# Applications of Ontology Design Patterns in the Transformation of Multimedia Repositories

Agnieszka Ławrynowicz[1] and Raúl Palma[2]

[2]Poznan University of Technology, Poznań, Poland alawrynowicz@cs.put.poznan.pl
[2]Poznan Supercomputing and Networking Center, Poznań, Poland
rpalma@man.poznan.pl

**Abstract.** The development of ontologies from scratch is a very inefficient approach. Hence, ontology development is increasingly being conducted by reusing existing ontological and non-ontological resources as it lowers the time and cost of developing new ontologies, avoids duplicating efforts and ensures interoperability. Similarly, the emergence of ontology design patterns has facilitated the reuse of best practices in ontology engineering, improving the quality of the developed ontology. In this paper we show how different ontology design patterns along with state-of-the-art ontology engineering methodologies help during the construction of ontologies, by describing the development process of the Digital Multimedia Repositories Ontology (DMRO). We show the applicability of the developed ontology by using it as the basis for the transformation of Videolectures multimedia repository to RDF data.

## 1 Introduction

The widespread use of ontologies in different applications and domains in the last years has led to an increasing interest of researchers and practitioners in the development of ontologies. To speed up the development process, existing resources in the form of ontologies and other non-ontological forms, such as thesauri, lexica and DBs, as well as best practices encoded in the form of ontology design patterns (ODPs) are increasingly being used. Reusing existing resources is an important aspect that is progressively better supported by the growing availability of ontology design patterns and ontologies for different domains, the availability of well established upper-level ontologies, and better support for transformation of non-ontological resources to ontological format. Besides speeding up the development process, reusing existing resources has many benefits, including lower development costs, interoperability with other ontologies and better quality of the developed ontology. Even more, we argue that in addition to the availability of such resources, ontology engineers can benefit from the availability of good use cases showing how to apply all of these resources in practical applications.

For example, one of the pilot applications of e-LICO project[1] aimed at describing multimedia artifacts in digital multimedia repositories, in terms of their usage, reviews and content type, as well as their relation with related events and agents, that would serve as background knowledge to be used in semantic data mining tasks. Although there are different ontologies available for describing multimedia artifacts (e.g.,

[1] http://www.e-lico.eu/

COMM, M3O), they focus mainly on a description of the artifacts characteristics with a detail that falls outside the application requirements, instead of focussing on the usage information of these resources which is key for data mining in multimedia repositories. Additionally, we can find ontologies and vocabularies describing parts of the required knowledge, such as agents or events (e.g., FOAF, SWRC), but they are generic or conceived for other applications, and are not related to multimedia resources. Consequently we developed the Digital Multimedia Repositories Ontology (DMRO), but reusing as much as possible knowledge from these and other ontologies and vocabularies.

Therefore, the goal of this paper is twofold: (i) to present the Digital Multimedia Repositories Ontology and show how it has been used in practice and (ii) to provide an example of the ontology development process where existing resources and ODPs were widely re-used. We show the applicability of the ontology by using it as the basis for the transformation of a multimedia repository coming from VideoLectures.Net portal to RDF data. VideoLectures.Net[2] is one of the largest scientific and educational video Web sites, mostly hosting lectures given by scholars and scientists at conferences, summer schools, and other events. The dataset used in our application of DMRO was prepared for the e-LICO data mining challenge [11] on recommender systems for the lectures.

The rest of this paper is structured as follows: Section 2 discusses the development of DMRO, Section 3 describes the modules of DMRO, Section 4 describes the application of DMRO in transformation of Videolectures.net data, Section 5 discusses the related work, Section 6 contains lessons learnt, and Section 7 concludes the paper.

## 2 Development of DMRO

After an analysis of existing ontology engineering methodologies (see [8,2] for a detailed comparison), we decided to follow the NeON Methodology for Building Ontology Networks ([7,2]) during the development of DMRO for several reasons: (i) its flexibility and adaptability to different development scenarios, specially those with focus on reusing (and reengineering) existing knowledge resources as well as best practices in ontology engineering ; (ii) the clear guidelines provided for every task with concise information cards, templates, heuristics and examples; (iii) and the technological support available for it through the NeOn Toolkit[3].

### 2.1 Requirements

In line with this methodology, we started the development of DMRO by collecting and analyzing an initial set of requirements using a structured document, called the Ontology Requirements Specification Document (ORSD)[4]. The ORSD document covers the following topics concerning the ontology: purpose, scope, implementation language, intended end-users, intended use cases, reusing ontology statements.

In particular, the goal of the engineering of DMRO is to use it for the tasks of: a) design of recommendation and personalization solutions for digital multimedia repositories; b) meta-learning/meta-mining on data mining experiments repositories; c) testing

[2] http://videolectures.net/
[3] http://neon-toolkit.org/
[4] Available at http://129.194.69.119/public/dmro/DMRO_ORSD.pdf

semantic data mining [9] algorithms. The scope of the ontology is on the applications in recommender systems, personalization, and adaptive faceted browsers of digital resources, and therefore the ontology being built is an application domain ontology.

DMRO should be as compatible as possible with established ontologies and vocabularies that cover relevant aspects in the domain of DMRO, such as Dublin Core[5], FOAF[6], RDF Review[7], OBO Relation Ontology[8], and OAI-ORE[9]. Consequently, it should reuse terms from these resources whenever is possible.

Additionally, in order to encourage the re-use of DMRO and facilitate its specialization for particular applications, it should follow a modularized approach, thus it is expected to build a set of small modules for representing DMRO concepts, such as: multimedia resources, users, events, reviews, Web Usage Mining related concepts, and the domain topics (in our case VideoLectures.Net topic category). Similarly, in order to ensure compatibility with existing tools and vocabularies, the current standard OWL 2 should be used as the implementation language.

The development was scheduled in two cycles/iterations within the lifetime of the e-LICO project. Each of these cycles was scheduled using the Gontt tool[10], which enables the graphical representation of an ontology project schedule in the form of a Gantt chart. Figure 1 illustrates the schedule for the first cycle.

The next section describes the resources re-used during the ontology construction.

### 2.2 Reusing Existing Resources

From the analysis of ontologies relevant for our domain and application, we couldn't find one single ontology covering all aspects of DMRO, but we found several ontologies modeling parts that could be reused. Hence, we decided to reuse individual statements instead of a whole ontology. Moreover, we decided to use an upper ontology that will model generic concepts that can be specialized for DMRO. The advantage of using an upper ontology will reflect in better interoperability and foster the reusability of DMRO. We also re-used and applied different ontology design patterns relevant for DMRO, which allowed us to follow the best practices in ontology engineering when modeling DMRO. Finally, we re-used a non-ontological resource: a dataset based on the data from Videolectures.net portal.

**Upper Ontology**

We used DOLCE Ultralite (DUL)[11] as an upper ontology. DUL is a very light version of DOLCE and DnS, which provides a simplification and an improvement of some parts of DOLCE Lite-Plus library[12], and Descriptions and Situations ontology [13]. DUL provides a set of upper level concepts that can be the basis for easier interoperability

---

[5] http://dublincore.org/documents/dcmi-terms/
[6] http://xmlns.com/foaf/spec/
[7] http://vocab.org/review/terms.html
[8] http://obofoundry.org/ro/
[9] http://www.openarchives.org/ore/
[10] http://neon-toolkit.org/wiki/Gontt
[11] http://ontologydesignpatterns.org/wiki/Ontology:DOLCE%2BDnS_Ultralite
[12] http://dolce.semanticweb.org
[13] http://www.ontologydesignpatterns.org/wiki/Ontology:DnS

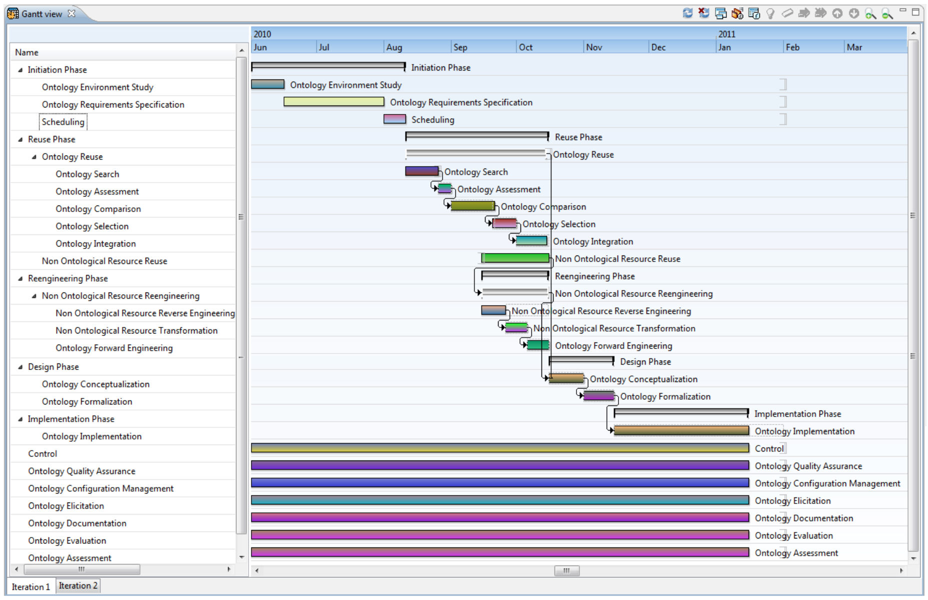


**Fig. 1.** First cycle of the development process of DMRO

among many middle and lower level ontologies. DOLCE-Ultralite falls within OWL-DL language, but it only uses OWL-Lite and disjointness constraints.

**Ontology Statements**

We identified relevant statements to be re-used from the following ontologies and vocabularies: (i) Dublin Core for modeling general metadata properties of resources; (ii) FOAF - for modeling the users, authors, participants, their personal data, and social aspects; (iii) SWRC ontology - for events; (iv) OBO Relation Ontology - for standard relations; (v) DCTYPE - for modeling collections; (vi) ORE - for modeling aggregations of resources; and (vii) RDF Review Vocabulary - for the reviews, comments and feedback. Additionally, we identified the following modules from the myExperiment Ontology: (viii) Viewings & Downloads - for modeling the usage of resources; (ix) Annotations - for modeling tags.

Finally, we have used IOLite extension of DUL[14] and CSnCs ontology[15], as an inspiration for modeling DigitalResources. IOLite is an ontology of information objects and realizations, plugin to DOLCE-Ultralite. CSnCS (Computer Science for Non-Computer Scientists) is a knowledge hierarchical repository of concepts in the domain of Information Technology for End Users, which also uses DUL as an upper ontology.

In order to select the statements to be reused we conducted a research to identify candidate ontologies, partially with the help of Watson plugin for NeOn Toolkit[16] (Figure 2 shows statements from CSnCS regarding DigitalResources), and followed a

[14] http://www.loa-cnr.it/ontologies/IOLite.owl
[15] http://www.let.uu.nl/lt4el/content/files/CSnCSv0.01Lex.zip
[16] http://neon-toolkit.org/wiki/Watson_for_Knowledge_Reuse

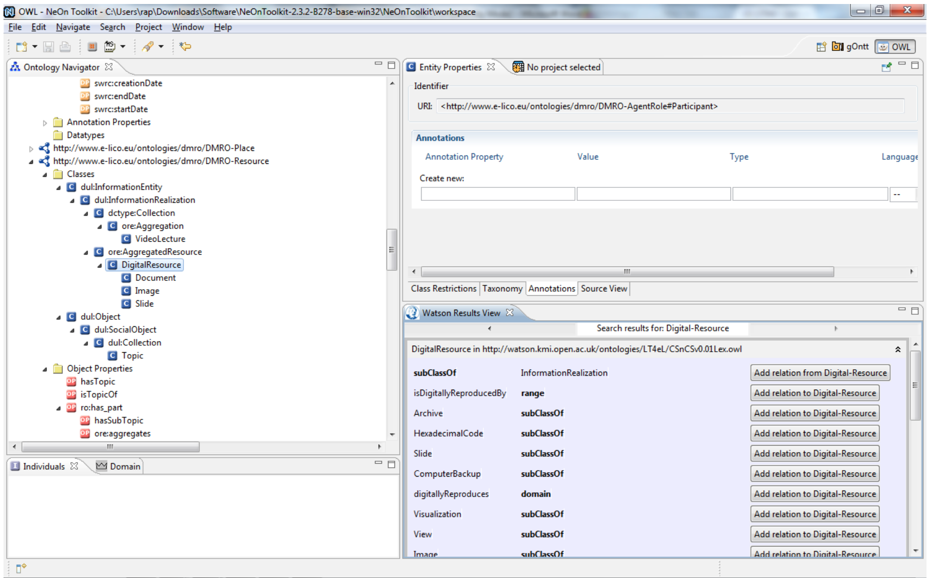


**Fig. 2.** Ontology statements search result in NeOn Toolkit

selection criteria consisting on the simplicity, coverage and popularity of the candidate ontologies, building from the dimensions proposed by the NeOn methodology for selecting ontological resources (i.e., understandability, integration effort and reliability).

**Ontology Design Patterns**

We re-used the following ontology design patterns from http://ontologydesignpatterns.org:

– Agent role[17] to represent agents and the roles they play. It was used with the following competency questions:
  - which agent does play this role?
  - what is the role that played by that agent?
– Participant role[18] to represent participants in events holding specific roles in that particular event. It was used with the following competency questions:
  - What is the role of this object in this event?
  - What is the object holding this role in this event?
  - In what event did this object hold this role?
– Tagging[19] to represent a tagging situation, in which someone uses a term, from a list of a folksonomy, to tag something (or the content of something). It was used with the following competency questions:
  - Who is tagging (the content of) what?
  - By using what term from what folksonomy?

17 http://ontologydesignpatterns.org/wiki/Submissions:AgentRole
18 http://ontologydesignpatterns.org/wiki/Submissions:ParticipantRole
19 http://ontologydesignpatterns.org/wiki/Submissions:Tagging

    - Which polarity has the tagging?
- Place[20] to talk about places of things. It was used with the following competency questions:
    - Where is a certain thing located?
    - What is located at this place?
- Topic[21] to represent topics and their relations. It was used with the following competency questions:
    - What is the topic of something?
    - What topics are included in this one?
    - What are the topics near to that one?

**Non-ontological resources**

We re-used a dataset based on the data snapshot from VideoLectures.net taken in August 2010, which consisted of a database dump into several CSV files, which contained data from 7 different tables (authors, categories, events, lectures_train, lectures_test, authors_lectures and categories_lectures). The snapshot contained 8,105 video lectures, where 5,286 lectures were manually categorized into a taxonomy of around 350 scientific topics. This dataset was used as background knowledge during the specification phase of DMRO development. In particular, the competency questions and the pre-glossary of terms in the ORSD document were prepared largely on the basis of the dataset as an inspiration.

## 3 Overview of DMRO

DMRO consists of a set of 6 interrelated modules, which are imported from the main ontology file available at http://129.194.69.119/public/dmro/DMRO.owl. Next, we shortly describe each of the modules.

**Resource.** This module describes multimedia resources (see Fig. 3). Despite using DOLCE-Ultralite (DUL) as upper ontology, we have also used 'IOLite', extension of DUL, as an inspiration for modeling the concept *DigitalResources*. We also considered other ontologies such as COMM or M3O for modeling multimedia resources; however their focus on a detailed description of these resources falls outside the scope of our ontology as the ORSD shows. We have re-used 'Topic' ontology design pattern for modeling the topics, and statements from DCTYPE, ORE and RO. We have also used ontology CSnCs 'Computer Science for non-Computer Scientists from Project LT4eL (http://www.lt4el.eu/)' as inspiration.

**Event.** This module describes events, and was created using DUL as upper ontology, and by re-using statements from SWRC and DC (see Fig. 4). Additionally, we have specialized the *SWRC:Lecture* class with various types of lectures identified during the specification phase.

**Agent-Participant-Role.** The module describes agents and their roles as participants in events, based on the ParticipantRole ontology design pattern (see Fig. 5). Besides using DUL as upper ontology, we have re-used statements from FOAF and

[20] http://ontologydesignpatterns.org/wiki/Submissions:Place
[21] http://ontologydesignpatterns.org/wiki/Submissions:Topic

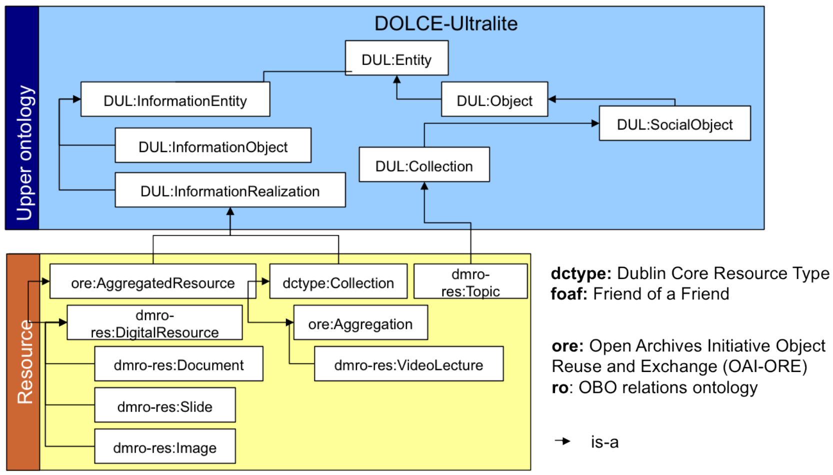


**Fig. 3.** An illustration of the part of DMRO's Resource module.

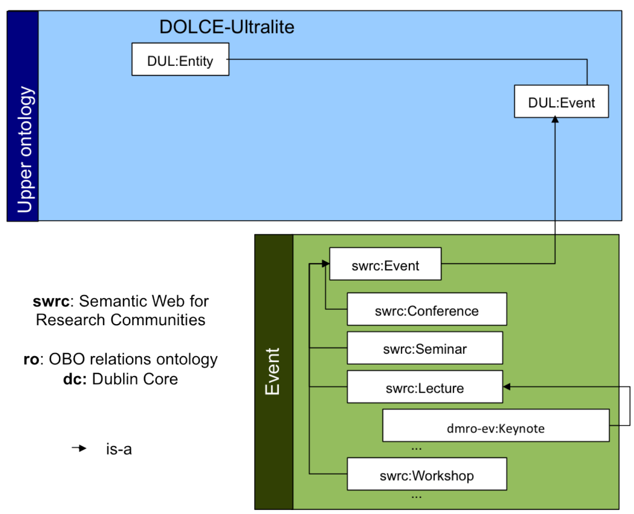


**Fig. 4.** An illustration of the part of DMRO's Event module.

RO. For example, instead of using the class *DUL:Agent* (which is a subclass of *DUL:Object*), we reused the more popular *FOAF:Agent* class and its subclasses. Referring to the suggestions of N.F. Noy and D.L. McGuinness on "an instance or a class" discussion[22] we decided to model agent roles, such as author and participant (e.g. dmro-apr:Presenter, dmro-apr:Author) as instances and not classes in the ontology as these are the most specific elements that are going to be represented in the knowledge base, i.e., they constitute the most specific elements in the answer to the competency

[22] http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html

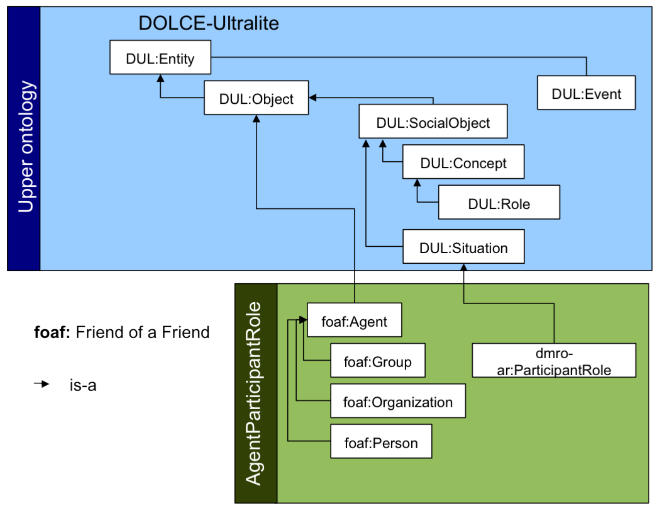


**Fig. 5.** An illustration of the part of DMRO's Agent-Participant-Role module.

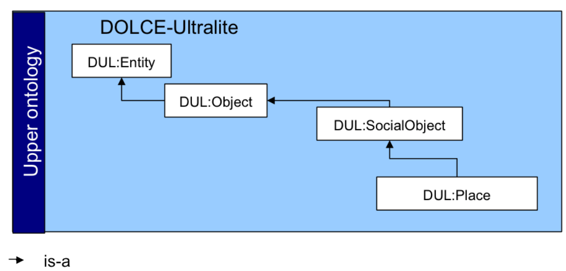


**Fig. 6.** An illustration of the part of DMRO's Place module.

questions related to agents roles (e.g., what is the role that played by that agent?). Before being taken, this design decision was discussed with the help of Cicero, an argumentation tool that is part of the e-LICO collaborative ontology development platform.

**Place.** This ontology module represents generic locations and was created based on the Place ontology design pattern, which is already implemented in DUL (see Fig. 6). So, we reused the relevant DUL statements and also some statements from RO (e.g., *located_in* property)

**Review.** This module was created by re-using the RDF Review Vocabulary and using DUL as upper ontology (see Fig. 7). We have also re-used statements from FOAF, DC and myExperiment Ontology. Mainly, the classes *Comment*, *Feedback* and *Review* from the RDF Review Vocabulary have been modeled as types of annotations (as in myExperiment Ontology).

**Annotation.** This ontology module was created by reusing statements from the Viewings & Downloads and Annotations modules of the myExperiment Ontology to model different resource annotations, including usage information and tagging, and us-

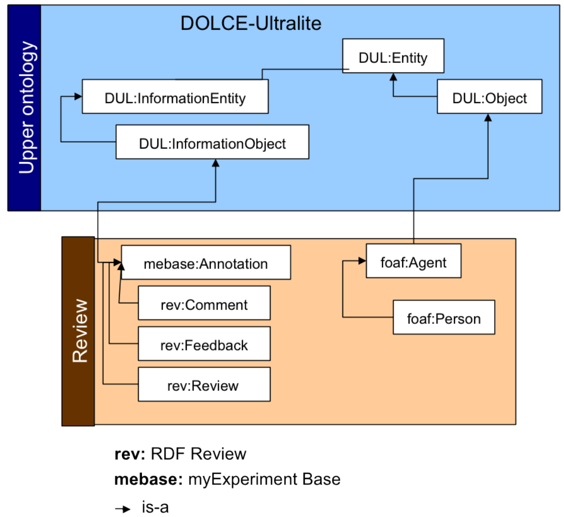


**Fig. 7.** An illustration of the part of DMRO's Review module.

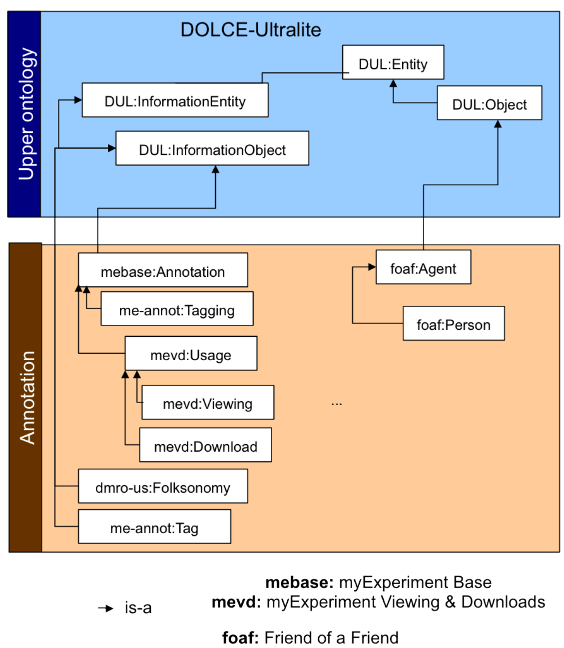


**Fig. 8.** An illustration of the part of DMRO's Annotation module.

ing DUL as an upper ontology (see Fig. 8). We also reused statements from the RO ontology, and some concepts from the Tagging ontology design pattern.

**Table 1.** Mapping for author table of Videolectures.net

| Column | Ontology Element | Element Type | OWL Axiom Type |
|---|---|---|---|
| id | foaf:Person | Individual | Declare & ClassAssertion |
| name | foaf:name | DataProperty | DataPropertyAssertion |
| email | DMRO-AgentParticipantRole:email | DataProperty | DataPropertyAssertion |
| organization | foaf:Organization | Individual & ObjectProperty & DataProperty | ClassAssertion & ObjectPropertyAssertion & DataPropertyAssertion |

## 4 Application of DMRO

DMRO has been used as the basis for (i) representing Videolectures.net topic hierarchy and (ii) transforming Videolectures.net dataset into RDF. The former was a simple test during which we generated instances of the concept *Topic* and re-used properties from SKOS vocabulary for the representation of hierarchical links (i.e., *skos:broader* and *skos:narrower*). The resulting knowledge base is available at http://129.194.69.119/public/dmro/DMRO-VLNetCat.owl.

For the transformation of Videolectures.net dataset into RDF we generated manually mappings between the terms from DMRO and terms from the dataset (e.g., columns in tables). The mappings range from simple alignments between a table column and a dataProperty in the ontology, to more complex alignments that created several ontology axioms for a value in a table column. Table 1 contains sample mapping that was created for the authors table[23]:

During preparation of the mapping, we have introduced several changes into DMRO (some of them are discussed in the Cicero argumentation tool within e-LICO collaborative ontology development platform) which allowed us to represent more accurately Videolectures.net data, and more generally digital multimedia repositories. The actual transformation was performed programmatically based on the conceptual mapping specified. The resulting knowledge base is available at http://129.194.69.119/public/dmro/DMROKB.owl and has already been used to test semantic data mining methods.

## 5 Related Work

Previous efforts in the multimedia domain, have led to the development of different ontologies, although most of them have been mainly focused on the detailed description of multimedia artifacts and only in few cases ontology design patterns were reused during the development process. For instance, COMM ontology [1], which is composed of multimedia patterns covering different media types (e.g., visual, audio or text), is based on DOLCE foundational ontology and the MPEG-7[24] standard. COMM reused two main design patterns: Descriptions & Situations (D&S) [5] and Ontology of Information Objects (OIO) [4], which were specialized to create multimedia design patterns

[23] The complete set of mappings is available at http://129.194.69.119/public/dmro/mapping.xls
[24] http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm

for decomposition of multimedia content into segments, the annotation of these segments, as well as basic patterns to formalize the notion of digital data and algorithm. M3O [3] aimed for rich presentations in the web (SMIL, SVG, Flash), covers multimedia, audio/music, image, video, and re-uses DUL. M3O also followed a pattern-based approach to ontology design, and used six patterns: Decomposition, Annotation, Information Realization, Data Value for representing complex values, Collection, and Provenance. Although these ontologies were useful examples in the domain, they were found too detailed and focused in describing multimedia resources, which was not the goal of DMRO as discussed in the introduction.

Other relevant ontologies include the Media Resource Ontology [12], a W3C initiative aimed at integrating data resources related to media on the Web that covers multimedia, audio/music, video, and provides mappings to various multimedia metadata formats (e.g., MPEG-7). Similarly, we can find also various initiatives that focus on the transformation of MPEG-7 standard to ontological format, such as [10] and [6].

## 6 Discussion

The NeON methodology was very useful in identifying relevant existing resources. The competency questions facilitated identification of relevant ODPs as well as ontologies to re-use. In the latter case, we found that pre-glossary of terms (a part of an ORSD document) is especially useful for this task. The terms extracted from competency questions (and answers to them) allowed us to efficiently search for ontologies.

The major lesson learnt and an idea for improvements concerns further re-using of identified ODPs and ontologies and deals with availability of guidelines on a vocabulary that could be used to instantiate or specialize chosen ODPs. Since ODP entities in principle constitute templates, they need to be further refactored while being incorporated to a developed ontology. This needs modeling decisions concerning the choice of a vocabulary and namespaces. We would find it very useful if such guidelines existed, for instance, pointing to most popular terms that are used by ontology engineers to instantiate a given ODP, possibly indicating also the domain of the ontology where they were used. The information on the domain is important due to differences in term popularity across different domains w.r.t. the same pattern. This would allow not only to re-use an ODP, but also to: i) refactor it such that most widespread vocabulary is used or/and ii) proper vocabulary for given domain is used. Similarly, we think that re-using ODPs would be greatly facilitated if example ontologies re-using a given ODP where listed, most preferably together with the information on a domain.

Currently, http://ontologydesignpatterns.org has placeholders for 'Known uses' of an ODP and for 'Examples (OWL files)', but they are often empty. Besides of that, we would like to stress the aspect of popularity of "known uses" and "examples" in the context of a given domain to make it easier for an ontology engineer to choose proper vocabulary. For instance, we chose to use myExperiment for representing tagging since we found the annotations ontology of myExperiment covering many concepts we wanted to model in the Annotation module, including tags. However, as one of the reviewers suggested, we could have rather used the Tag ontology (http://www.holygoat.co.uk/owl/redwood/0.1/tags/) which is already widespread in Linked

Data on multimedia such as in DBTune datasets, if we were aware of its popularity. Listing it as an example accompanying the ODP would help in making this choice.

## 7 Conclusions

This paper presented the Digital Multimedia Repositories Ontology and its application in the transformation of the data from Videolectures.net into a knowledge base represented in RDF. Importantly, by discussing the ontology development process and its application, it also provided a use case of re-using existing ontological and non-ontological resources and ODPs for ontology construction.

The DMRO-based RDF version of the Videolectures.net dataset provides proof-of-concept of the coverage of the DMRO terms, and their suitability to represent required knowledge on the Videolectures.net use case. It may be used to test semantic data mining approaches, some of them already developed within e-LICO, as well as in the experiments in the digital multimedia repositories.

**Acknowledgments** The research reported in this paper has been supported by the EU e-Lico project (http://www.e-lico.eu/, FP7 STREP ICT-231519).

## References

1. R. Arndt, R. Troncy, S. Staab, L. Hardman, and M. Vacura. COMM: Designing a well-founded multimedia ontology for the Web. In *In Proceedings of the 6th International Semantic Web Conference (ISWC2007), Busan, Korea*, pages 11–15, 2007.
2. M. del Carmen Suárez-Figueroa. *NeOn Methodology for Building Ontology Networks: Specification, Scheduling and Reuse*. PhD thesis, Universidad Politécnica de Madrid, Spain, 2010.
3. D. Eißing, A. Scherp, and C. Saathoff. Integration of existing multimedia metadata formats and metadata standards in the M3O. In *Proceedings of the 5th international conference on Semantic and digital media technologies*, SAMT'10, pages 48–63, Berlin, Heidelberg, 2011. Springer-Verlag.
4. A. Gangemi, S. Borgo, C. Catenacci, and J. Lehmann. Task taxonomies for knowledge content D07. Technical report, Metokis Project, 2004.
5. A. Gangemi and P. Mika. Understanding the Semantic Web through Descriptions and Situations. In *Proceedings of ODBASE03 Conference*, pages 689–706. Springer, 2003.
6. R. Garca and O. Celma. Semantic integration and retrieval of multimedia metadata. In S. Handschuh, T. Declerck, and M. Koivunen, editors, *Proceedings of the ISWC 2005 Workshop on Knowledge Markup and Semantic Annotation (Semannot'2005)*, volume 185, pages 69–80. CEUR Workshop Proceedings, 2005.
7. A. Gómez-Pérez and M. del Carmen Suárez-Figueroa. Scenarios for building ontology networks within the NeOn methodology. In *K-CAP*, pages 183–184, 2009.
8. A. Gomez-Perez, M. Fernandez-Loez, and O. Corcho. Methodologies and methods for building ontologies. In *Ontological Engineering*, Advanced Information and Knowledge Processing, pages 107–197. Springer Berlin Heidelberg, 2004.
9. M. Hilario, A. Kalousis, N. Lavrac, A. Lawrynowicz, J. Potoniec, and A. Vavpetic. Semantic data mining tutorial at ECML/PKDD'2011.
10. J. Hunter. Adding multimedia to the Semantic Web: Building an MPEG-7 ontology. In *SWWS*, pages 261–283, 2001.
11. T. Šmuc, N. Antonov-Fantulin, and M. Morzy, editors. *Proceedings of the ECML/PKDD Discovery Challenge Workshop 2011*, Athens, Greece, September 2011.
12. www.w3.org. Ontology for media resource 1.0. http://www.w3.org/TR/mediaont-10/, 2012.