Abstract: Formal modelling is a challenging and expensive task, especially for people not familiar with design techniques and formal languages. In this paper we present the modelling experience within the APOSDLE EU-project, and we describe the methodology we have developed to support an integrated modelling process of the ontologies and workflows inside APOSDLE. Our approach is based on two main pillars: (i) the usage of Semantic MediaWiki (SMW) for the collaborative development of informal, but structured, natural language descriptions of the models to be created, and (ii) a tight, and as much as possible automatic, integration of the SMW with tools for formal modelling, in order to re-use the informal descriptions for formal models creation.

Key Words: modelling methodology, semantic web technology, APOSDLE project

Category: I.2.4

1 Introduction and Motivations

The spread of Semantic Web technology, and Service Oriented Computing had the effect that more and more tools nowadays rely on the availability of formal models of specific domains (for instance in the form of an OWL ontology), of business processes (for instance in the form of a BPMN, or YAWL workflow) and learning goals to achieve particular competencies. Formal modelling is a challenging and expensive task, especially for people not familiar with formal design languages. As a consequence, organisations
interested in using a system are first asked to develop the formal models of their domains without having the expertise to do so. In this paper we present the experience of dealing with this problem within the EU project APOSDE\(^1\), and the methodology we have developed to support an integrated modelling process of ontologies and workflows inside this project. APOSDE aims at developing a software platform to support the process of learning@work, that is learning within the context of the immediate work of a user and within her current work environment. To deliver a user with context-sensitive learning material, tailored to her specific needs, the APOSDE system needs to know not only the profile of the user, but also about the context in which the user is acting, spanning from the tasks a user can perform, the learning goals required to perform the tasks, and a description of the domain of affairs (application domain) of the organisation. Most of this knowledge is contained in the APOSDE Knowledge Base, that we illustrate in figure 1.

![Figure 1: The APOSDE Knowledge Base](image)

Of the four models contained in the Knowledge Base, the APOSDE Categories (used to classify tasks, learning goals and learning material) are APOSDE-inherent structures, while the task model, domain model, and learning goal model are entirely organisation/domain dependent and need to be provided every time the system is deployed for a new domain. If we consider the typical application environment of our system, we can assume that:

1. most of the organisations won’t have these formal models already available;
2. most likely, they will not even be interested in defining such models, as their main interest is in setting up a tool enhancing their work environment’s productivity;
3. we need to model in an integrated way both specific static domains and learning goals, in the form of OWL ontologies, and business processes, in the form of YAWL workflows. This requires the organisation to become acquainted with more than one formal representation language.

\(^1\) [http://www.aposdle.org](http://www.aposdle.org)
4. we can not always rely on skilled knowledge engineers inside the organisation to take care of the whole modelling process, but we have to encourage, train and support knowledge engineers and domain experts from the organisation to be the main actors of the modelling phase.

If we look at the situation we have described above, we can easily realise that while certain aspects and problems of the modelling process are specific to APOSDLE (e.g. identification of adequate APOSDLE categories), the big problems we had to deal with encompass the specificity of the APOSDLE knowledge base, but have to do with the general problem of supporting the integrated modelling of an application domain, processes and learning goals, and to do it in a collaborative manner, without relying on a single knowledge engineer within an organisation taking care of the entire process, but on a modelling team composed of different knowledge engineers (possibly with different skills) and domain experts.

Despite the number of methodologies available (for a review of the state of the art, please see [Ghidini et al., 2007]), we have found that none was adequate to support, in a satisfactory manner, the collaborative and integrated modelling of knowledge, processes and learning goals described above. Therefore, starting from ideas and techniques available from existing methodologies, we have developed a specific methodology to support the creation of the APOSDLE knowledge base. While certain aspects of this methodology are tailored to the specific requirements of APOSDLE, the novelty of our overall approach concerns the development of tools for informal modelling, and their tight integration with tools for formal modelling, which can provide a concrete support to the integrated modelling process of knowledge, processes and learning goals in a general domain. Our approach is based on two main pillars: first, the usage of Semantic MediaWiki (SMW) [Krotzsch et al., 2005] for the collaborative development of informal, but structured, natural language descriptions of the domains to be modelled, and second, a tight, and as much as possible automatic, integration of the SMW with tools for ontology and workflow modelling. This in order to re-use the informal descriptions for automatic ontology and workflow creation.

In this paper we provide an overview of the methodology we have built, abstracting as much as we can from the specific settings of APOSDLE. The paper is structured as follows: first we present a brief overview of the entire methodology (Section 2), and then, due to lack of space, we concentrate on the part concerning the informal and formal modelling of a domain ontology (Section 3). We end with some concluding remarks (Section 4).

2 The Entire Methodology

In this section we present a brief overview of the methodology that we propose to support the integrated modelling process of ontologies and workflows. This methodology consists of five distinct phases which cover the entire process of models creation:
Phase 0. Scope & Boundaries. At the beginning, questionnaires and workshops are used to identify an appropriate domain to be modelled. Once the appropriate application domain has been chosen, its scope is determined with the help of use case scenarios. The results of this phase are documented in a MediaWiki².

Phase 1. Knowledge Acquisition. In this phase, knowledge about the domain to be modelled is acquired both (i) from experts of the domain, and (ii) from available digital resources relevant for the domain. Knowledge elicitation from experts is achieved using well-known and established techniques like interviews, card sorting and laddering, while knowledge extraction from digital resources is based on state of the art algorithms and tools for term extraction (e.g. see [Pammer et al., 2007]).

Phase 2. Informal Modelling. Starting from the knowledge acquired in Phase 1, an informal, but structured and rather complete, description of the different models which will constitute the knowledge base is created. The descriptions of the informal models are obtained by filling some predefined templates in a SMW. The use of the SMW allows to describe the elements of the different models in an informal manner using Natural Language. However, at the same time it allows to structure the descriptions in a way that they can be easily (and often automatically) translated in formal models, without forcing the members of the modelling team to become necessarily experts in the formal languages used to produce the formal models.

Phase 3. Formal Modelling. In this phase the informal descriptions of the models are transformed in formal models in a way which is as much automated as possible. The result of this phase is an OWL ontology describing the application domain, an OWL ontology describing the learning goals, and YAWL workflows modelling tasks.

Phase 4. Validation & Revision. Finally, the knowledge base created so far is evaluated and possibly revised. The methodology provides support to automatically check, via SPARQL queries, different properties both of the whole knowledge base, and of its single components. The results of these checks are evaluated and used to revise the knowledge base, if needed.

Due to the lack of space, in the next section we describe in details Phase 2 and Phase 3 of the integrated modelling methodology, focusing on the informal and formal creation of the domain model only (see [Ghidini et al., 2007] for a more detailed description of the whole methodology). We also report our experience and some lessons learnt applying the integrated modelling methodology in the context of APOSDE.

3 Modelling a Domain Ontology

Informal Modelling. Starting from the (flat or already partially structured) list of elements obtained and documented during the knowledge acquisition phase in a MediaWiki, the modelling team filters them, retaining only the relevant ones. The elements

² www.mediawiki.org
considered in APOSDLE are only concepts, but the approach could be extended to consider relations and individuals as well. In order to help the modelling team deciding whether a concept is relevant or not w.r.t. a particular domain, guidelines are provided. Examples questions used during the deployment of the methodology in APOSDLE are:

1. Is this domain concept useful for retrieval?
2. Are there resources dealing with this domain concept, or is it reasonable to expect resources dealing with this domain concept in the future?
3. Does this domain concept help to differentiate between resources?
4. Does this domain concept refer to a learning goal of a hypothetic APOSDLE user?
5. Does this concept help APOSDLE to support the acquisition of a learning goal?

Once the skeleton list of elements (concepts) is ready, we provide it as input to our informal modelling tool: the SMW. The idea is to use a pre-defined template to automatically create a page for each one of the concepts of the domain ontology and let the modelling team to fill the templates, thus providing the information needed to create the ontology. The reason to use a SMW is that it allows the modelling team (composed of domain experts and knowledge engineers) to provide the descriptions about the elements of the domain model in Natural Language. Differently from using a word processor, the Natural Language descriptions inserted in a SMW can be structured according to the pre-defined templates, and with the help of semantic constructs like attributes and relations. Therefore, the informal descriptions in Natural Language contain enough structure for (semi-)automatic translation in OWL ontologies, thus allowing to reuse the informal descriptions for automatic ontology creation.

*Template for the domain concepts.* Figure 2 shows a screenshot of the template we have used for describing concepts in APOSDLE. The template is tailored to the information we needed to obtain for APOSDLE. The bottom-level part of the template concerns the name and description of the concept, and take into account the aspects of

![Domain concept template](image)

**Figure 2:** The domain concept template in the SMW
multi-linguality we had to face in APOSDLE. These elements ("English Description", "English Name", "Name", "Short Description", "Synonyms") are modelled as String attributes in the SMW. The top-level part of the template concerns the relation of the concept with other concepts in the (still informal) ontology. In order to help the modelling team we predefine in the template some typical relations such as "Is a" and "Is part of". We also provide a general "Is related to" relation and ask the modelling team to instantiate (rename) it to specific, domain dependent, relations, when possible. All these relations are modelled as Relations also in the SMW.

Although the information required in the template is very basic, it allows to guide the modelling team to provide all the information that was needed by the particular application. In addition mentioning explicitly the "Is a" and "Is part of" relations prevented the modelling team to incur in one of the typical modelling mistakes of non expert designers when they use the graphical environment of ontology editors to create is-a taxonomies, that is, to actually use a mixture of is-a, is-part-of, and other specialisation relations in the creation of the taxonomy, thus making the taxonomy more similar to a directory structure than to an is-a taxonomy. The usage of the methodology in different domains could require the extension of these templates or the creation of new ones. Creating templates in the SMW is not a difficult operation. The challenging aspect here is to be able to design appropriate templates which can guide the modelling team in the process of providing appropriate descriptions of the different elements.

Another advantage of using a SMW are checks (also in automatic) to evaluate the quality of the informal descriptions produced. In addition, the domain concepts’ relevance and unambiguity can be checked using the SMW. The verbal descriptions of concepts have proved to be very useful to help with this task.

**Formal Modelling.** The content of the SMW, created in the informal modelling phase, is then automatically translated into OWL to produce the formal ontology. The idea behind the translation is to transform each concept description (concept template) of the informal domain model into the corresponding OWL class. The starting point for creating automatically the OWL ontology from the informal domain model is the built-in SMW Export pages to RDF functionality. Using this functionality, it is possible to generate a document in OWL/RDF format containing information on the relations and attributes used in the pages selected to be exported. However, the model of knowledge used by SMW is quite different from the one we developed for the modelling methodology. In a nutshell pages of SMW are by default seen as instances, and not as classes. Therefore the straightforward application of the Export pages to RDF functionality produces, for each concept template, an instance of a top class smw:Thing, instead of an OWL class. Similarly, the "is a" relation is mapped by the Export pages to RDF functionality to an object property named is a, while in our approach this relation needs to be mapped to the RDFS subClassOf relation. For this reason we developed a Java tool which post-processes the files obtained using the Export pages to RDF functionality, and generates an OWL ontology.
Once the OWL ontology is created, the modelling team can correct the hierarchy and the relations in the informal domain model. Missing or redundant concept correlations are updated in the formal model accordingly. Note that the informal model is not kept up-to-date with changes made in the formal model at this stage because reflecting the changes made into an OWL ontology back to SMW is not a trivial task. We still have to address it, partly because the interaction with the formal model should be narrowed down to checks of the hierarchy, of the granularity of concepts and relations, and of formal checks of consistency and detection of unsatisfiable concepts.

A similar approach is proposed by the methodology for modelling processes and learning goal. The template we proposed in APOSĐLE for informal task description, asks for possible mappings between a task and required domain elements of the domain model. These mappings are then used in the formal modelling phase to create the learning goals model with an ad-hoc tool developed for APOSĐLE, the TACT tool. At the current development stage of the methodology and the supporting tool for it, the YAWL task model is manually created from the informal descriptions provided in the SMW.  

**Lessons Learnt.** After using the methodology for the development of the APOSĐLE knowledge base for three APOSĐLE application partners, we have collected qualitative feedback from different members of the modelling teams. On the positive side, the SMW allows for sharing information among all the members of the modelling team. Furthermore, templates helped guiding the people in charge of modelling to provide complete descriptions of the elements of the models. Finally, by using the SMW as informal modelling tool, the transfer from the informal domain model to a formal domain ontology was almost effort free, and the interaction with the formal language was kept very low. Last, but not least, the entire modelling process is well documented in the SMW. On the negative side, the SMW does not provide any built-in textual or graphical overview of the overall structure that can be easily used to show the taxonomy of domain concepts. Although some visualisation tools for MediaWiki are available, they can not be easily adapted to work with our approach. In addition the SMW does not support a user-friendly revision of filled templates. This makes the refinement of the informal models, and the use of an iterative process towards modeling, quite laborious. The integration of tasks with formal modelling tools and the automatic translation into YAWL workflows was more challenging than the one for the domain model. While the SMW makes easy to represent and export declarative knowledge about concepts, in the form of RDF triples, processes have also a procedural aspect which is difficult to reconstruct starting from the (declarative) description of the single task components. Summarized, the SMW’s coherent interface to describe the elements of the different data models proved to be quite useful for informal modelling. SMW also helped in producing an effective integrated development of the models. Solving the issues of visualisation and revision of the models or support for an automatic translation of the task model from the SMW to YAWL would make the tool even more user-friendly.
4 Conclusions

In this paper we introduced a methodology to support the integrated modelling process of ontologies and workflows within APOSDLE. Using this methodology, three APOSDLE Application Partners built their own Knowledge Base:

- CCI has developed an OWL ontology with 144 concepts and 14 relations, an OWL ontology with 304 learning goals, and a YAWL model containing 42 tasks for the environmental consultancy domain (REACH area);
- EADS has developed an OWL ontology with 43 concepts and 28 relations, an OWL ontology with 99 learning goals, and a YAWL model containing 24 tasks for the electromagnetism simulation domain;
- ISN has developed an OWL ontology with 43 concepts and 28 relations, an OWL ontology with 211 learning goals, and a YAWL model containing 40 tasks for the innovation and knowledge management domain.

The methodology has also been successfully applied by two APOSDLE Technical Partners to build two additional Knowledge Bases for academic/demonstrative purpose:

- City University London has developed an OWL ontology with 78 concepts and 52 relations, an OWL ontology with 114 learning goals, and a YAWL model containing 78 tasks for the RESCUE (requirements engineering) domain;
- University of Twente has developed an OWL ontology with 71 concepts and 23 relations, an OWL ontology with 50 learning goals, and a YAWL model containing 19 tasks for the statistical data analysis domain.

Our novel approach of developing tools for informal modelling, and their tight integration with tools for formal modelling, can provide a concrete support to the modelling process of ontologies and workflows in other domains, e.g. some preliminary encouraging results have been obtained in a biomedical domain.

Acknowledgements

This work has been partially funded under grant 027023 in the IST work programme of the European Community.

References