Ontologizing Metadata for Assistive Technologies

The OASIS Repository*

Alexander Garcia Castro and Immanuel Normann
Department of Linguistics and Literature
University of Bremen, Germany
{cagarcia, normann}@uni-bremen.de

Joana Hois and Oliver Kutz
SFB/TR 8 Spatial Cognition
University of Bremen, Germany
{joana, okutz}@informatik.uni-bremen.de

Abstract

This paper discusses the design challenges encountered when building an ontology repository for an application framework devoted to assistive technologies that can be browsed and queried in a highly heterogeneous and expressive way. As a main vehicle to achieve this goal we propose to use metadata and meta-reasoning. We analyze how metadata can be used in the context of open repositories of ontologies, and how it can and needs to be extended in various ways. In particular, we study a redesign of the Ontology Metadata Vocabulary OMV by restructuring and enriching it with the ABC ontology and domain-specific categories for assistive technologies. Examples of an elderly assistance system illustrate how this extended OMV can support more sophisticated reasoning and querying over repositories of ontologies for assistive applications.

1. Introduction

As The Semantic Web (SW) envisions a metadata-rich Web where human-readable content will have machine-understandable semantics there has been an increasing number of OWL ontologies [18] responding to those knowledge representation requirements. Wang et al. collected 1275 files, both OWL and RDF schemas, in 2005; a more recent counting, based on web crawling, gave an impressive result of over 6000 validated OWL ontologies (Backer et al., unpublished data); by the same vein Swoogle [6] hosts 2,563,125 Semantic Web Documents (SWD) [16]. These growing numbers, which reflect the intrinsic need of the SW for ontologies, have fostered a number of research projects aimed at supporting re-usability, better modularization as well as intelligent storage and retrieval for the encoded knowledge. To this end the design and development of an agreed upon metadata for describing ontologies is critical. Several repositories should be able to facilitate not only the discovery of reusable components, entire ontologies or just portions of them, but also interoperability across repositories. In a recent effort to unify the description of ontologies, the Ontology Metadata Vocabulary Consortium [14] proposed a set of descriptors that follows the principles of the Dublin Core. This is a step in the right direction as most ontologies exist without any additional information in the form of metadata. We advocate the use of OMV and support further refinements and extensions of this proposal. Although the extensions we propose for the Ontology Metadata Vocabulary (OMV) [14] are, in principle, domain independent, our main interest lies in supporting repositories with a particular focus on supportive applications for elderly and disabled people. As part of the OASIS project (Open architecture for Accessible Services Integration and Standardization) [12], we are currently developing a repository for ontologies aiming to describe spatial-temporal scenarios as well as medical and technological information related to elderly and disabled populations, i.e. users with special needs. Within this context we are working on a repository of ontologies that provides structured access and easy-to-extend descriptions for the ontologies it hosts. The following principles are important when using metadata for structuring ontology repositories: (i) ontology standards should be kept intact; (ii) the metadata-core is connected to various meta-descriptions through alignments mediators; (iii) meta-descriptions structure specific parts of knowledge; (iv) meta-descriptions need to support query languages and reasoning; (v) meta-descriptions may again be ontologies.

In this paper we follow and adopt the OMV initiative; it is here understood as a core metadata language. We argue that it needs to be extended in two dimensions: horizontally, by adding specific metadata meeting our application needs (i.e. motivated by OASIS), and vertically, by refin-

*For the work reported in this article we gratefully acknowledge the financial support of the European Commission through the OASIS project (Open Architecture for Accessible Services Integration and Standardisation) and the Deutsche Forschungsgemeinschaft through the Collaborative Research Center on Spatial Cognition (SFB/TR 8). The authors would like to thank John Bateman for fruitful discussions.

ing a specific part of this extended metadata by means of an alignment with a purpose-specific meta-description or ontology. We are particularly interested in the OMV metadata that addresses temporal aspects and is related to the evolution of ontologies, we refine this section vertically. As there are several similarities between the concept of a digital library and that of an ontology repository, we selected, for the purpose of this extension, the ABC model [18]. The ABC ontology was initially proposed as a core ontology for digital libraries.

2. Related Work: Repositories and Metadata

Repositories, within the context of the SW, should offer more than just data storage. The Ontolog community, a virtual community of practice of ontology experts, discussed the matter and agreed that the purpose of an Open Ontology Repository (OOR) is to provide an architecture and an infrastructure that supports: a) the creation, sharing, searching, and management of ontologies, and b) linkage to database and XML Schema structured data and documents [12]. Currently there are some ontology repositories over the web, however none of them complies with those requirements agreed upon during the last Ontolog Summit [13]. For instance, Swoogle provides a single entry-point to several semantic web documents (ontologies), but does not offer any validation, as there is no quality control over the exposed material; nor does it facilitate query or editing operations. Swoogle's query approach for finding ontologies is based on (sub-)string search and link-based reference counting; once the document has been found it does not support any further operation. It also allows the composition of queries via the REST interface. OntoSelect [2] offers a similar approach; it presents the user with a basic overview of web-accessible ontologies. The collection can be browsed by: ontology name (derived from owl:Ontology/rdfs:comment); format (from the ontology URL); human language (from rdfs:label); number of labels, classes, properties, or included ontologies (owl:imports). Currently OntoSelect hosts 1530 ontologies. The TONES repository, developed as part of the TONES project [5], hosts 185 ontologies. It aims to provide a reasonable amount of ontologies for testing purposes, emphasizing reasoning techniques. This repository also supports the REST interface for programmatic access. Ontologies can be selected and sorted by means of metrics for expressivity, class and property restrictions and axioms, logics, and individuals. A novel approach is provided by Rubin et al. [7] with Bioportal. Not only does it provide access to several ontologies, but it also facilitates online editing operations such as annotation of ontologies in the form of marginal notes currently only available for classes. In [15], a lightweight metadata ontology for an ontology repository of a multi-agent system is presented. The ontology consists of four classes: Conceptualization, Ontology, Person, and Representation. The Ontology is described by a title, version number, language, author, and textual description. The Person defines the author of an ontology, while the Conceptualization class defines an abstract view, on which the ontology is based. The class Representation specifies the encoding of Ontology, Person, and Conceptualization. This repository also supports the REST interface.

Although existing ontology repositories aim to provide access to semantic web documents by means of similar query facilities, they diverge in the methods and techniques employed for gathering these documents and making them available; each one of them interprets and uses metadata in a different manner. For instance, Swoogle defines three categories of metadata; (i) basic metadata, which considers the syntactic and semantic features of an ontology, (ii) relations, which consider the explicit semantics between individual ontologies, and (iii) analytical results such as SWO/SWDB classification, and ontologies [6]. Both, TONES and OntoSelect, also rely on structural metadata; however, the use of this metadata is limited to a subset of it. As Bioportal supports the involvement of communities of practice it makes use not only of structural metadata but also of that metadata describing how the community has engaged. For instance, descriptions of those who have defined a new relationship by means of a marginal note in a way that it facilitates to establish rankings of confidence.

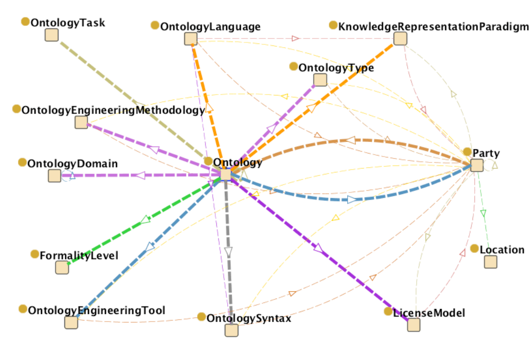

Figure 1. A view of the OMV

The ontology metadata standard OMV aims to enhance the retrieval, identification, and reuse of ontologies from the Web effectively and efficiently. Modularity is one of its design principles: OMV distinguishes between the OMV Core and various OMV Extensions. We will consider here only OMV Core and refer to it, for convenience, simply as OMV. The main class in OMV is Ontology. With its datatype properties URI, version, and resourceLocator it allows to describe an ontology file in a particular version at a particular physical location. Possible relations between instances of the class Ontology are: useImports, hasPriorVersion, isCompatibleWith, and isBackward-Compatible. Moreover, there are ten properties relating Ontology to ten other classes as depicted in Fig. 1.

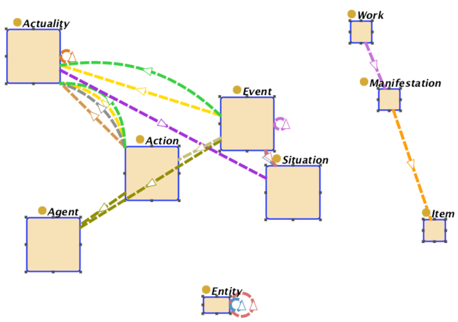

Figure 2. A figure of ABC

The ABC metamodel was developed within the Harmony international digital library project. It is intended as a basic model and ontology that provides the notional basis for developing domain specific ontologies. ABC was initially applied to metadata descriptions for complex objects of museums and libraries. A core intent in ABC is the ability to model the creation, evolution, and transitions of objects. Inspired from the theory of Petri Nets a process or evolution is modeled as a network of events and situations where transition links only exist between events and situations. These abstract concepts provide contexts for more concrete objects: Actions performed by agents occur in the contexts of events and manifestations of works exist in situations. Fig. 2 depicts that portion of the ABC ontology.

The need for meta-reasoning has been noted many times, and there are several approaches on how to support reasoning and querying on the meta-level being a recent proposal [17]. Similar to our approach, the authors interpret a domain ontology O independently from its meta-ontology M, which they call meta-view (and assume to be an OWL ontology as well). However, to facilitate maintenance, they store the meta-view within O itself as an annotation, and define an operator $\mu(O) = M$ that can extract the meta-view M from O on demand. Apart from this rather technical detail, the main difference with our approach is a shift in focus of what the meta-reasoning is about: in our approach, metadata is data about ontologies as a whole, and the meta-reasoning is intended to filter out ontologies from a large family of ontologies that fit, for instance, a particular purpose or task. In [17], the meta-view contains a reification of the axioms and annotations of a (fixed) ontology O, and meta-reasoning is intended to evaluate and select pieces of information, e.g. axioms, from within the ontology O.

3. The Need for Ontologized Metadata

We agree with both, the OMV and the ABC, as well as those metadata models proposed by the investigated repositories in that the syntactical structure of the ontology is a valuable, and viable, metadata source. We have closely followed the steps proposed by Garcia et al. [4], namely (i) identification of competency questions, (ii) reusable ontologies, (iii) domain analysis and knowledge acquisition, iv) iterative building of ontology models, (v) formalization, (vi) evaluation. We started with the definition of competency questions [4, 9, 1] for which time and temporal changes were an issue. For instance, (i) for which versions of ontology X has developer A contributed? (ii) Which action caused existence of two variants (e.g. RDF/OWL) of the same version of ontology X? (iii) Has that ontology been reused for the development of a new ontology? (iv) Who was the developer who has taken ontology X as basis for the new ontology Y? As ours is a proposed model, we evaluated it against the competency questions.

We used these questions also as a framework to conduct an informal evaluation that could produce information as to where the OMV model was not sufficient; scenarios for which modularization was essential were then laid down, one significant feature for our scenario was the specialization of metadata; for instance, the need for additional categories for describing assistive technologies. We continued with the identification of reusable upper-level ontologies that could be used to better modularize our domain and also that could make it easier for us to model temporal aspects. We analyzed DOLCE [3], ABC, and CIDOC-CRM [11]. It was decided to use ABC because it focuses on ontologizing metadata by proposing a core ontology for it. This has several benefits; firstly, it allows for untangling the metadata, acquiring thus a more modular structure. For ABC temporal aspects are central to their theory; more specifically, modeling the changes over time for digital media. Furthermore ABC provides a simple and easy-to-adapt model that is closer to our domain digital media for ABC, ontologies for us. DOLCE was too general for our purposes, i.e. ontologizing metadata. Although CIDOC-CRM provides special features for modeling time and temporal events, the nature of change is intriguingly different. The motivation of the ABC model is to represent how objects change over time, while CIDOC-CRM focuses more on changes in context and ascription than the transformation of the object [8].

The domain analysis and knowledge acquisition was done in parallel with the re-definition of the OMV model. After analyzing both models in detail with regard to merging and restructuring strategies, we identified and mapped those ABC classes that are most significant and relevant. Mappings were being constantly validated against the relevant scenarios. Fig. 2 illustrates a fragment of the ABC ontology and how it models time and temporal changes. ABC takes from the Functional Requirements for Bibliographic Records (FRBR) standardization effort [10] the definition: “Work is an abstract notion of an artistic or intellectual creation”. A Work as an abstract notion is not tangible, but it always has a realization (hasRealization) which is called Manifestation in ABC. Assertions on Manifestations depend on the context (inContext) of a Situation. Situations can change only when an Event happens; that is to say, a situation follows an event and an event precedes a situation which essentially allows for modeling temporal aspects of change. Therefore events are always annotated with timestamps. An event not only changes situations, but also may produce (hasResult) new manifestations. Moreover, events are composed (hasAction) of Actions, which are performed in the presence (hasPresence) of Agents.

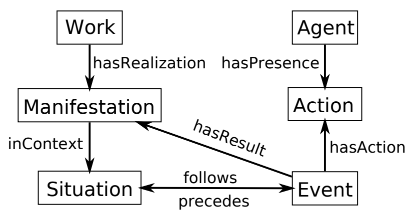

Figure 3. A portion of the ABC model

The ABC model allows to answer questions such as: how many variants of the file XY (=manifestation) exist after (=event) the maintainer A (=agent) migrated (=action) to the new format that was adopted as a new standard by the market (=situation)? The type of manifestation we have in mind within the context of OMV is an ontology as a tangible object (e.g. OWL file). In a real world scenario an ontology developer, similarly to a software engineer, creates, modifies, and/or deletes ontology files. This is an important process known as versioning; ABC models it, and allows for further refinements. However, ABC is not tailored for such kind of representation; we could not find a practical advantage in having Situation as a concept on its own within the context of combining OMV with ABC. It is sufficient to model changes of manifestations instead of modeling changes of situations where manifestations are present; we are dealing with ontologies as tangible objects, not with high-level abstraction - such as situation from ABC. Similarly, we consider Event as an unnecessary abstract concept because the actions directly change manifestations. Fig. 3 illustrates a tailored version of the ABC fragment where we also change the rather passive property name hasPresence to the active name performs. The tailored ABC model can be summarized as: A Manifestation is the realization (hasRealization) of a Work that is subjected (hasResult) to Actions performed by an Agent. We simplified this model so it could better represent temporal changes in ontologies. Fig. 3 also illustrates the simplified model.

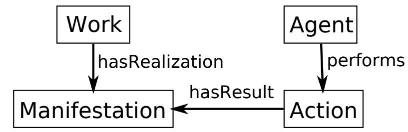

Figure 4. An initial step towards ABCing OMV

We were investigating fragments of ABC that could be used to model these temporal aspects. A tailored version of that particular fragment was consequently proposed. None of the OMV classes/properties represent temporal aspects related to the class Ontology. OMV models temporal aspects as data type properties; namely, version, modificationDate and creationDate. Intended instances of this class are concrete objects like OWL-files hosted on a specific server with a specific version number. However, the OMV model does not allow representing which of these files belong to a common history of changes, i.e. an abstract notion of ontology that is the umbrella for all its manifestations in concrete OWL files. Moreover, an OMV Party which corresponds to an Agent in ABC is “affiliated” to the history of those files. To make best use of both ontologies the ABC and the OMV we suggest to merge them as depicted in Fig. 5.

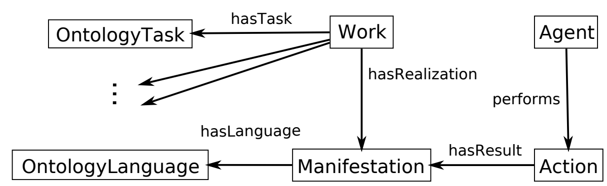

Figure 5. The ABCed model

The major change we introduced to the original OMV model was the refactorization of the class ontology into two classes: Work and Manifestation. The intended instances of Ontology in OMV are in fact all manifestations. Apart from that, we used from ABC the class Action which connect manifestations. First and foremost this allows us to model the evolution of ontologies. The refactorization also makes it possible to have a clearer definition for the object properties that are being used by the class Ontology in OMV; for instance, a concrete file (=manifestation) is written in a certain OntologyLanguage. However, a class such as OntologyTask may refer to the work as a whole, rather than to one of its manifestations. OMV resembles an extensive checklist; some of the classes proposed by OMV may not be necessary for the purpose of describing an ontology. For instance, it remains an open question if classes such as OntologyEngineeringMethodology are necessary. Very few methodologies are explicit in the accurate description of those methods and techniques they suggest. Furthermore, the application of a methodology usually involves the redefinition of some of the original steps/methods/techniques proposed, how could the methodology be reported in a way that the description could enrich the metadata being used for the ontology repository? Repositories such as sourceforge.com have defined their own set of metadata and their query facilities make use of these metadata in order to facilitate more accurate searches. These software repositories, however, are different from OORs, as they provide functionalities mainly for storing projects. They are less aiming at programmatic access via APIs or dependencies on the repository software, which is necessary for consuming ontologies from repositories.

4. Use Case

To give an illustrative example we want to describe the evolution of the transportation ontology by means of the ABC-ontology as illustrated in Fig. 6. We will use object identifiers as in: i) from ABC: WK for Work, MN for Manifestation, AC for Action, and AG for Agent, ii) from OMV: OL for OntologyLanguage, OD for OntologyDomain, and iii) from an OASIS module: OUG for OASIS User Group. A developer (AG1), whose interest is accessible transportation (OD1), wants to develop an ontology (WK2) accordingly. Initially, AG1 re-uses (AC1) the latest version (MN1) of a transportation ontology (WK1) from the repository. He writes the first draft (MN2) of the redefined transportation ontology in RDF (OL1). After submitting MN2 to the repository, another developer (AG2) is engaged in the work (WK2), who adds (AC2) information about OASIS user groups (see below). Therefore, MN2 needs to be written in OWL 1.1 (OL2) and results in a new version (MN3). After reviewing MN2, AG1 and AG2 decide to split the ontology into two versions (MN4 and MN5). MN4 specifies accessible transportation for blind users (OUG1). This ontology (WK3), for instance, specifies blind-specific services at airports and train stations. MN5 specifies accessible transportation for wheelchair users (OUG2). This ontology (WK4), for instance, specifies wheelchair-specific services, such as elevator facilities. This example also shows how we used the competency questions presented in Section 3.

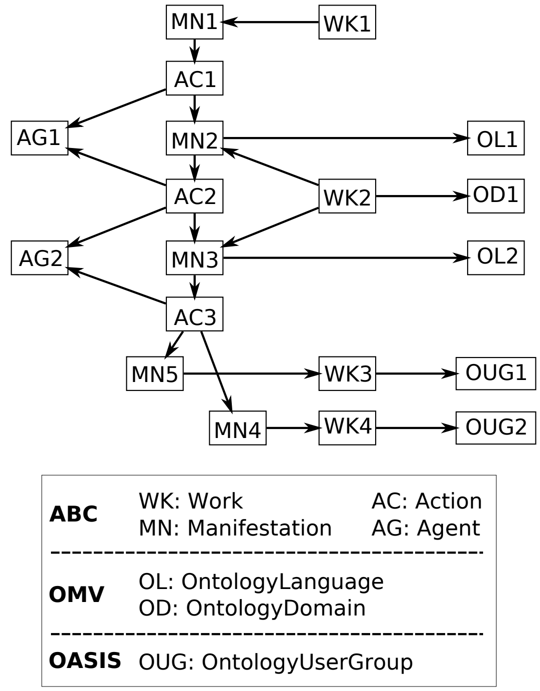

Figure 6. Case Study

5. Our Specific Scenario

For application-specific purposes, the OMV model should be extended with additional subcategories, i.e. vertical refinement. In the OASIS project, aspects for supporting independent living, health monitoring, and social relationships are specified by ontologies. The OASIS repository needs to be enriched with metadata about the OasisUserGroup, Device, Health, Transportation, and others so that queries can be processed, for instance, (a) Are there versions of the transportation ontology addressing issues related to impaired users?, (b) For which OASIS devices and/or user groups were these ontologies designed?, or (c) Which transportation ontologies can be used with mobile devices for way-finding (as in route planning)? The first competency question requires information about OntologyDomain (transportation and mobile devices) and OntologyTask (way-finding). For this case it would be sufficient to define instances (e.g. Transportation and MobileDevice) for the original OMV classes OntologyDomain and OntologyTask. However, the second question illustrates that queries related to specific types of OntologyDomain depend on a structured representation of the domain. Hence, for the OASIS repository of ontologies for assistive technologies, refinements of the OMV categories OntologyDomain and OntologyTask have to be implemented.

As seen in the example above, specific user groups (OasisUserGroup) are defined by the metadata. This category defines the user group for which the ontology is intended to be used. This category of OasisUserGroup is specialized into several subtypes, such as different elderly, impaired, or stakeholder classes. Also, in the example above, the transportation ontology has different versions (MN4, MN5) for either wheelchair users (OUG2) or blind users (OUG1). Furthermore, Device is specified as a subclass of OntologyDomain. In the OASIS project, assistive technologies are employed on different devices, such as mobile phones, handhelds, PDAs, wearable clothing, stand-alone computers, or integrated devices in stationary systems. Ontologies that support short-range trip planning, for instance, for visiting friends, the art gallery, or a hospital department, will provide applications with different modules for maps, route instructions, navigation or way-finding depending on the kind of device. Ontologies are therefore selected specifically by their device, based on information from the metadata. Given this refinement of OasisUserGroup and Device for OntologyDomain, the example (b) above can be supported. Similarly, not only Device and OasisUserGroup but also other specializations for OntologyDomain are needed. For instance, Environment, SocialRelationship, Tourism and Health. The granularity of the refinement as well as the required orchestration of the resulting metadata will be the result of future experiences and work on the OASIS repository. The competency question (c) illustrates that both (i) versioning information provided by the ABC extension of OMV together with (ii) the OASIS extension of OMV are necessary.

6. Outlook

The query capabilities of the investigated repositories remain limited and are restricted mostly to the set of metadata being used; this constrains the queries mostly to syntactical features presented in the ontology. The programmatic access to the repository provided by Swoogle facilitates the construction of more complex queries that involve, for instance, a subset of the available ontologies in Swoogle. However, the lack of a standard programmatic access to other repositories makes it difficult to do cross-queries examinations. For instance, simple queries such as which is the most commonly used object property for which there is this X domain and this Y range defined within the domain of transport ontologies across K, L, and M repositories are not possible. Query functionality can thus be improved by having better structured, richer and more extensible metadata. As we envision that several repositories of ontologies will be established, one important aspect that needs to be preserved is interoperability. This can be facilitated by having a core metadata that facilitates the development of specialized descriptors while maintain the coherence of the core; enabling thus interoperability. This will also facilitate efficient reasoning and maintenance of the entire ecosystem of metadata. In our work, we have analyzed which modules should refine or specialize OMV in order to support OASIS-specific requirements, i.e. a repository of ontologies for assistive systems. By refining OMV with relevant modules addressing particular issues of the related repository as described above, it should be easy to adapt different modules for repositories storing other kinds of ontologies.

References

[1] E. Bontas, C. Tempich, and Y. Sure. Ontocom: A cost estimation model for ontology engineering. In *ISWC'06*, 2006.

[2] P. Buitelaar, T. Eigner, and T. Declerck. Ontoselect: A dynamic ontology library with support for ontology selection. In *Demo Session ISWC'04*, Hiroshima, Japan, 2004.

[3] C. Masolo et al. Ontologies library. Technical report, WonderWeb Deliverable 2003, ISTC-CNR: Padova, Italy, 2003.

[4] C.A. Garcia et al. The use of concept maps during knowledge elicitation in ontology development processes - the nutrigenomics use case. *BMC Bioinformatics*, 7:267, 2006.

[5] D. Calvanese. TONES - Thinking ONtologiES (Project overview). In *Tones Industrial Advisory Team Workshop*, Edinburgh, UK, 2006.

[6] L. Ding. Swoogle: A search and metadata engine for the semantic web. In *13th ACM Conf. on Information and Knowledge Management*, 2004.

[7] D.L. Rubin et al. Bioportal: A web portal to biomedical ontologies. In *Symbiotic Relationships between Semantic Web and Knowledge Engineering (AAAI Spring Symposium Series)*. Stanford Univ. Press, 2008.

[8] M. Doerr, J. Hunter, and C. Lagoze. Towards a core ontology for information integration. *Journal of Digital Information*, 4(1), 2003.

[9] A. Gómez-Pérez, M. Fernández-López, and O. Corcho. *Ontological Engineering*. Springer-Verlag, London, 2004.

[10] IFLAI. Functional requirements for bibliographic records. `www.ifla.org/VII/s13/frbr/frbr.pdf`, 2008. [cited July, 2008].

[11] ISO Editor. Information and documentation – a reference ontology for the interchange of cultural heritage information. Technical report, ISO, 2006.

[12] OASIS. Open architecture for accessible services integration and standardization. Technical report, `http://www.oasis-project.eu`, 2008. [cited 2008 June].

[13] OntologySummit. Ontology summit 2008 communiqué: Towards an open ontology repository 2008. Technical report, `http://ontolog.cim3.net/cgi-bin/wiki.pl?OntologySummit2008_Communique`. [cited 2008 June].

[14] R. Palma, J. Hartmann, and P. Haase. Ontology metadata vocabulary for the semantic web. Technical report, OMV Consortium, 2008.

[15] J. Pan, S. Cranefield, and D. Carter. A lightweight ontology repository. In *2nd Int. Joint Conf. on Autonomous Agents and Multiagent Systems*, 2003.

[16] Swoogle.umbc.edu. Swoogle's statistics of the semantic web. [cited 2008 June], 2008.

[17] T. Tran, P. Haase, B. Motik, B. C. Grau, and I. Horrocks. Metalevel information in ontology-based applications. In *AAAI-08*, Chicago, 2008.

[18] D. Wang, B. Parisa, and J. Hendler. A survey of the web ontology landscape. In *ISWC'06*, Athens, GA, USA, 2006.