



Federal Ministry  
of Economics  
and Technology



**THESEUS**

New Technologies  
for the Internet of Services



Innovation Policy, Information Society, Telecommunications

## The THESEUS Research Program

New Technologies for the Internet of Services

[www.theseus-programm.de](http://www.theseus-programm.de)

**Compilation and design**

Scholz & Friends Agenda

**Print**

Richard Thierbach, Buch- und Offset-Druckerei

**Photo credits**

BMW, Corbis, Getty, Fotolia, Masterfile

**Publisher**

Federal Ministry of Economics and Technology  
Public Relations  
10115 Berlin

**In collaboration with:**

THESEUS Program Office  
c/o empolis GmbH  
An der Autobahn  
Postfach 180  
33311 Gütersloh

**Last revised**

November 2008



The Federal Ministry of Economics and Technology has been awarded the berufundfamilie® audit certificate for its family-friendly HR policy. The certificate is granted by berufundfamilie eGmbH, an initiative of the Hertie Foundation.



Federal Ministry  
of Economics  
and Technology



**THESEUS**

New Technologies  
for the Internet of Services

Innovation Policy, Information Society, Telecommunications

# The THESEUS Research Program

New Technologies for the Internet of Services

# Table of Contents

Preface .....	05
1 THESEUS – New Technologies for the Internet of Services .....	06
2 ALEXANDRIA – knowledge platform on the Internet .....	08
3 CONTENTUS – technologies for the multimedia library of the future .....	10
4 MEDICO – smart searching in medical databases .....	12
5 ORDO – order in a digital world .....	14
6 PROCESSUS – improving how corporate knowledge is leveraged .....	16
7 TEXO – infrastructure for Web-based services .....	18
8 THESEUS – basic technologies .....	20
9 THESEUS – Joint Research .....	23
10 THESEUS Talents 2008 – ideas competition to foster young research talent .....	24
11 THESEUS SME 2009 – call for SME-operated projects .....	25
12 Partners in the THESEUS program .....	26

## Preface

Three letters spell our future: ICT. With information and communication technologies (ICT), we can secure jobs and safeguard economic prosperity in the years to come. Therefore, we are promoting development in precisely this area. To this end, the Federal Ministry of Economics and Technology (BMWi) launched the THESEUS Research Program, which acts as a critical link between scientific excellence and economic and business potential. THESEUS is directed towards new technologies for the Internet of Services, and the program was declared a beacon project at the first IT summit of the Federal Government.

We will soon be presenting interim results of THESEUS at the third IT summit. Market leaders from the world of enterprise software, the manufacturing industry and medical technology are actively involved in the THESEUS program. However, not only big names play a role in the project. THESEUS is also specifically geared towards small and medium-sized enterprises, an aspect of the program to which we attach considerable importance. With the complementary THESEUS SME incentive measure, these midmarket businesses are encouraged to submit their ideas to shape the project going forward.

Moreover, as part of THESEUS, leading research institutes and elite universities are working to develop semantic technologies that better harness and utilize complex knowledge on the Internet. And while we greatly value cutting-edge research, we are also deeply committed to nurturing the next generation of researchers. The THESEUS TALENTS ideas competition opened the research program to young, talented individuals who want to become part of the Internet of the future.

With a total financial volume of some 200 million euros, approximately half of which is provided by the BMWI, THESEUS is the biggest ICT research project of the Federal Government. With THESEUS, we want to become the leaders in information and communication technologies, which would benefit Germany as a business location on the whole.

The results of the THESEUS project will bring particular advantages to our citizens by facilitating better knowledge access and providing new and more efficient services on the Internet. Enterprises

can exploit the results of THESEUS to optimize their business processes and develop new services and business models. This will create new markets, and companies can tap into new business fields. With our funding program, we ultimately want to provide incentive for the development and expansion of R&D in businesses to stimulate economic growth and job creation. The research activities and further sophistication of the technologies for the Internet of Services will provide fresh impetus to the research sector in the long run.

This brochure provides you with an overview of the THESEUS program. You are invited to become an active part of THESEUS and to benefit from everything the program can offer.

**Federal Ministry of Economics and Technology**

## 1

## THESEUS – New Technologies for the Internet of Services



**The Internet provides us with new data and information every day. However, with increasing data volumes we are faced with the challenge of giving users easy and efficient access to this immense knowledge base. Partners from the world of science and business are working together on finding a solution to this problem under the framework of THESEUS, Germany's biggest IT research program.**

Just as the legendary Greek hero Theseus found his way out of the Minotaur's labyrinth, so too should the solutions developed in the THESEUS research program point the way through the growing volume of data present on the Internet. Under the umbrella of the THESEUS program, 30 research partners from science and business are developing new technologies and applications with the aim of easing access to information, linking data to new knowledge and creating the basis for the development of new services on the Internet.

The technologies developed in the course of the THESEUS program hold enormous potential as they are key to making an "Internet of Services" an actual reality. While services such as online shopping, booking flights or research services are only available today as individual instances on the Internet, such offerings could be automatically grouped and linked

in an Internet of Services. In the future, users would simply have to tell their PCs "I want to move from Berlin to Hamburg", and the computer program would automatically find suitable listings for apartments, coordinate the move and even register the change of address with the local authorities.

### **Semantic technologies as the basis**

To make this vision a reality, new semantic technologies are being developed in the THESEUS program that are capable of analyzing, classifying and linking information on the basis of the information content. In this way, not only can computer programs find information using keywords or contextual fragments, as is common practice today, but can also automatically determine the meaning of the information, put the data in the context of other items of information, model the information as classification systems (ontologies), and apply specific rules to draw logical inferences from the contents.

The development of basic technologies is pivotal to any progress made in this area. For example, in the context of the THESEUS program, participating researchers are working on innovative processes for the automatic generation of metadata – similar to key data or indices – to accelerate multimedia

document processing, and are also concentrating on the development of new graphic user interfaces.

### **Trials in application scenarios**

To ensure new products and services are created as quickly as possible using the basic technologies, partners participating in the project put prototypes of the products and services to the acid test in six application scenarios. These scenarios determine how the technologies can be exploited for innovative tools, services and business models.

In the ALEXANDRIA application scenario, the Web 2.0 approach – which involves linking users in virtual communities and the growing importance of user-generated content – is extended to include innovative knowledge management methods so users can more easily access content on the Internet to suit their individual needs.

The CONTENTUS application scenario uses semantic technologies for the digitization of cultural assets in the form of texts, graphics, and audio and video recordings. The goal of CONTENTUS is to better preserve cultural assets for future generations, and facilitate rapid and comprehensive access to this knowledge across multimedia formats.

As part of the PROCESSUS application scenario, researchers examine how the knowledge available in a company can be better leveraged. To this end, a platform is being developed that integrates resource planning and business process management, and makes it easier to compare products, solutions and business partners. As a result, companies can cut their costs and optimize their entire business processes.

The MEDICO application scenario aims at improving the quality of medical diagnoses on the basis of imaging processes. Semantic technologies support physicians in rapidly and reliably detecting abnormalities in medical image data, and improve their diagnoses and therapy plans by quickly analyzing image databases and specialist literature.

ORDO, the fifth application scenario, concentrates on the development of innovative approaches

and technologies to organize digital information. The research is geared to developing tools that organize information automatically, making it easier to handle and analyze large data volumes. Anyone working with large volumes of difficult-to-manage data, such as patent information, on a regular basis can benefit from such tools.

To ensure that the offerings on the future Internet of Services can be used in a secure but open environment, researchers in the TEXO application scenario are developing an integrated platform for provisioning, managing and combining Internet-based services. This platform will make it easier for providers and users alike to develop and avail of these services, and it constitutes an important milestone on the road to the next generation of the Internet.

### **Key impetus for Germany as a business center**

With the research conducted in the THESEUS program, new technologies for the Internet of Services are created, and complex knowledge is mapped for the first time ever. As a result, the participating researchers are enabling the easy utilization of knowledge available on the Internet, and are laying the foundations for the development of new applications. In this way, THESEUS not only strengthens Germany's position in the area of Internet research, but also ensures that businesses in Germany are helping to shape the Internet of Services with innovative products and services, and will benefit from this market of the future.

In competitions it holds, THESEUS invites external scientists and medium-sized enterprises to take part in this program at an early stage and put the intermediate results to work.

The THESEUS program will last five years and receive some 100 million euros in funding from the Federal Ministry of Economics and Technology (BMWi). A further 100 million euros are mobilized by participating partners from the fields of industry and research, with the result that approximately 200 million euros in total are being injected into innovative, future-oriented research activities – technologies for the Internet of Services and for Germany as a center for business and investment.

## 2

## ALEXANDRIA – knowledge platform on the Internet



The city of the same name on the Egyptian coast housed the most famous library of ancient times. Its precise cataloging system was a pioneer in easing information access. Inspired by this role model, researchers in the ALEXANDRIA application scenario are developing a knowledge platform on the Internet that combines completely innovative interaction features to make it easier to tap and utilize knowledge.

As its starting point, the ALEXANDRIA application scenario focuses on the growing importance of Internet users in Web 2.0. Apart from simply using search engines, e-mail and chat services, Internet users today also actively make their knowledge available on the Web. Common examples of this new mode of interaction include blogs, the Internet encyclopedia

Wikipedia or social networks like Facebook or Xing. These platforms would never be possible without users who provide and exchange knowledge and information.

### Compiling and accessing knowledge

The aim of the researchers working on ALEXANDRIA is to better link the knowledge available in Web 2.0 and make it easier to access. For this purpose, a knowledge platform is being developed that not only gathers information and integrates the knowledge generated by users, but also helps better manage the growing volumes of data. This knowledge platform is dedicated to supporting users publish, process and search for content. The platform itself is to be an intuitive, easy-to-use system. To this end, researchers

are developing semantic technologies and classification systems capable of processing large volumes of data and logically linking them to one another.

### Integrating users' knowledge

A further objective of ALEXANDRIA is to integrate the Internet users and their knowledge as an integral part of the planned intelligent knowledge platform. One possible way of achieving this goal is through tagging, the term used to assign keywords to content, such as documents or images. Such data enrichment – also known as metadata – can be performed manually by users, or automatically by the system. In this way, technological solutions can detect and correct typing errors, while people can better and more quickly correct incorrect meanings of similar or identical terms. The knowledge platform created in the ALEXANDRIA application scenario aims to draw on the specific strengths of both approaches.

### Comprehensive access to knowledge

To demonstrate the possibilities such a knowledge platform holds, ALEXANDRIA's first sample application focuses on the topic of "famous Germans". Within an easy-to-navigate user interface, researchers demonstrate how the technologies developed make it easier to access and retrieve knowledge and information. A number of famous people are displayed on a screen. If users want to quickly find out who one of the people shown is, they can read automatically extracted data and five key facts on that person. In a continuous voting process, users decide what information appears under the five key facts by individually ranking the relevance of the data. More in-depth information that users can also view can also be rated in detail. Involving the users and tapping their knowledge improves the quality of the information presented, and ensures that the data users consider to be most relevant are actually displayed.

To demonstrate the possibilities of linking data content, a "relationship browser" shows the relationships of the eminent person shown to other famous contemporaries. The relationships presented are automatically extracted from the data records. Users can correct relationships that were incorrectly assigned

by the automatic process, or add relationships the browser did not find. This example illustrates the pivotal role users play in creating, enriching and rating content in ALEXANDRIA. Using the technologies developed, it should be easier to harness this user knowledge with semiautomatic processes and organize information as it is needed, with the most important items first.

### Objectives of ALEXANDRIA

- ▶ Develop an easy-to-use, intuitive knowledge platform on the Internet
- ▶ Actively involve users
- ▶ Support users in publishing and processing content
- ▶ Combine manual and automatic metadata generation



### 3 CONTENTUS – technologies for the multimedia library of the future



**Germany's 30,000 libraries, museums and archives house an incredible wealth of knowledge in the form of millions of books, pictures, tapes and films. Researchers involved in CONTENTUS are examining how this cultural heritage can be made available to as many people as possible, and how we can preserve this information for future generations.**

One of the biggest challenges of our knowledge-driven society is how to give users easy, cross-media access to knowledge and cultural assets. In the framework of CONTENTUS, concepts and technologies are being developed for an infrastructure that enables cultural institutions and information providers come closer to achieving this goal. These new technologies make it possible to automatically process and semantically link large volumes of data, be that in the form of texts, images or video and audio recordings. The work of CONTENTUS is closely harmonized and coordinated with the "German Digital Library" initiative of the Federal Government. The comprehensive description of content using metadata plays a central role here. Metadata can be compared to

keywords or indices and are deployed to make search and retrieval easier. Linking the various digital knowledge records on the Internet, and integrating user-generated content, is another task facing the researchers. To this end, they are developing intelligent algorithms that automate all necessary processes and allow users enrich digital content with their own information. This results in interconnected next-generation virtual media libraries.

#### **Next-generation multimedia libraries**

Under the umbrella of CONTENTUS, researchers are developing concepts and technologies which lay the foundation to make next-generation multimedia libraries a reality. These multimedia libraries link the collections of libraries, media archives and broadcasting institutions to a new information structure that unites providers and users via the Internet. In doing so, new processes are created so that users can add their knowledge to the existing multimedia knowledge base. At the same time, the data are semantically linked to an innovative knowledge network. In this way, providers can make their multimedia

collections available and optimize data structure and quality.

One example clearly underlines the benefits this next-generation multimedia library has to offer: A search for “Lucullus” returns numerous multimedia hits including documents, images, films or audio data. Thanks to the use of semantic technologies, topic-related content is also listed. As a result, scores or recordings of the “Trial of Lucullus” would be displayed in addition to other works of author Bertolt Brecht and composer Paul Dessau, such as “The Good Person of Szechwan”. If requested, pieces from the most famous interpreters of Brecht’s music are displayed, as are links to commercial music providers that sell Brecht’s songs online. While people using classic archives would usually have to run intensive searches before finding individual works or parts of collections, users of the planned multimedia library benefit from rapid search processes and useful results spanning all media formats.

#### Advantages for users and providers

With CONTENTUS, users can search and navigate more easily through digitized cultural heritage in the future. They can also easily share their results and findings with other users and providers, and add links to enrich content. As part of CONTENTUS, a special platform for scientific work with digital content is also being created which enables providers, such as cultural institutes, make their multimedia collections accessible to a wider audience. Technologies are also being created for the qualitative processing and semantic connection of multimedia data, which means that even smaller libraries and archives can become part of the planned information structure.

#### Research and development under the flag of knowledge

Numerous technologies and new concepts that are crucial to developing an extensive knowledge network are being united for the first time in the CONTENTUS program. In a six-step process, analog data media are digitized and readied for next-generation multimedia libraries. High throughput methods help digitize large volumes of analog

content quickly and efficiently. After this step, the researchers develop technologies to gauge and optimize the quality of the imported material. Text, keyword, voice or language recognition processes analyze the content to then generate metadata that describes it. This type of indexing is crucial to be able to find digitized books, films or recordings in the vast World Wide Web. Step four uses these metadata to interlink content items as part of the semantic linkup process. As illustrated in the “Lucullus” search example, related items are suggested to users. If they so choose, they are then taken on a journey to discover and thoroughly understand their selected topic. In addition to classic full-text searches, users could also run searches like “which of Thomas Mann’s children were also writers and what did they write?” The fifth step involves users and experts who can add to the content. The final step completes the process: interlinked, multimedia access to information – a milestone in the development of a Web-based knowledge infrastructure.

#### Objectives of CONTENTUS

- ▶ Safeguard cultural heritage
- ▶ Develop concepts and technologies for next-generation multimedia libraries
- ▶ Integrate user-generated content
- ▶ Create new semantic and cross-media search options

## 4

## MEDICO – smart searching in medical databases



Imaging methods, such as ultrasound or computer tomography (CT), are critical to medical diagnosis and therapy plans as they make it possible to detect diseases at an early stage and initiate the right treatment. Vast amounts of heterogeneous information in the form of documents, images, laboratory data etc. are generated by these processes. Up to now however, there has been no instrument available that intelligently structures this information and makes it accessible. Researchers involved in the MEDICO application scenario aim to bridge this gap.

How can a computer learn to interpret images, recognize similarities and appropriately relate them to text-based content? These are the questions addressed in the THESEUS MEDICO application scenario. Using semantic technologies, the participating researchers are developing applications for simple and networked searches in medical databases. Ultimately, their aim is to support doctors and those working in the health sector by intelligently compiling all the relevant patient information from image and text-based results.

#### New possibilities for diagnosis and therapy

MEDICO not only recognizes anatomical structures like bones, tissues and organs. It also classifies data automatically, compiles reference images and treatment reports from multiple databases and thus identifies pathological abnormalities. The benefits of MEDICO can be clearly illustrated taking the example of lymphoma diagnosis and treatment. A lymphoma patient is being treated at a local hospital. The physician in charge of the treatment wants to know whether the chemotherapy has started to take effect. MEDICO automatically compares recent CT images to those taken before therapy commenced, and extracts relevant information from these images. The result: The chemotherapy seems to be working since the lymph nodes have not increased in size. However, the program finds new lesions on the spleen that need to be investigated. The attending physician does not have any similar cases at the local hospital that could provide more information. To clarify the damage to the spleen, the physician uses another feature of MEDICO that digitally transmits the images and test results of the patient to the data center of

a large university hospital. The program compares the data of the lymphoma patient from the local hospital to the numerous cases of lymphoma that have been treated at the university clinic. The oncologist in charge of the treatment can use this vastly broader information base to receive precise recommendations on further tests to be carried out and possible therapy. Furthermore, by combining the MEDICO system with the ORDO component of THESEUS, the physician can run a rapid search in the latest scientific literature on the subject of spleen lesions in lymphoma patients. Optimum treatment for the patient can thus be identified by drawing on up-to-the-minute research findings.

#### Support for physicians and patients

The intelligent interpretation and retrieval of diverse images and texts present the medical sector with a vast range of opportunities. Attending physicians have an efficient and precise tool at their disposal while patients benefit from the fact that their doctors can draw on a wide intelligence base when making diagnosis and therapy-related decisions. In addition, MEDICO will help hospitals save time and money as unnecessary examinations and tests, which are often exhausting for the patient, can be avoided.

The researchers involved in the MEDICO component framework are initially concentrating on semantic searches and the structure of information in medical image databases. However, MEDICO holds even greater potential than this. As soon as the ethical and data protection issues have been clarified, patients and the pharmaceutical industry alike can benefit from rapid and efficient access to relevant studies which can further improve diagnosis and therapy quality. Insurance agencies and official bodies could also exploit the technologies developed in the future, and conduct considerably more accurate demographic or statistical studies on the basis of anonymous data.

#### Giving data a useful structure

The technologies developed with MEDICO should enable computers to generate information automatically from the content of the images shown, merge this information with other medical data

and structure the results to identify recommendations for the physician in charge of treatment. To achieve this goal, numerous components still have to be developed, including pattern recognition methods, computer-aided recognition systems or clinical decision-making tools. However, the central task comprises modeling a suitable ontology, which involves developing a classification system that can categorize and link the wealth of information on the basis of medical terminology. The metadata needed are generated automatically by MEDICO and assigned to the specific image data. In MEDICO, the image content is structured on the basis of the human anatomy. Ontologies developed by medical experts such as “RadLex” or the highly granular “Foundational Model of Anatomy Ontology” are used to describe the image content.

With the development of ever-improving diagnosis and treatment methods, the amount of information doctors have surrounding their patients and their disease patterns grows. MEDICO interprets, links and compares medical image and text information quickly and easily, and thus plays a key role in advancing medical science.

#### Objectives of MEDICO

- ▶ Intelligently search and structure medical information such as images, test results and texts, and make information available for different medical situations
- ▶ Provide support in making diagnosis and therapy-related decisions
- ▶ Improve speed and efficiency in accessing medical knowledge in day-to-day clinical situations and research scenarios



## 5 ORDO – order in a digital world



**Human knowledge is available in so many formats, with information stored as text, graphics, video and audio recordings in different file types and on diverse data storage media. As part of the ORDO application scenario, researchers are developing a system that can better handle this heterogeneity.**

Modern information technology makes our lives easier but sometimes also less transparent. Consider the Smith family: they have just come back from vacation and want to document their trip. But a few minutes into the job, things are not as easy as they sound. The holiday snaps are saved on the digital camera, greetings from the beach are stored as text messages on the cell phone, and daughter Jessica kept a travel diary on her laptop. How can Dad combine all this information to create an entertaining presentation to show friends? How can he clearly arrange the data and decide what is important and what is not?

### **Organizing knowledge better**

ORDO aims to develop new services and software tools that help users organize and prioritize information and search results. Semantic processes are used and implemented for this purpose. This results in technologies that provide users with all the data whose content matches their search query, and allow users to distinguish between relevant and irrelevant information, identify connections and view background information. In our example, the new technologies would give Dad an immediate overview of what holiday information was stored on the specific data storage media. He could visualize the information easily, rate its relevance and probably decide not to include his daughter's travel diary in the presentation after all.

This overview of the information available is generated automatically since ORDO technologies “tidy up” on their own. Furthermore, solutions are being

developed to process and graphically illustrate large volumes of data. This results in a comprehensive and individual knowledge management system that enables users to always find the right information.

#### Practice run in research and development

The research results of ORDO are being specifically tested in pilot projects to see whether they meet the needs of research and development departments – such as those in the chemical and pharmaceutical industry – that have to handle large volumes of data on a regular basis. In addition to analyzing primary data, such as patents, these industries also have to assess extensive secondary information, including research reports and other scientific publications. Furthermore, e-mails, archives, internal Wikis and blogs also have to be screened. The ultimate goal of the research work conducted in ORDO is to develop methods and technologies that make all these stores of data more efficient to use and allows them to be linked to content from the Internet.

#### Extensive research options

One example from the world of science clearly illustrates the benefits of ORDO technologies: a pharmaceutical researcher is looking for information on the ingredient acetylsalicylic acid. Comprehensive search results should illustrate the agent's chemical formula, explain how the ingredient works and provide a list of pharmaceuticals containing the drug, including images of specific products. Up to now, researchers would have had to compile all this information themselves, retrieving data bit by bit from a chemical database, a pharmaceutical archive and even marketing documents of the drug manufacturer.

With ORDO, the researcher would be able to direct his/her query at a central platform that then coordinates the search. The computer program used can automatically find the chemical formula for acetylsalicylic acid in a database and then apply the formula to locate a graphic of the substance's molecular structure. All this information is returned to the scientist in a transparent format. Researchers can arrange the search results according to their particular needs and individually choose what information is relevant and where they want to learn more.

#### Information management using semantic technologies

With ORDO, computers will soon be able to identify concepts and relationships in text documents, extract facts and generate summaries and synopses automatically. Compared to existing recognition systems, these features constitute a huge step towards more efficient information management. Furthermore, these technologies can also be used in the future for analyzing and organizing multimedia content.

The ambitious ORDO project is based on the enhancement and sophistication of semantic technologies that enable a computer to automatically recognize data content and meaning, and put these data items into a specific context. This is particularly important when organizing and analyzing unstructured data. A further focus of ORDO is the development of efficient processes to create classification systems used to model, manage and organize data. Combined with innovative linguistic and statistical methods, these ontologies can be implemented more effectively still, and can even be generated automatically in some instances – a decisive step in making the Internet a more organized place.

#### Objectives of ORDO

- ▶ Semantically classifying and prioritizing information
- ▶ Recognizing concepts and relationships in texts, extracting facts and generating summaries automatically
- ▶ Visualizing the connection between items of information



## 6 PROCESSUS – improving how corporate knowledge is leveraged



**In light of the global competition for innovative products and services, corporations have to address the challenge of optimizing how they put their knowledge to use. As part of the PROCESSUS application scenario, a solution is being developed that ensures the right employee gets the right information at the right time. This paves the way to reducing costs and optimizing business processes.**

Knowledge is a fundamental cornerstone for the competitiveness of German industry. However, a lot of critical information is hidden in internal documents, such as contracts, analyses, plans, schedules, business reports and e-mails. Consequently, employees often only have limited access to the valuable knowledge these documents store.

While conventional data compilation systems can analyze documents, they cannot put the documents into a general context. Currently, the data sources are

only assessed by their content and then saved in different databases for the specific target groups. The systems do not yet record the general business processes where this information can be exploited. This is where PROCESSUS comes into play as it specifically provides every employee with the knowledge that is particularly relevant to his or her individual case.

The challenges facing PROCESSUS are many and varied. Conventional data retrieval systems must be adapted to include functions for analyzing dynamic workflows. Diverse databases have to be interconnected so it is no longer important where information is physically stored. Different specialist terms have to be analyzed and harmonized by a computer program implementing semantic technologies. Work process analysis also plays a pivotal role in delivering perfectly tailored information to users. As soon as PROCESSUS understands the context in which the information is needed, the results are adapted to this context.

### Field test in mechanical engineering and information technology

The concepts and solutions devised in PROCESSUS are being tested in the fields of mechanical engineering and information technology. In mechanical engineering, the focus is on applications in drive and automation engineering. The mechanical engineering sector, in particular, has highly complex business processes and extremely diverse user groups. A typical process chain in this industry comprises preliminary design, construction, production, sales and customer service. The same product is processed in very different contexts in these individual steps, with specialists in the construction team primarily handling technical information while their colleagues in sales deal more with market and competitive analyses. As a result, diverse data have to be made available in the context of the specific process step.

In such situations, PROCESSUS applies a completely new approach towards accessing information: users describe their problem and, while processing the query, the computer includes solutions already successfully implemented in the same context. In this way, the solution presents the user with new insight and new possibilities and opportunities, such as ideas on how to use existing products in nontraditional markets. For example, on the basis of a PROCESSUS-aided search, a sales agent at a company that manufactures paint production systems realizes how easily a machine component, such as an arm or a labeling unit, can be used in the production of beverage bottling systems.

This intelligence can also just as easily be applied to the area of information technology when linking diverse computer applications in a service oriented architecture. In the PROCESSUS program, an infrastructure for managing content is being developed for the software industry to classify software components by their ability to solve problems and not by their technological features. From this “toolkit”, suitable parts can be easily combined for an application thanks to the content described by PROCESSUS.

These two examples demonstrate how the technologies developed as part of the PROCESSUS application scenario make it possible to link industry and

process knowledge more quickly, and develop new products, solutions and business models. This gives companies a competitive advantage as they become quicker and more agile now that staff can always tap knowledge when and where they need it, and companies can save money on research and development. When it comes to quality, businesses also have an edge over the competition since the knowledge made available is of a particularly high caliber. Consequently, PROCESSUS helps companies leverage their knowledge optimally to better face down competition on the global market.

### Synergy with TEXO

While the results of PROCESSUS are useful on their own, they hold enormous potential when combined with the technologies developed in the TEXO application scenario in particular. Given their modular design, the PROCESSUS components can be deployed as individual services on a TEXO platform. In this way, the information provided by PROCESSUS can be called up individually and in context using TEXO, and knowledge available in businesses can be used more effectively.

### Objectives of PROCESSUS

- ▶ Extend conventional content management solutions to include process orientation
- ▶ Make it easier to access relevant corporate knowledge
- ▶ Reduce costs and increase efficiency through the use of semantic technologies



## 7

## TEXO – infrastructure for Web-based services



While the sale of music and video content on the Internet is commonplace nowadays, the same cannot be said for trade in services on the World Wide Web. For this reason, the TEXO application scenario focuses on developing an infrastructure that makes it easier to combine and avail of services on the Internet – an important step on the path to the Internet of Services.

Searching for suitable services on the Internet can be a complex and tedious task for consumers and businesses alike. When processes get complex, the fact that different partners are responsible for different subservices, and such services cannot be combined practically, complicates matters even further. For example, any foreign company looking to open an office in Germany will soon realize what a laborious task setting up a business is. Choosing the location, adding the company to the register of business names, analyzing the market and hiring suitably qualified staff are tasks that require a lot of time and resources. For many small and medium-sized enterprises, in particular, it would be an enormous relief if all these steps could be handled by a central point of contact. But which providers can offer all the necessary services? And how can they be found?

### Framework for an innovative services industry

TEXO is developing an integrated online platform that can solve these problems. This lays the foundation for a new services industry on the Internet where services can be bought and sold like commodities, and combined to form new product offerings. The platform enables trade in technical services and conventional products such as consulting services, while accommodating both the interests of the service providers and the needs of the customer.

### Wide range of benefits for providers

The possibility of offering services over the Internet opens up huge opportunities for midmarket businesses in particular. They can present their services worldwide and win new customers. Companies can also expand their portfolio by combining their services with those of other providers, and thus access new markets with innovative service packages. TEXO provides the technological framework to ensure that new services can be developed at a faster pace and offerings from partner companies can be integrated more easily. As a result, small and medium-size enterprises especially can concentrate on their core

competencies and save time and money, since tasks that do not form part of their main business profile can be simply outsourced to external service providers.

### **New opportunities for customers**

Both consumers and businesses can benefit as customers from the new service platform. The following three examples illustrate the advantages associated with TEXO.

#### **Example 1: The “eco calculator” for businesses**

To be able to export its products, a business needs a certificate that documents the recyclability of the goods. Only selected providers can actually issue such a certificate. The company runs a search on the TEXO platform and quickly finds the information it needs as various “eco calculators” are listed as services for product certification. The selected “eco calculator” is automatically integrated into the business working environment and can be used here to calculate the product recycling rate as a prerequisite for certification. Thanks to TEXO, businesses can thus find the services to suit their needs without conducting complex and time-consuming searches, and can avail of the services on a user-friendly platform.

#### **Example 2: Official paperwork handled at a central point**

When a child is born, the parents are faced with many tasks, including visits to government offices, compulsory medical checkups, e-mails to friends and so much more. With the TEXO platform, the requisite services from businesses and government offices can be grouped together and offered on one single Web site. This means that the parents can perform all the necessary tasks at a central location using an intuitive user interface. For example, they only have to enter their name and address once and automatically receive information on pediatricians in their neighborhood, as well as all the essential forms from the competent authorities.

#### **Example 3: Simplified business startup**

The European Union’s Services Directive aims to support EU citizens in setting up a business by making

it possible to process all the associated official paperwork through one point of contact. The TEXO platform allows users to flexibly select and combine the services of various government offices to create tailor-made packages for the individual entrepreneurs and track the status of the application. With TEXO, an individual and dynamic set of services is developed that is perfectly suited to the specific needs of entrepreneurs wishing to open an office in Germany.

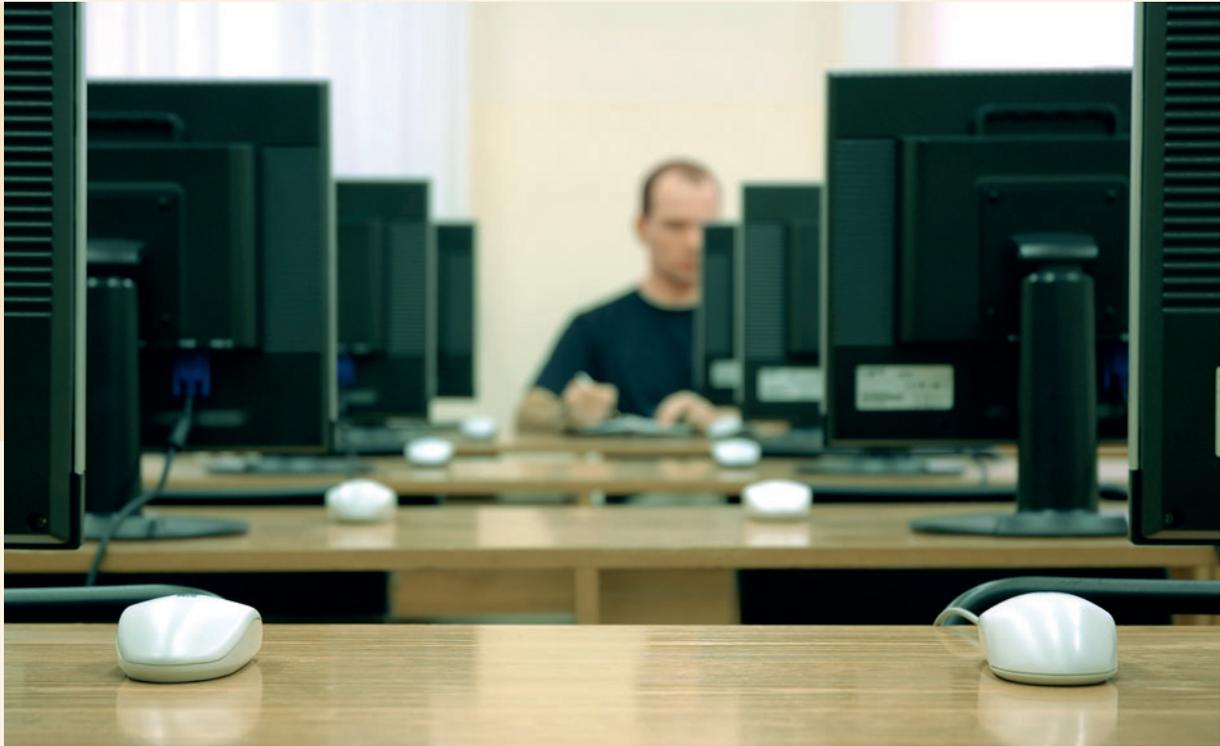
#### **The basis: SOA**

Service oriented architecture (SOA) is the technological backbone to TEXO. SOA is a technology used in IT that is capable of grouping different computer applications to form one single application. The researchers working on TEXO are using SOA as their basis to develop an infrastructure that applies semantic technologies to describe the meaning of information and services in such a way that is comprehensible for a computer. This sets the stage for a new services industry on the Internet.

#### **Objectives of TEXO**

- ▶ Create a service platform on the Internet
- ▶ It should be possible to trade services like commodities and combine them to create new offerings
- ▶ Provide the basis for an open and secure environment for trading services on the Internet
- ▶ New business models, particularly for SMEs

## 8 THESEUS – basic technologies



**The Internet holds more knowledge than all the archives and databases in the history of humanity. The THESEUS program concentrates on developing the necessary basic technologies and standards to improve and ease access to this knowledge base in the future.**

The primary focus of THESEUS is the development of semantic technologies that are capable of registering and evaluating the content and meaning of information. In this way, computer programs can automatically analyze the content of information like texts, graphics or audio and video recordings, link the information and draw logical conclusions – a critical development on the path to the Internet of Services.

Innovative basic technologies and technical standards have to be developed to make this evolution of the Internet possible. The basic technologies being explored in THESEUS are developed and tested in six application scenarios by industry partners in the THESEUS consortium. The purpose of the test is also to discover how these technologies can be leveraged

as quickly as possible for new tools, services and business models on the Internet.

### **An outline of the basic technologies:**

#### **1. Automatic generation of metadata**

The THESEUS research partners are developing new methods to register media content. These methods are based on metadata, which refers to information on an object that is recorded in a database and acts as a pointer towards this object when a search is run. An example would be the name of the author of a book, or the period in which a film is set. The new processes make it possible to create metadata for various media content, including texts, photos, and audio and video files. One of the subjects of the development project focuses on finding ways to automatically generate metadata. Another part of the project concentrates on using semantic technologies to record the context of specific media content. The aim is to retrieve content referring to the same topic from different sources, compile this information and make as much data available to the user as possible.

## 2. Rapid processing of multimedia documents

To reduce the wait time when searching through complex multimedia databases, researchers in the THESEUS program are designing high-efficiency algorithms for metadata. The algorithms make it possible to rapidly scan the contents of images and videos even in databases containing several hundred thousand objects. Another priority area is the compression of data, and particularly the compression of graphic files. Given that the quality and resolution of digital images will continue to improve, new encoding processes are needed to be able to compress and save the data. To improve image and graphic searches, image recognition systems are also being developed that will enable computers to automatically register the object illustrated in a photo or video.

## 3. Innovative ontology management

The THESEUS research program is developing new methods and tools for managing ontologies so information can be analyzed and linked efficiently. The term ontology refers to a classification system that describes the contents of data in greater detail and provides information on the data's context. The developers aim to create highly informative ontologies right down to automated logical reasoning which a program can perform itself. As with metadata, different knowledge stores can thus be linked to enable faster search processes that return better, more effective results.

## 4. Machine learning

A further THESEUS project focuses on the development of scalable technologies capable of structuring metadata and modeling ontologies. These technologies allow applications to use a computer's resources efficiently to perform as many computing steps as possible in a minimum amount of time. This uses innovative processes for determining relationships in texts, graphics and audio and video files that are based on machine learning methods. In machine learning, a computer not only processes information but also recognizes relationships and patterns between the individual items of information. The computer can then automatically apply this knowledge to other data.

## 5. Situation-sensitive dialog processing

Before a computer can act on behalf of a user, it must first understand what the user wants. To enable this dialog between humans and machines, researchers in the THESEUS program are developing new functions that can be flexibly deployed in different applications. An individual mode of communication can be selected depending on how the user wants to communicate with the computer, and the reason for this communication. Innovative algorithms enable multimodal user interfaces which can be controlled through gestures and human speech. Such interfaces make it possible for users to formulate their queries and requests intuitively and refine the request through verbal dialog with the system. A special component inside the computer – the semantic mediator – acts as the interface between the multimodal user interface and the various metadata sources. It converts a verbal request into a semantically correct data record which the system requires to run a search.

## 6. Innovative user interfaces

New graphic user interfaces are also being developed in the THESEUS program to make it easier to identify and understand the relationships between different data, metadata and documents. In this way, the results of a query can be illustrated as a knowledge network, which indicates how the search results are related to the actual search string and to one other. This, in turn, provides users with a clear overview of a specific topic and helps them find the information they need more quickly. Such an approach not only benefits private users but also helps businesses, since they can apply defined standards to easily integrate such user interfaces and tailor them to meet their specific requirements, while saving on development costs in the process.

## 7. New ways to identify content

The research partners are also devising new solutions for clearly and unambiguously identifying content even if slight changes have been made to the data. For example, film-makers could track different versions of a film in the production process. The solutions can also be applied to protect copyright and marketing rights to intellectual property, thereby

ensuring that the rights of the holder of the copyright are not violated despite the increased availability of multimedia content on account of the Internet.

#### **8. Evaluation of basic technologies**

The quality of the basic technologies developed under the umbrella of THESEUS is appraised and analyzed by experts. For example, new technologies for speech and image recognition, or for the automatic assignment of metadata, are examined in terms of their reliability, functionality and suitability. The results of this assessment are then integrated into the research and development process to help optimize the end result. The appraisals and expert reports are also used to develop technical standards.

## 9 THESEUS – Joint Research



Many research efforts are being conducted around the world with the aim of developing and enhancing the Internet. Joint Research is a THESEUS module that creates a bridge to these international project partners. It identifies interesting projects, initiates contact with participating researchers and establishes a global network that benefits all the scientists involved.

THESEUS Joint Research is commissioned by the BMWi and supports THESEUS research projects in terms of international networking. The Fraunhofer Institute for Software and System Technology (ISST) and consulting agency ESG Consulting are responsible for developing and managing Joint Research activities.

Joint Research compares the technologies, applications and business models developed in THESEUS to similar European and international initiatives. It performs comparative analyses to identify similarities and differences between the THESEUS projects and other projects centering on the areas of research that make up the THESEUS program. The BMWi and the THESEUS consortium examine the results to draw strategic conclusions for positioning and developing the THESEUS program. In doing so, THESEUS Joint Research pays particular attention to global considerations and plans to develop the Internet of the future.

Joint Research closely tracks the solutions applied in these projects that are aimed at developing new

offerings for information and services to revolutionize the Internet and make it a Web-based knowledge infrastructure.

The ]LEVERAGE.KNOWLEDGE[ motto of THESEUS Joint Research reflects the expectations associated with Internet-based knowledge infrastructures and the technical solutions needed.

Joint Research comprises user services and functions that are to be made available on the Internet of Services. The significance of semantic technologies for the successful introduction and utilization of the Internet of Services is also appraised. Where possible, the engineering concepts and processes developed and implemented in designing the Internet of the future are presented and discussed by the Joint Research teams.

By linking THESEUS to economic and social players in Germany, the project aims to initiate knowledge transfer to various fields of application. For this purpose, high-profile publicity events are held, publications are issued and the results of the Joint Research are made available on the Internet.

## 10 THESEUS Talents 2008 – ideas competition to foster young research talent

**Bright ideas, creative minds.** In its “THESEUS Talents 2008” ideas competition, the THESEUS consortium gave young researchers and programmers the opportunity to submit their own ideas and actively shape the research work conducted in the THESEUS program.

Junior scientists, students, pupils and young developers were encouraged to submit ideas for using semantic technologies under the THESEUS framework, and thus play an active role in developing the Internet of Services. Applicants could choose from almost 30 subject areas related to the basic technologies and application scenarios of THESEUS.

From image classification processes that automatically categorize pictures and graphics on the basis of defined features, to suggestions for new services on the Internet – the possible ideas and approaches were as varied as the THESEUS research program itself. The jury had to choose from over 150 entries and rated the projects submitted on the basis of their quality, originality, innovativeness and feasibility.

First place went to Sonja Kraus, a computational linguistics student from Grevenbroich, who impressed the jury with her “semantic strategies” project. This focused on enriching content using metadata – known as tagging – which, in turn, makes it easier to search and link content. Sonja developed an easy-to-use tool that provides users with automatically generated suggestions to assign tags to content on the Internet. Assigning metadata is easier and quicker as a result – an important step on the path to the Internet of the future.

### Prize money and sponsorship

A total of 50,000 euros in prize money was awarded. In addition to the cash prizes, the 14 winners were also given the opportunity to implement their ideas while completing internships at the companies and research institutes of the THESEUS consortium. In this way, the young scientists get hands-on experience working in a network of reputable partners, and are mentored and supported by experts in the field.



The THESEUS Talents competition that was held by the Federal Ministry of Economics and Technology in association with the THESEUS consortium thus played a pivotal role in specifically fostering up-and-coming IT talent, and in opening the research program to new ideas.

### THESEUS Talents – facts and figures

- ▶ Ideas competition for young scientists in Germany
- ▶ Tasks based on specific issues from over 30 topics concerning basic technologies and the fields of application
- ▶ A total of 50,000 euros in prize money
- ▶ Prize winners sponsored by businesses in the THESEUS consortium

# 11

## THESEUS SME 2009 – call for SME-operated projects

Given their speed and flexibility in developing and tapping new markets, small and medium-sized enterprises (SMEs) are the forerunners in many areas of technology. However, these very businesses often do not have enough time or resources to invest in research and development. Against this backdrop, they can benefit all the more from the partnership and collaboration opportunities offered by the “THESEUS SME 2009” call for projects.

The “THESEUS SME 2009” call for SME-operated projects gives small and medium-sized enterprises in Germany the opportunity to trial the technologies developed in the THESEUS program and translate them into new products and processes. High application orientation is the focus of the call for projects, which is run by the Federal Ministry of Economics and Technology (BMWi) in association with the THESEUS consortium. Under the framework of the call, the BMWi is making some 10 million euros in funding available for a period of up to three years.

### Wanted: innovative applications and processes

Interim results from the first program phase, which are based on the development and trial of THESEUS basic technologies in application scenarios, lay the foundations for integrating small and medium-sized enterprises. Drawing from the research results available, the businesses have the opportunity to develop and test new technologies, products and processes. The call for projects promotes application-oriented projects such as the development of new services and business models. SMEs with up to one thousand employees can apply and submit their projects either individually or in collaboration with other partners. A forum is available on the Internet for companies interested in sourcing suitable partners.

The call for projects is a two-phase process. In the first phase, interested companies enter an outline of their project by December 1, 2008. An independent jury analyzes these entries in terms of their originality, feasibility, commercialization opportunities and sustainability. This appraisal also takes the skills and abilities of the applicant into account, as well as the relevance of the entry to the objectives of the THESEUS program. In the follow-up consultation phase,

the applicants selected by the jury can firm up and elaborate their project outline in the course of workshops with the THESEUS consortium, and use this as the basis for applying for funding.

### More growth and employment

“THESEUS SME 2009” is central to promoting groundbreaking research in small and medium-sized enterprises. The early transfer of scientific results from THESEUS to SMEs improves the businesses’ competitiveness and drives the multiplier effect, as the results can also be used more quickly by other enterprises. This provides important stimulus for more innovation, growth and employment in Germany.

For more information on the call for projects, visit: <http://theseus-programm.de/en-theseus-mittelstand-2009.html>

### The objectives of “THESEUS SME”

- ▶ Open the program to witness greater participation from small and medium-sized enterprises
- ▶ Early transfer of scientific findings to the midmarket
- ▶ Greater participation of SMEs in the results of THESEUS
- ▶ Extend the THESEUS program by developing and testing innovative technologies for new products and processes
- ▶ Boost the impact and sustainability of the THESEUS program

## 12 Partners in the THESEUS program

- ▶ German National Library, Frankfurt am Main
- ▶ Deutsche Thomson oHG, Hanover
- ▶ German Research Center for Artificial Intelligence (DFKI) GmbH, Kaiserslautern
- ▶ empolis GmbH, Gütersloh
- ▶ Festo AG & Co. KG, Esslingen
- ▶ Research Center for Information Technologies (FZI), Karlsruhe
- ▶ Fraunhofer-Gesellschaft (9 institutes)
  
- ▶ Institut für Rundfunktechnik GmbH (IRT), Munich
- ▶ intelligent views GmbH, Darmstadt
- ▶ Ludwig-Maximilians-University (LMU), Munich
- ▶ Lycos Europe GmbH, Gütersloh
- ▶ mufin GmbH, Berlin
- ▶ moresophy GmbH, Munich
- ▶ ontoprise GmbH, Karlsruhe
- ▶ SAP AG, Walldorf
- ▶ Siemens AG, Munich
  
- ▶ Technical University of Munich, Garching
- ▶ Technical University of Darmstadt, Darmstadt
- ▶ Technical University of Dresden, Dresden
  
- ▶ Erlangen University Hospital – Radiological Institute, Erlangen
- ▶ University of Karlsruhe, Karlsruhe
  
- ▶ VDMA Gesellschaft für Forschung und Innovation (VFI) (German Engineering Federation), Frankfurt am Main





This publication is available free of charge as part of the public relations work of the Federal Ministry of Economics and Technology, and may not be sold. It may not be used by political parties or campaigners or electoral assistants during an election for the purposes of campaigning. In particular, it is forbidden to distribute this publication at campaign events or at information stands run by political parties or to insert, overprint, or affix partisan information or advertising. It is also forbidden to pass it on to third parties for the purposes of electoral campaigning. Irrespective of when, in what way, and in what quantity this publication reached the recipient, it may not be used even when an election is not approaching in a way that might be understood as suggesting a bias in the federal government in favour of individual political groupings.