



---

D9.1 Project Description – vM36

Kalina Bontcheva (University of Sheffield)

EU-IST Specific targeted research project (STREP) IST-2004-026460 TAO  
Deliverable D9.1 (WP 9)

---

WP9 Project Management v M36  
Nature: Report  
Dissemination: PU  
Contractual date of delivery: 28 February 2009  
Actual date of delivery: 15 April 2009

## **TAO Consortium**

This document is part of a research project partially funded by the IST Programme of the Commission of the European Communities as project number IST-2004-026460.

### **University of Sheffield**

Department of Computer Science  
Regent Court, 211 Portobello St.  
Sheffield S1 4DP  
UK  
Tel: +44 114 222 1930  
Fax: +44 114 222 1810  
Contact person: Kalina Bontcheva  
E-mail: K.Bontcheva@dcs.shef.ac.uk

### **Mondeca**

3, cité Nollez  
75018 Paris  
France  
Tel: +33 (0) 1 44 92 35 03  
Fax: +33 (0) 1 44 92 02 59  
Contact person: Jean Delahousse  
E-mail: jean.delahousse@mondeca.com

### **University of Southampton**

Southampton SO17 1BJ  
UK  
Tel: +44 23 8059 8343  
Fax: +44 23 8059 2865  
Contact person: Terry Payne  
E-mail: trp@ecs.soton.ac.uk

### **Sirma Group Corp., Ontotext Lab**

Office Express IT Centre, 5th Floor  
135 Tsarigradsko Shosse Blvd.  
Sofia 1784  
Bulgaria  
Tel: +359 2 9768 303  
Fax: +359 2 9768 311  
Contact person: Atanas Kiryakov  
E-mail: naso@sirma.bg

### **Atos Origin Sociedad Anonima Espanola**

Dept Stream-Public Sector  
Atos Origin Spain, C/Albarracin, 25,  
28037 Madrid  
Spain  
Tel: +34 91 214 9321  
Fax: +34 91 754 3252  
Contact person: Nuria De Lama  
E-mail: nuria.delama@atosorigin.com

### **Dassault Aviation SA**

DGT/DPR  
78, quai Marcel Dassault  
92552 Saint-Cloud  
Cedex 300  
France  
Tel: +33 1 47 11 53 00  
Fax: +33 1 47 11 53 65  
Contact person: Farid Cerbah  
E-mail: Farid.Cerbah@dassault-  
aviation.com

### **Jozef Stefan Institute**

Department of Knowledge  
Technologies Jamova 39  
1000 Ljubljana  
Slovenia  
Tel: +386 1 477 3778  
Fax: +386 1 477 3131  
Contact person: Marko Grobelnik  
E-mail: Marko.Grobelnik@ijs.si



## **Publishable Executive Summary**

### **Transitioning Applications to Ontologies – TAO**

<http://www.tao-project.eu>

Semantic Web technology, ontologies, and semantically described web services have now demonstrated their maturity and usefulness in a number of applications. However, their wide-scale adoption is hampered by the relatively high adoption costs, mainly through the need of employing specialist knowledge engineers, the manual ontology creation effort for the domain experts, and the need to deal with diverse, heterogeneous legacy data and software.

The migration problem is particularly difficult in legacy applications, which tend to be:

- Built with languages and data models that are now out-dated
- Badly structured and hard to maintain
- Badly documented and understood
- Difficult to integrate with each other and with new systems
- Need for migration towards Web 2.0 applications and services

The goal of TAO has been to make transitioning existing 'legacy' applications to ontologies fast and effective, thus allowing companies to:

- Build a reusable transitioning process;
- Minimise consulting time during migration and integration;
- Minimize costs;
- Reduce integration overheads and limit risk.

We created an open source infrastructure to aid transitioning of legacy applications to ontologies, through automatic ontology bootstrapping, semantic content augmentation, and generation of semantic web service descriptions. The work is grounded in the TAO transitioning methodology and the tools are integrated into the TAO Suite. In this way, TAO enables a much larger group of companies to exploit semantics without having to re-implement their applications.

The results are validated in two high-profile case studies: a comprehensive open source platform (with thousands of users) and a data-intensive business process application at Dassault Aviation (managing a multi-million business).

## **Impact and Exploitation**

This project offers a low-cost migration path for legacy applications to knowledge technologies and is accessible to both SMEs (which are cost sensitive) and large enterprises (with huge investments in complex and critical IS). The results have been



## D9.1 / Project Description

validated in two high-profile case studies: a comprehensive open source platform (with thousands of users) and a data-intensive business process application (managing a multi-million business).

Two of the TAO project partners (University of Sheffield and Ontotext) obtained over 750,000 euros in follow-on commercial funding from the Austrian company Matrixware to apply two of the TAO tools (content augmentation and the large-scale knowledge store) to the problem of Large-Scale Semantic Annotation of Patents. The goal is to exploit this TAO technology and annotate terrabytes of data in several days of supercomputer time. The project starts in April 2009.

The TAO ontology learning tools are now used by the company behind <http://videlectures.net> for the automatic classification of video materials posted to their web site.

Last but not least, as part of their case study, Dassault Aviation carried out standardisation work aimed at that industry, which arose from TAO's contribution to aviation domain ontologies and transitioning of legacy databases

### **TAO's main innovation**

There are three main technological innovations in the project:

1. Bootstrapping via semi-automatic acquisition of domain ontologies. The main research objective of TAO is to identify the major characteristics of the data sources required for learning domain ontologies suitable for annotation of legacy content and services; choose the right extraction methods; and support human corrections of automatically learned domain ontologies as part of the application transitioning process.
2. Augmentation and integration of legacy content (databases and documents used by applications) relative to the domain ontologies to enable ontology-based information access. A new heterogeneous knowledge store was developed to support scalability and heterogeneity. Efficient support for a combination of structured/semantic queries and keyword search queries is now in place, as are a number of content augmentation tools which enrich legacy data with semantics.
3. Transitioning Methodology and Infrastructure. TAO provides an innovative infrastructure for transitioning legacy applications to semantic- and service-based ones via semi-automatic bootstrapping.

### **TAO's results**

The main implementation outcome is the TAO Suite<sup>1</sup>. It is a user-friendly infrastructure, implementing the TAO methodology for the transitioning of existing applications. It is

---

<sup>1</sup> The online version of TAO Suite is available at: <http://www.tao-project.eu/researchanddevelopment/demosanddownloads/tao-suite.html>



## D9.1 / Project Description

based around TAO's innovative, open-source tools, which aid the transitioning of legacy applications towards ontologies. We cover a number of key steps:

- learning the domain ontology from legacy content;
- transitioning of relational databases to ontologies;
- automatic augmentation and semantic annotation of legacy content;
- storage, reasoning and semantic queries of the semi-automatically derived knowledge in a large-scale, heterogenous knowledge store.

### **Results: The Industrial Perspective**

The TAO Suite and its component technologies were presented to potential clients from diverse domains and needs during an industrial workshop organised at ATOS Origin, in Paris on 27th of January 2009. The clients understood very well that the proposed outcomes of the project did not constitute a shrink-wrapped, ready-to-use solution to the transitioning problem. They were pleased to discover and learn about TAO's methodological principles and technical products that could be re-factored to cover their specific needs. We received excellent feedback from this event, one comment being that they were waiting for such enabling technology in their business cases and another being that it was the first time they were seeing complementary solutions (ontology learning, content augmentation, knowledge storage and queries, WSDL annotations, SOA architecture etc.) harnessed together to facilitate the overall process of transitioning.

In other words, from this external industrial perspective, the project has been successful in developing and integrating the necessary enabling technologies for transitioning legacy applications to ontologies, without making a too rigid stance on what software architectures or semantic service formalisms must be adopted.

### **Dissemination**

With respect to dissemination, activities oriented non-scientific, industrial targets, but alongside this we have continued our scientific publications. TAO also presented its first tutorial at the European Semantic Web Conference (ESWC)<sup>3</sup> and another informal one, focused around the TAO Suite was delivered to the ServiceFinder project. TAO also organized an industry-oriented workshop in January 2009 which attracted a strong interest and produced very positive feedback on the technological achievements of the project.

Scientific dissemination has also been strong. In 2008 alone, the project published 11 papers, gave 1 tutorial at a major international conference, published numerous online video lectures, had collaborative meetings with other relevant projects, issued several press releases and had presence at all important scientific conferences and workshops in the field (e.g., ISWC, ESWC, WWW). TAO also targeted dissemination at industry-oriented events such as the European Semantic Technology Conference ESTC'08 and SemTech'08. A TAO [Wikipedia article](#) was also created to increase outreach.



## D9.1 / Project Description

### **Further details**

See the TAO website at <http://www.tao-project.eu>.

Also a showcase of the project is available at <http://www.tao-project.eu/index/TAO-Showcase/index.html>.

### **Administrative details**

TAO (FP6-026460) is a specific targeted research project of the European Union's 6th Framework Programme - call 4.

TAO started on 1 March 2006 and finished on 28 February 2009.

The overall budget is around 4.4 million euro.

7 partners from 5 European countries are involved in the project.

### **List of participants**

- University of Sheffield, UK
- University of Southampton, UK
- Atos Origin S.A.E, Spain
- Jozef Stefan Institute, Slovenia
- Mondeca, France
- Ontotext Lab, Sirma Group Corp., Bulgaria
- Dassault Aviation, France

### **Contact person**

#### **Project coordinator**

Dr. Kalina Bontcheva

Department of Computer Science

University of Sheffield

Regent Court

211 Portobello

Sheffield S1 4DP, UK

Phone: +44 114 222 1930

Fax: +44 114 222 1810

Email: [K.Bontcheva@dcs.shef.ac.uk](mailto:K.Bontcheva@dcs.shef.ac.uk)