prototype. During the OrganiK project the system was developed and evaluated on a small scale. Despite the technical improvements which were identified and carried out during the partner-specific case studies, there are specific enhancements which are required before being able to bring products to the market based on this technology. These enhancements concern the following scale-up activities:

- Carrying out targeted improvements and optimizations in specific components. For instance, extensions to functionality, or improvements to performance.
- Exposing certain components as services (web APIs). This will allow the capabilities of these components to be accessible to software systems that are layered either on top of the KM-PaaS platform (extensions) or outside the platform (integration).
- Re-packaging and re-configuring the OrganiK KM System distribution. This will allow the platform to be hosted on a third-party virtualized cloud infrastructure (IaaS) provider, with elastic scalability and automatic load balancing capabilities.

Methodology for managing the scale-up activities

For the needs of planning the scale-up activities in KnowHow, we have chosen to conceptualize requirements around the high-level notions of *Themes* and *Epics*, as illustrated in the figure below.

[12] http://www.w3.org/2004/02/skos/



As shown above, 'Improving KM-PaaS components' represents a strategic *Theme* for the KnowHow project. 'Optimization of the recommender component' represents an Epic mapped to the said *Theme*. 'Improving KM-PaaS components' is mapped to five *Epics* in total. The same holds for Theme 2 'Exposing KM-PaaS services'. Theme 3 'Enabling scalable deployment of the KM-PaaS' maps to two Epics.

In the following section, we will give an overview of the three *Themes*. Detailed documentation of the *Epics* according to the MoSCoW method[13] can be found in deliverable D2.1.

Theme 1: Improving KM-PaaS components

The social networking and intelligent information processing capabilities of the OrganiK platform have received consistently positive feedback from the SMEs that participated in the OrganiK project. However, there are certain features in relation to these capabilities, which need to be improved, by way of extensions to functionality, or improvements to performance. For example, it would be highly desirable to decrease the effort required for the initial set-up of the platform, which currently involves some rather specialized manual work for bootstrapping the taxonomy learning component with initial data. Specifically, it would be preferable if administrators could simply provide the text sources on which the taxonomy learning component is to be statistically trained, and let the system handle the process of producing and registering the underlying SKOS taxonomy as appropriate. Similarly, it would be highly desirable to improve the performance of the content analyser component by employing some third-party open source libraries for Natural Language Processing which are more optimized than the ones currently employed.

Theme 2: Exposing KM-PaaS services

[13] D. Clegg and R. Barker. Fast-Track: A Rad Approach (Case Method). Addison Wesley. 1994.

The SMEs participating in KnowHow see great potential for the OrganiK platform as a software system which can support both (i) the development of vertical social business applications as extensions layered on top of this system; (ii) the enhancement of external social business applications or other external information systems with capabilities for social networking and intelligent information processing. Both usage scenarios call for scaling-up the OrganiK platform to a proper software development platform (i.e. a piece of software that third parties can build on by extending it, adding functionality, or integrating it to other software systems). This means that certain components of the platform which are presently tightly-coupled with the platform's core (i.e. the Drupal content management system) will need to be decoupled from it and exposed as reusable services through web APIs. This will allow the respective capabilities to be accessible to software systems that are either layered on top of the platform (extensions), or situated outside the platform (integration).

Theme 3: Enabling scalable deployment of the KM-PaaS

The KnowHow consortium sees great value in leveraging the advantages that the cloud computing model has to offer today, and aims to "cloudify" the OrganiK KM platform such that it creates a Knowledge Management Platform as a Service (KM-PaaS) system. The KM-PaaS approach offers many advantages for knowledge-intensive SMEs. Besides technological capabilities, such as multi-tenancy, platform independence, scalability, and elasticity, cloud computing facilitates reduced initial costs that lower the entry barrier for SMEs to deploy an extensive knowledge management solution. The cost advantages associated with cloud computing are because cloud computing users do not need to own any computing resources (server, storage, etc.) to support the capabilities they require. Through flexible cost models, cloud users are usually billed on a highly granular level for the usage (e.g. server hour, storage per month, etc.) over a certain period of time. To achieve this, the enhancements necessary are to re-package the existing distribution of the OrganiK platform in a form that is "cloud-ready". More specifically, creating appropriate Virtual Machine Images to allow the platform to be hosted on a third-party Infrastructure as a Service (IaaS) provider like Amazon Elastic Cloud Compute or Rackspace Cloud Servers, with elastic scalability and automatic load balancing capabilities. This requires reconfiguring the platform such that the storage back-end can be deployed across a number of different cloud storage service providers, and likewise, all the exposed services can be deployed to appropriate containers within a Virtual Machine.

1.4 Set up and execution of four KnowHow demonstration trials

1.4.1 Collaborative Innovation – CAS netWorks

The CAS trial in the KnowHow project focuses on social business platforms for collaborative innovation targeting associations. Within the scope of the project, CAS has implemented, operated, and validated a collaborative innovation platform enriched with the KnowHow capabilities for the largest ICT-oriented association in Baden-Württemberg.

The platform for supporting and facilitating social business collaboration among associations has functionalities for different roles. On the one hand, the organizers working at the trial association, and on the other hand the members. Members can be associations, companies, employees or people looking for job opportunities. In the following, the use cases based on the requirements of the different user groups are briefly presented.

Use Case 1: "Member Clustering": The organizers at CyberForum want to identify groups of members with certain characteristics.

Use Case 2: "Member Identification": Users want to identify members with specific skills. This can be useful to find partners for a project, i.e. collaboration spaces, (with skills the member itself does not have).

Use Case 3: "Identifying Competitors": Members want to identify competitors, i.e., members with semantically identical skills.

Use Case 4: "Semantic Search": Members want to identify different items within the CyberForum and within Collaboration Spaces. Items can be knowledge artefacts, knowledge assets, events, or other members.

The integration of the KnowHow KM-PaaS query engine within the CAS trial is shown in the following figure.

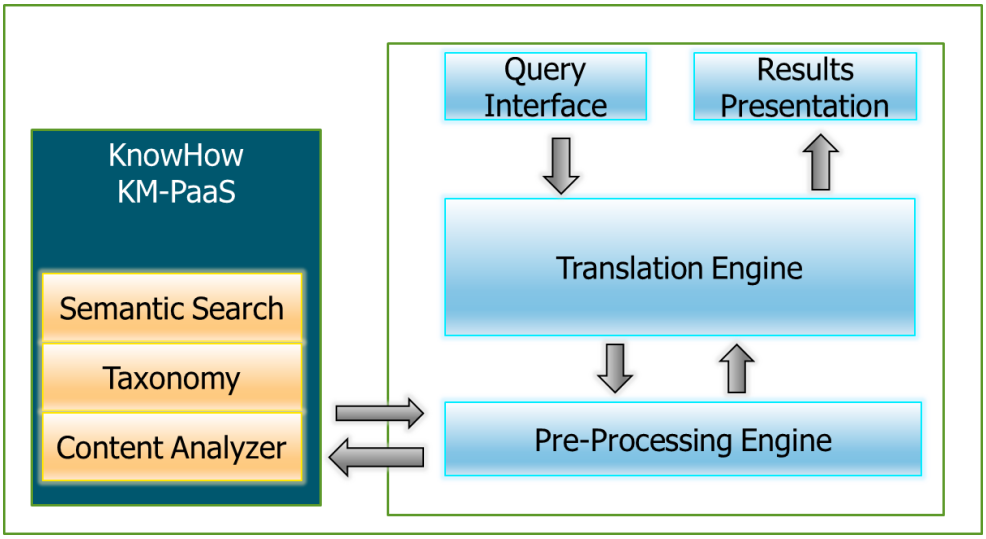


Figure 0.4 KM-PaaS Integration – Using the Query Engine

1.4.2 Social Business – LTC KnowHow KM-PaaS

Within the scope of LTC's trial the aim was to demonstrate the improved usability of knowledge management processes for the large community of language service providers (LSPs). LSPs can benefit from enhanced technology applications to support the core, collaborative business of multilingual services. LTC aimed to demonstrate the KnowHow KM-PaaS platform as a potential feature of any workflow and business process system for the multilingual organisation.

In the context of the social enterprise, the LTC trial has developed, implemented and validated the KnowHow applications for the collaboration portal of the existing LTC management platform (see figure below). The system is a business process management system for the language industry, to coordinate and increase efficiency in carrying out translation and other types of language services projects.

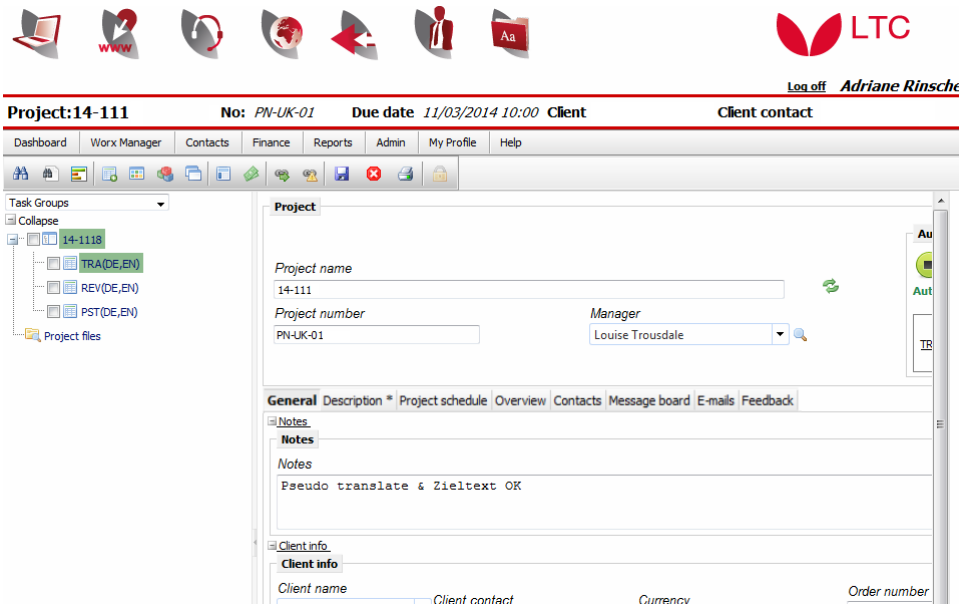


Figure 0.5 LTC Worx business management system, project management portal

Using content analysis, recommendation, and semantic search technology, knowledge assets from customers and external LSPs can be intelligently captured, filtered, stored, and reused. The entire process of quality multilingual services becomes integrated, collaborative, and more efficient. Importantly, the social components of KnowHow have been adapted to increase interactions with internal staff members such as client account managers and colleagues with specific tool or subject-area knowledge.

Certain social enterprise features will be demonstrated in accordance with commercial requirements. The validation of demonstration trials will be assessed as knowledge transfer becomes more integrated within internal processes and external project management. The different stakeholders common to the desired solution will include the following main user roles:

- Project managers undertake multitasking and organising different aspects of projects and orders, including deadlines, mediating between stakeholders, ensuring problems are resolved and leveraging linguistic tools. Extra effort experienced in coordinating requests, applying useful information for a specific project, utilising language technology tools, and finding information, results in lost operational performance.
- Internal suppliers provide language services such as translation and revision among others. Internal service providers are knowledge workers from the multilingual service sector, with vast experience in generic translation services and technologies. They execute translation projects in close collaborative setup with the customer and the project manager.
- External suppliers are freelancing language service providers with expertise in certain subject areas. They work closely with internal staff and are part of an ongoing collaborative network. Their services are often required for specialized translation projects involving particular technology tools or managing specific multilingual assets.
- Clients/requesters can be closely involved in many translation projects. They require specific production conditions and are familiar with knowledge regarding context, terminology and operational needs. These clients facilitate the customization of important project-specific factors and also influence the monitoring and communication of projects.

The following key use cases help to explain the requirements of the different user groups.

- **Finding assistive information** such as tool guides for suppliers, company task guidelines and best practices (such as how to create an invoice), client-specific processes, similarly translated material, terms and other management queries. Semantic search will enable users to find a wider and more accurate range of resources. This scenario will involve the use of semantic enrichment to better access information using categories and taxonomies and enhanced semantic search to present relevant information.

- **Suggesting resources of interest** based on user behaviour. According to which items are viewed by the user, the areas of interest through subscription to particular terms, and the sharing of resources made easier by social bookmarking, more useful information is made available at relevant times. This would leverage the KnowHow KM-PaaS to use knowledge gained whilst performing previous projects and reduce effort in obtaining similar results.

- **Collaboration on important tasks** for translation, terms and other management queries. By using wiki pages and leadership of senior users, important resources for terminology or client data can be recorded and checked as the standard information. This will remove duplication and ensure more consistent quality control.

- A desired use case, which may need to investigate technical capabilities further concerns suggesting resources, **based on content analysis**. For instance, KnowHow content analysis will help identify the topic of a document, and then recommend useful knowledge assets, resources and if possible suitable service providers related to the topic and requirements in question; then, suggest a suitable supplier based on the corresponding content subject.

The next figure explains a common event-driven process. During the initial phase of the language project a client requests a translation specifying operational requirements and the quotation, schedule and supporting materials are defined before the project is created by the project manager.

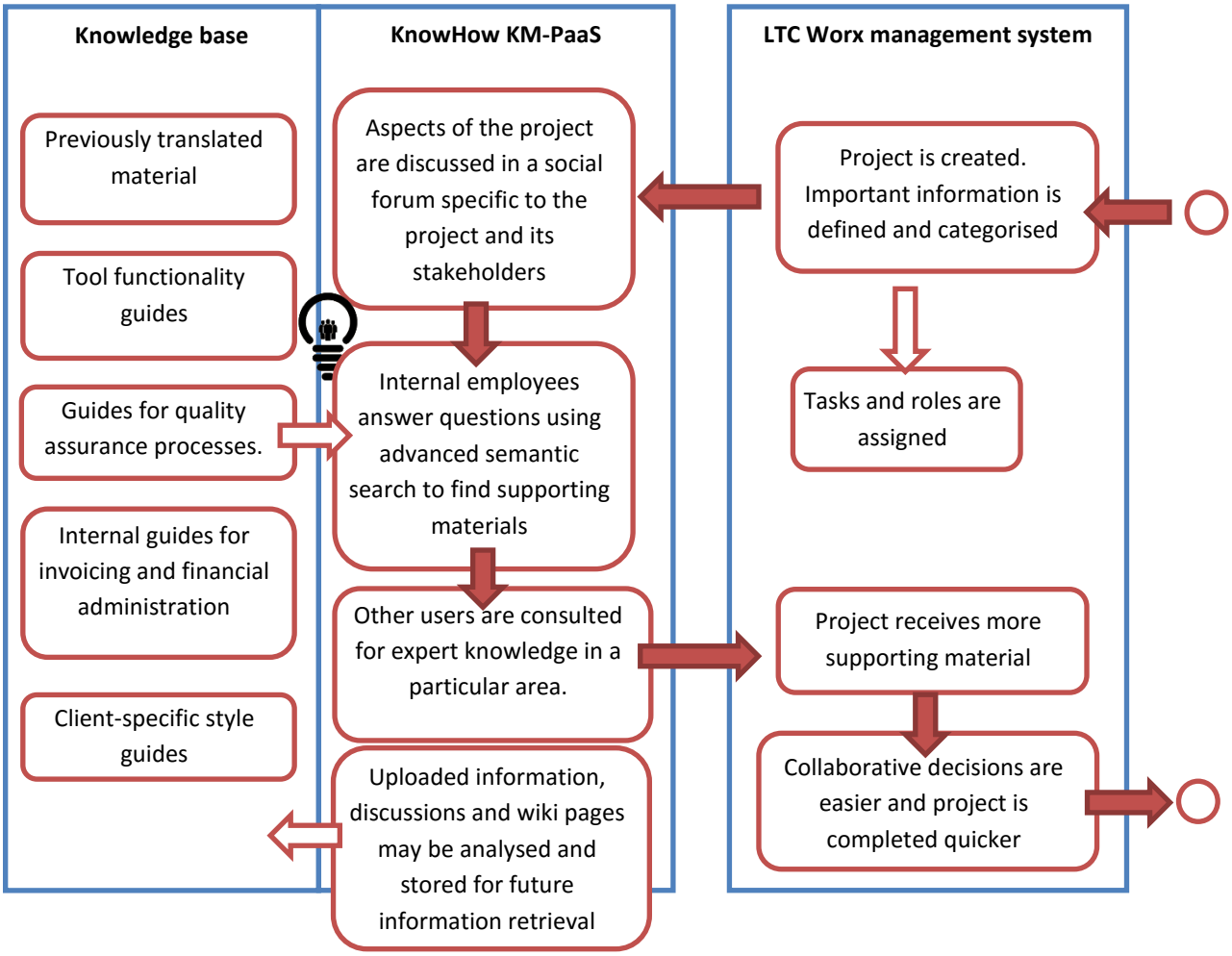


Figure 0.6 Common event-driven use case process chain

1.4.3 Supply Chain Networking

Syria Informatica produces and develops an internet portal of the genus Enterprise 2.0 with CMS Content Management System platform and comprehensive features for e-commerce. With the KnowHow Project, Syria Informatica aimed to demonstrate the KnowHow KM-PaaS features, when integrated into its own CMS platform, to exploit the benefits of its intelligent information analysis features. To carry out this demonstration, and to organize the launch of this platform, Syria Informatica has performed its trial in collaboration with consortium of 70 companies, dealing within the automotive aftermarket.

The trial offered the following opportunities for Syria, regarding commercialisation in the near future

- Enable the implementation of a reference model, to be used for the future market launch of the platform.
- Allow Syria to check the innovative intelligent analysis features of KnowHow KM-PaaS, adopting the system in the context of a real operating company with a before-after analysis of the results.
- Allow Syria to implement the platform and its components, adapting them in a broader context, though always focused on SMEs.
- Investigate the resolution of the taxonomy issues related to the adoption with Italian language.
- Allow the analysis of the impact and the changes in the business processes of enterprises adopting the tool
- Allow to test and improve the overall usability of the system

In this context, we have collaborated with two main stakeholders

The **sales and marketing department** is responsible for all marketing activities (outgoing and incoming) with the aim of increasing the growth of the consortium in the whole Italian country and for the provision of services and business opportunities to the members. Its main task is the selection, negotiation, adaptation and sharing of information regarding products and commercial agreements, which are made available to members through the consortium informative portal.

The network of the members: they are the real main beneficiaries of the consortium activities. The companies of aftermarket industry, become part of a collaborative network that provides them with consulting services, marketing, representation and commercial brokerage together with support and after sales service.

The process at the base of Syria trial concerned the management of communications of trade agreements and promotions between consortium suppliers and consortium members. This process can be improved thanks to the KnowHow PaaS features about the recommendation of the best information / promotions / products / commercial agreements to the members that fulfil the needs of individual members, basing them on the analysis of their profiles and their needs thanks to their social activities into the consortium informative portal.

The implemented work process takes care of the simplification and automating of information management process for promotions and campaigns, thanks to the intelligent analysis and recommender system which is equipped with the KnowHow KM-PaaS platform.

The user's profile analysis system, assess the interests of each member and will directly suggest them with relevant information / promotions / campaigns.

Suppliers are able to upload their relevant information directly into the platform, therefore lightening the workload of the marketing department in terms of the tasks of collecting and disseminating information. Information entered directly by suppliers can be automatically forwarded / recommended to each members according to their profile / interests.

Consortium members are using a tool that shows in a personalized way the most relevant information, reducing the waste of time for consultation and improving the timing and quality of their activities.

The marketing department is able to access all this information to improve the quality of its analysis. The entire process can be automated, leaving more time to improve the quality of agreements with suppliers and management / support / assistance to the members.

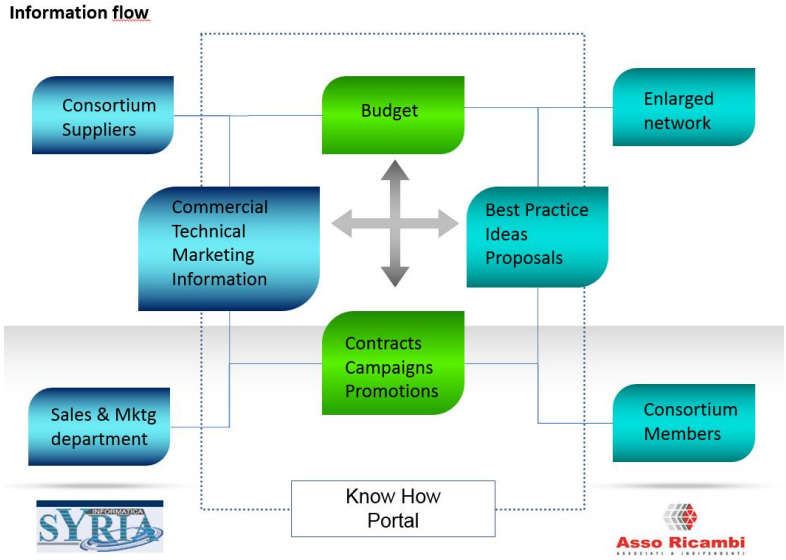


Figure 0.7 Information flow within the SYRIA trial

1.4.4 Knowledge-Based Trade Fair Services

The trial of ASN targets trade fair organizers. ASN has set up and operated a trade fair platform, providing knowledge management services to trade fair exhibitors and trade fair visitors. The platform has been developed in collaboration with the largest trade fair organizer in the area, and tested in the frame of two exhibitions delivering feedback and allowing for improvements in an iterative manner.

The scope of the trial was the collection of information of exhibitors and visitors, creation of associated profiles and matching offers and requests in order to increase the impact of the fair for all stakeholders (i.e. visitors and exhibitors) and secure existing customer base for the trade fair organizer. The user might create visitor profiles directly or with linkage to social media profiles. Detailed information of the visitor is collected during an initial validation process. At the same time, the trade fair organizer might use the analysis of the collected information for improving future fairs, increase internal process efficiency and perform targeted social marketing activities.

The ASN trial has implemented four use cases, namely "employ social media", “personalization and recommendation”, "utilize semantic search" and "arrange a meeting". For this, it has utilised the KM PaaS services for recommendation, semantic search and content analysis. The high-level architecture of the ASN trial is shown in the following diagram.

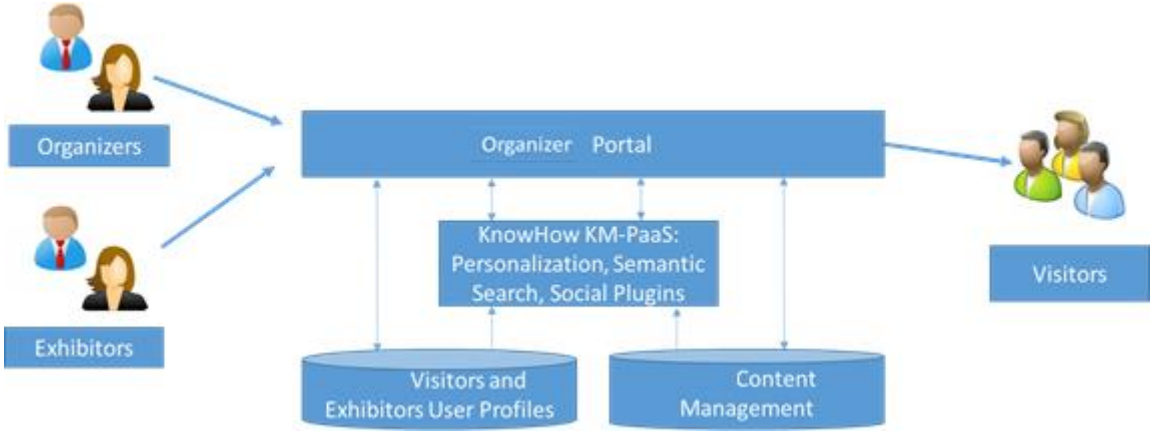


Figure 0.8 Architecture of ASN trial

In the following, we will describe the solution of the ASN trial. The system serves three types of stakeholders: (a) the trade fair organizer, (b) the trade fair exhibitor, (c) the trade fair visitor.

The trade fair organisers

The trade fair organisers manages and organizes the fairs, certifies visitors and exhibitors and controls the overall platform. He is interested in adding value to the fair, according the demands of his customers, his main success criterion is the impact created for exhibitors. Therefore, he is interested in employing social media to create customer profiles, rate exhibitions and provide feedback during and after a trade fair. The successful demonstration of the ASN trial will result in making the fairs more dynamic, allow for better meeting the expectations of their customers (exhibitors and visitors) and deliver valuable feedback for improving future trade fairs.

Trade fair exhibitors

Trade fair exhibitors are participating in the fair. Their main interest is to find new customers, initiate new business or to disseminate their brand. Hence, they are interested in contacts with the right persons, i.e. visitors of the exhibition. For this, they need to populate and maintain an updated profile with their offers, capabilities, products or interests.

Trade fair visitors

Trade fair visitors are representing companies from the domains presented in the exhibition. They visit the fairs not mainly to keep up to date, but seeking different types of business collaborations. To this extent, it is important to find suitable partners, i.e. exhibitors with the appropriate portfolio. Hence, they need to be informed on exhibitor profiles in a certain level of detail, i.e. description of their offerings.

Apart of some administrative functionality, the ASN trial is mainly supporting the following three use cases:

Use Case 1: "Employ social media": The trade fair organizers will link their fairs with social media for performing marketing activities, validating visitors, prefill customer profiles, offer rating of trade fairs and collect feedback.

Use Case 2: "Personalisation and recommendation": Exhibitors will complete and confirm their profile, including type of business, expertise, company information, product portfolio, contacts in the booth etc. In combination with visitor profiles collected during the certification process, matching algorithms will deliver recommendations to visitors, i.e. suggestions which fair and which exhibitor or booth they should visit.

Figure 0.9 Mock Up screen on Use Cases 1 and 2

Use Case 3 "Semantic Search": In regular terms, actual information will be analysed and main concepts / terms will be identified. When users search for information, a semantic search is performed and related content is displayed e.g.: Searching for fairs with the word 'computing' will provide as result also fairs in 'electronics'.

Figure 0.10 Mock Up screen on Use Case 3