

MAKING HETEROGENEOUS ONTOLOGIES INTER- OPERABLE THROUGH STANDARDISATION

A META ONTOLOGY LANGUAGE TO BE STANDARDISED: ONTO- LOGY INTEGRATION AND INTEROPERABILITY (OntoIOp)

Christoph Lange^{1,2}, Till Mossakowski^{1,3,4}, Christian Galinski^{5*} and Oliver Kutz^{1,3}

1. University of Bremen, Germany

2. Computer Science, Jacobs University Bremen, Germany

3. SFB/TR 8 “Spatial cognition”, University of Bremen, Germany

4. DFKI GmbH, Bremen, Germany

5. International Information Centre for Terminology (Infoterm)

Gymnasiumstrasse 50, 1190 Vienna, Austria, +43-1-4277 58026, cgalinski@infoterm.org

ABSTRACT

Assistive technology, especially for persons with disabilities, increasingly relies on electronic communication among users, between users and their devices, and among these devices. Making such ICT accessible and inclusive often requires remedial programming, which tends to be costly or even impossible. We, therefore, aim at more interoperable devices, services accessing these devices, and content delivered by these services, at the levels of 1. data and metadata, 2. datamodels and data modelling methods and 3. metamodels as well as a meta ontology language.

Even though ontologies are widely being used to enable content interoperability, there is currently no unified framework for ontology interoperability itself. This paper outlines the design considerations underlying OntoIOp (Ontology Integration and Interoperability), a new standardisation activity in ISO/TC 37/SC 3 to become an international standard, which aims at filling this gap.

KEYWORDS

International Standards, Content Interoperability, Content Repositories, Structured Content, Standardised Data Models, Standardised Content, Accessibility, Ontologies, Ontology Languages

1 INTRODUCTION

The *Recommendation on software and content development principles 2010* formulated at the 12th International Conference on Computers Helping People with Special Needs (ICCHP 2010) addresses the important aspect of interoperable content from an eAccessibility and eInclusion perspective. Content interoperability enhances the possibilities to search, retrieve, and recombine content from different points of view. Content interoperability enhances the personalisability of ICT devices, which is especially important for persons with disabilities (PwD). Content interoperability can only be achieved on the basis of – preferably international – standards. However, although a number of pertinent standards exist, some of these need revision, and additional standards are needed for new requirements.

Content interoperability means the capability of content entities to be combined with or embedded in other (types of) content items on the one hand, and to carry enough *context* information for being extensively re-used (i.e. *re-usability* extended towards *re-purposability*). Different items of structured content are increasingly combined with or embedded in each other. The fact that structured content must increasingly be multilingual and also multimodal only adds to the complexity. Therefore, a systemic and generic approach to data modeling and content management has become the order of the day.

So far the standardisation efforts concerning structured content focused on

- Data categories and metadata used in the conceptual design of individual entries of structured content,
- Data models and data modeling methods,
- Metamodels to make competing data models interoperable.

Some kinds of structured content, such as coding systems (e.g. for names of countries, currencies, languages or safety symbols) are so important that the content items themselves are internationally standardised. If *ontologies*, in the meaning of knowledge representation, are included, the above-mentioned metamodels need to be extended towards meta ontologies and ontology languages, in order to provide the possibility to make ontologies interoperable.

An ontology is a “formal, explicit specification of a shared conceptualisation” [1], which represents a shared vocabulary and taxonomy that models a domain — that is, the definition of concepts and other information objects, as well as their properties and relations. An ontology language – different from mere knowledge representation – provides a metamodel for such formal, explicit specifications of a shared conceptualisation.

Particularly in the field of eAccessibility&eInclusion, the use and re-use of all kinds of content across different technical platforms is a must. On the other hand, today, strongly heterogeneous content is still more the rule than the exception. While in the past the development focus was on tools (i.e. devices, computer hardware and software), it is increasingly recognised today that “communication” ultimately is the most important issue, namely

- communication with and among persons with disabilities (PwD) directly or through ICT devices,
- communication between PwD and the devices they use, and
- communication among these devices.

Needless to say that the above-mentioned approach has a big impact on software development, as can be gathered from the documents MoU/MG/05 N0221¹, [2] and MoU/MG/05 N0222 [3]. Modularity and comprehensive interoperability, capability for multilinguality and multimodality based on open standards are increasingly required. Therefore, metadata, data models, messages, protocols, conversion of all sorts, multilinguality (incl. cultural diversity), multimodality, design for all (DfA) etc. have become the objective of standardisation efforts in industry, first of all in eBusiness, by specialised organisations or in public institutions. [4]

Since a number of years, the number of standards for services and structured content is growing exponentially. PwD are among those who will benefit most from this development. However, in reality, as experienced in the OASIS project mentioned in Section 2, and elsewhere, probably most of the existing repositories of structured data are not consistent even within a given repository, and contradictory between different repositories. Most of them are not based on proper metadata and data modelling methods, and therefore not integratable, not reliable, and full of deficiencies. This is unacceptable especially in applications that support PwD, particularly in our aging societies, where more and more people suffer from multiple impairments. Implementing best practices in existing content repositories, such as in eBusiness², and improving data quality and reliability by higher standards for quality assessment and certification, are necessary steps – but not sufficient. The current reality shows that assistive services are, and will most likely always be, relying on a multitude of heterogeneous content repositories with their underlying models and metamodels. The proposed International Standard “OntoIOP” (cf. Section 3) acknowledges this diversity and aims at bridging such different models and metamodels.

2 IMPLEMENTATION CASE: THE OASIS EU PROJECT

OASIS (Open architecture for Accessible Services Integration and Standardisation) is an ongoing European large-scale Integrated Project within the 7th Framework Programme. It started in January 2008 for a period of four years. OASIS aims at an innovative reference architecture, based upon ontologies and semantic services, which will allow plug and play and cost-effective interconnection of existing and new services in all domains required for the independent and autonomous living of the elderly and enhancing their Quality of Life (for this and the following see [5]). Despite including Ambient Assisted Living (AAL) components, the focus on services means that the OASIS approach is generic and thus applicable to eApplications beyond the functional and technical requirements of AAL.

OASIS will provide a complementary solution to the direction now being pursued in the Semantic Web approach to ontology design: whereas in the Semantic Web re-usability is pursued across open ontologies, OASIS wants to achieve re-usability within strict modularity, by means of a Common Ontological Framework (COF). COF follows the above-mentioned

¹ MoU/MG: Management Group of the ITU-ISO-IEC-UN/ECE Memorandum of Understanding on standardisation in electronic business

²For eBusiness content repositories, we refer to the standards developed by ISO/TC 184/SC 4.

MoU/MG recommendation, which states as basic requirements for the development of fundamental methodology standards concerning content interoperability the fitness for

1. multilinguality (also covering cultural diversity),
2. multimodality and multimedia,
3. eAccessibility (also covering eInclusion), and
4. multi-channel presentations,

which have to be considered at the earliest stage of software design and then throughout all software development cycles. [2] As side effects in software design are a source of major system instability, development and maintenance costs, avoiding them for ontology design will be a major innovative contribution of considerable benefit.

3 OntoIOp (ISO/TC 37/SC 3 WI 17347)

Although in current ontology standardisation initiatives much has already been taken from established software engineering practice, surprisingly little of what is known about modularity and structuring has so far been applied. Part of this gap can be traced back to differing starting points: adopting a description logic basis, as typical within Semantic Web oriented information modelling, has restricted the development of more powerful and generic approaches to supporting modularity. This is largely due to the fact that existing metamodel specifications and ontology definition standards assume that the ontologies produced are essentially compatible down to the exchange of terms and filling in respective knowledge gaps. But this ‘assumption’ of ontological compatibility frequently fails to hold. It also does not match current practice nor expectations when standardisation is considered across technical communities.

In order to fill the current gaps in modular ontology design and to augment ongoing standardisation efforts with an essential layer of standardised modularity and structuring guidelines, a new working item (WI 17347) on “Ontology Integration and Interoperability (OntoIOp)” has been initiated in ISO/TC 37/SC 3 “Systems to manage terminology, knowledge and content”. The initiative brings together new results in the international state of the art in ontology-based interoperability. This includes results from several large-scale initiatives. Thus the proposed International Standard OntoIOp aims at bridging the above-mentioned gaps in standards and guidelines.

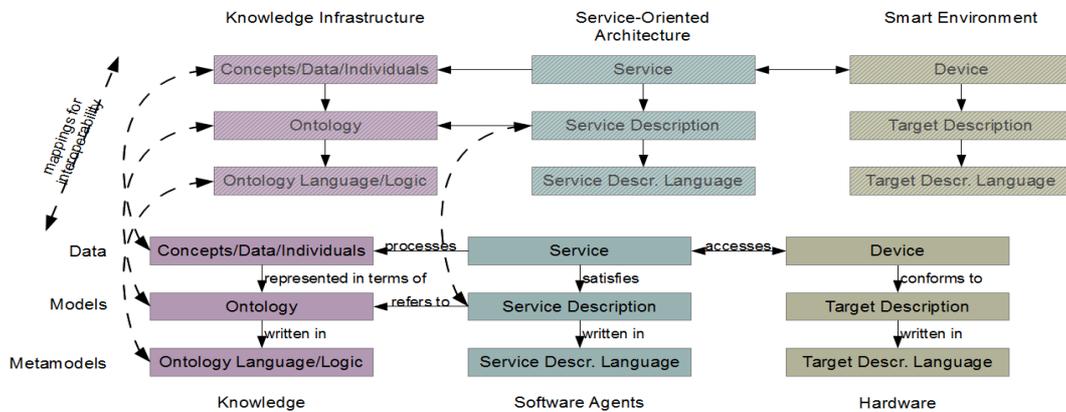


Fig. 1: Interoperability of Knowledge, Service, and Hardware

The general picture of interoperability that we propose is shown in Fig. 1. Read from left to right, the story of the picture is as follows: ontology-based knowledge infrastructure provides the conceptual background for service-oriented architectures, and software services in turn interact with hardware targets (devices) in smart environments (such as ambient assisted living homes). Read from top to bottom, the story is that individual concepts belong to ontologies, which in turn are formulated in some ontology language, such as OWL or Common Logic. Likewise, services adhere to service descriptions, which are in turn formulated in a service description language, such as WSDL or DIANE. Finally, hardware devices (called targets) conform to target descriptions, and e.g. the Universal Remote Console consortium has formulated an XML format for such target descriptions [6].

The important insight is that interoperability – the third dimension in this diagram – needs to address all these levels. Concerning the vertical line, let us stay at the left column (knowledge infrastructure) for a moment. Interoperability of data means that data is converted from one format into another one as required (e.g. JPEG to PNG). At the ontology level, there is the need to relate different ontologies using different terminologies and concepts. Finally, ontologies may have been formulated in different ontology languages, and these languages need to be related in order to gain interoperability. Indeed, interoperability between two knowledge representations may involve mappings at all three levels simultaneously: data, ontologies, and ontology languages. The mappings arising at these levels are represented by the dashed lines. For ontology-based web services and for devices, the situation is similar, but mappings have not yet been studied to the same extent as for ontologies.

Concerning the horizontal line, interoperability of services and service composition can often be enhanced by grounding the service descriptions in some ontology (or different but interop-

erated) ontologies. Interoperability of targets is often ensured by encapsulating as software services and then using service composition techniques.

The OntoIOp standardisation effort has been started with this big picture of interoperability in mind, but soon forces have been concentrated to the ontology part of the picture – this bears enough complexity in its own. As indicated by the structural similarities, we hope that the lessons learnt there will be useful for the service and device level as well.

3.1 Motivation of OntoIOp

We now provide details for the ontology part of the above diagram but refer to [7] for a more in-depth elaboration. As a matter of fact, we face a multitude of ontology languages. While the OWL standard [8] has led to an important unification of notation and semantics, still many diverse formalisms are used for writing ontologies. Some of these, such as UML class diagrams [9], can be considered fragments and notational variants of OWL, while others, such as Common Logic [10], clearly go beyond the expressiveness of OWL. Moreover, not only the underlying logics are different, but also the modularity and structuring constructs, and the reasoning methods.

Many ontologies, in AAL and other domains, are written in description logics such as *SROIQ(D)* (underlying OWL 2 DL) and its fragments. These logics are characterised by a rather fine-tuned expressivity, exhibiting (still) decidable satisfiability problems, whilst being amenable to highly optimised implementations. However, in many practical cases the expressivity has to be extended beyond the scope of standard DLs – to first-order or even second-order logic. As an example for the former, consider simple devices such as light switches, which only need a simple domain model for their own operation, whereas the model required for integrating many such devices in a complex AAL environment can be provided by grounding the simple individual ontologies in a foundational ontology, such as DOLCE [11], with its abstract, generic notions of space, time, parthood, etc.

As OWL mainly targets applications in the Semantic Web and related areas, it cannot be expected to be fit for any purpose: there will always be new, typically interdisciplinary application areas for ontologies where the employed (or required) formal languages do not directly fit into the OWL landscape. Heterogeneity (of ontology languages) is thus clearly an important issue. This does not only include cases where the expressivity of OWL is simply exceeded (such as when moving to full first-order logic), but, ultimately, also cases where combinations with or connections to formalism with different semantics have to be covered, such as temporal or spatial logics; cf. e.g. [12].

Ontology interoperability is closely related to ontology matching and alignment [13]. Here, we will not consider the process of *finding* suitable matchings and alignments, but rather aim at a language for *expressing* them. Ontology interoperability is also related to ontology modularity and structuring; see [7] for further details.

3.2 Requirements for a Future Standard

We here propose a solution to the above issues based on the concept of heterogeneity: facing the fact that several logics and formalisms are used for designing ontologies, we suggest heterogeneous structuring constructs that allow to combine ontologies in various ways, in a systematic and formally and semantically well-founded fashion.

The proposed standard will specify a *distributed ontology language* (DOL), which will serve as a language for distributed knowledge representation in ontologies and interoperability among ontologies, and in the long run also among services and devices.

It is important to stress that we do not aim at “yet another ontology language”. Rather, DOL will be a meta-language for *integrating* existing ontology languages, such that one can continue to use existing tools. DOL shall allow for including all practically relevant languages – OWL together with its sublanguages (called profiles), the UML class diagrams, Common Logic (and its dialects), and others [7] – but also possible future languages. This is achieved via *conformance criteria* inspired by Common Logic [10]. We are currently specifying criteria for how each of the following can conform with DOL: a logic underlying an ontology language, a concrete serialization of an ontology language, a document (which may contain fragments in different ontology languages), and an application that processes such documents.

DOL will allow the user to integrate and compare ontologies written in different languages. Mappings between ontologies, as they occur in ontology matching and alignment [13], will be first-class citizens. Moreover, the standard will also accommodate for

- translations between ontology languages, as objects expressible in the language;
- heterogeneous ontologies combining parts written in different languages; and
- ontologies distributed over different internet locations.

DOL will be defined with a formal syntax and criteria for further serialisations to conform. In particular, we shall target machine processing by specifying conforming XML and RDF serialisations called Distributed Ontology Interchange format (DIF). Existing ontologies in existing XML- or text-based serialisations shall be made valid DOL ontologies with a minimum amount of syntactic adaptation.

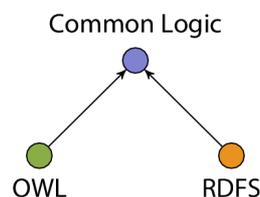


Fig. 2: Mapping two ontology languages into a third one

For giving DOL a well-defined formal semantics, we will define and standardise a core graph of appropriate ontology language translations (see below), which is open for extension by translations to further conforming languages. Every pair of ontology languages should have ontology translations into a common target language (cf. Fig. 2).

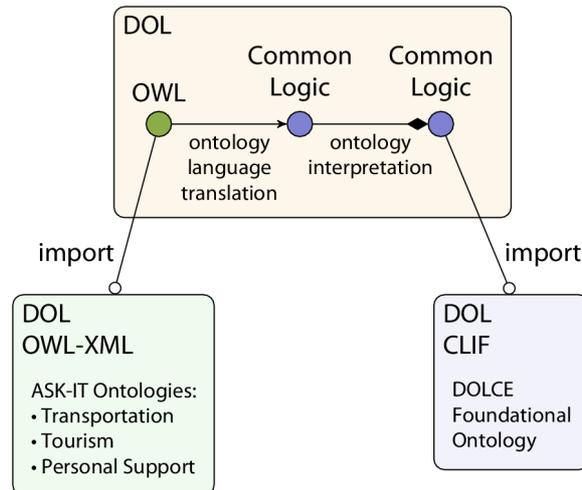


Fig. 3: Mapping between ontologies formulated in different ontology languages

3.3 A Very Brief Sketch of the Hyperontology Framework

The definition of the semantics of the distributed ontology language DOL leads to technical challenges, since all the individual ontology languages come with their own syntax and semantics, which need to be integrated properly. Addressing this problem of logical translation between ontology languages, [14] has presented the theoretical foundations for the distributed ontology language DOL on the ‘translation layer’. The general idea here is that such translations will allow users to use their own preferred ontology formalism, whilst being interoperable with other formalisms.

At the heart of our approach, therefore, is a graph of ontology languages and translations, enabling users to:

1. relate ontologies that are written in different formalisms;
2. re-use ontology modules even if they have been formulated in different formalisms;
3. re-use ontology tools like theorem provers and module extractors along translations.

A detailed discussion of the various translational relationships between (almost) all known ontology languages can be found in [14]; Fig. 4 illustrates the translational relationships. Leaving out technicalities (but see [14]), a ‘regular’ translation between two ontology languages means that the syntax and semantics of one logic can be translated into another. This means that, typically, the former is a fragment of the latter. A standard example would be OWL which, via the standard translation, can be considered a fragment of first-order logic. The languages shown in Fig. 4 range from sub-Boolean (OWL fragments) to sub-second-order (Common Logic, but extending full first-order by some second-order constructs).

Notice that there are often several translation between two distinct languages, namely one given by an immediate arrow (the ‘natural’ translation), and one given by composing several other translations. The reason for giving translations that could alternatively also be obtained via composition of already existing translations is that the ‘natural translations’ typically have better properties in terms of comprehensibility and succinctness. The choice of a translation therefore impacts in particular computational properties and tool re-usability.

3.4 Embedded Documentation

Accessibility is not just enhanced by offering novel services that assist end users by *performing* actions on their behalf, but also by *explaining* the functionality of existing services. To that end, we require ontologies to not just enable knowledge representation in an explicit, machine-comprehensible way, but also to provide human-comprehensible annotations of its concepts. Such embedded documentation obviously supports knowledge engineers and service developers in using an ontology, but services can also expose the documentation to their end users at runtime.

State-of-the-art ontology languages hardly address documentation. Common Logic [10], for example, merely allows for attaching unstructured comments to its phrases (modules, sentences, imports, or text). OWL [8], thanks to its roots in the RDF metadata model, allows for supplying additional types of comments (*annotation properties*) and supports complex multilingual annotations, but there are fewer subjects within OWL ontologies to which such annotations can be attached. More complex kinds of annotations, e.g. of subsets of an ontology⁵, are not currently supported by any language – let alone the software engineering practice of “literate programming”, where natural language documentation and formal expressions are freely interwoven and reference each other in a fine-grained way, enabling the generation of both a reference manual and compilable/executable code (here: a formal ontology) from the same source.

Building on initial results obtained with a proof-of-concept ontology language with rich documentation capabilities [16], DOL intends to enable ontology engineers to document arbitrary fragments of existing ontologies without having to translate them. To that end, DOL will extend the annotation facilities of existing ontology languages with a vocabulary for non-

⁵unless such subsets have been modeled explicitly as ontology modules

intrusive *standoff markup*, pointing to the exact subjects of annotation from external documentation files or from special internal comments. The DOL annotation sublanguage will be built on established Semantic Web standards, including RDF for representing annotations in general, RDFa [17] for fine-grained embedding of annotations into XML serialisations of ontologies, and XPointer URIs [18] as a non-destructive means of referencing text or XML pieces, where the original ontology language does not provide built-in identification mechanisms.

Finally, we will compile a list of vocabularies recommended for annotating ontologies, comprising generic metadata vocabularies such as Dublin Core [19] as well as ontology-specific vocabularies such as the Ontology Metadata Vocabulary (OMV [20]).

4 SUMMARY AND OUTLOOK

Content integration – whether in the form of virtual or real data integration – and content interoperability must be based among others on

1. consistent methodology standards for data models and data modelling,
2. coordinated standardisation of several kinds of structured content,
3. standardised identification systems for individual pieces of information,
4. standardised transfer protocols and interchange formats, in order to be efficient and reliable,
5. standardised metamodels and a standardised meta ontology language.

Persons with disabilities are among those who will benefit most from these standards. They will not only be able to use information across technical platforms, but also gain access to information and knowledge in general, which would also improve their access to education.

There are several technical committees sharing standardisation work on the above-mentioned aspects. OASIS attempts to combine the results of a number of past projects with best practices concerning structured content development and maintenance as well as with respect to content integration. We have sketched a new “distributed ontology language”, DOL, enabling integration of a number of existing and future ontology languages, as well as ontology mappings and language mappings, leading to sustainable interoperability among ontologies. A formal, machine-comprehensible semantics is provided by the Common Ontological Framework and its sophisticated heterogeneous structuring mechanisms, whereas an annotation and documentation framework improves comprehensibility for human users. Future work will have a closer look at the interoperability of ontologies, services and devices as well.

If there are standards, standards-based certification is possible. Especially with respect to eAccessibility&eInclusion there is a definite need for certification, validation or verification of data, which possibly can largely be done through web services. These standards and certification schemes would

1. first of all benefit PwD,
2. benefit also small content and service providers,

3. be affordable and
4. fit the kind of service, the technical state-of-the-art at the service providers' side and the expectations of the clients.

Coordination and harmonisation efforts supported by the EU Commission have a positive effect on the development of technical, organisational and content interoperability standards as well as standards-based content repositories, which will benefit first of all persons with disabilities.

ACKNOWLEDGEMENTS

Work on this paper has been supported by the EU FP7-funded project “Open architecture for Accessible Services Integration and Standardisation” (OASIS), by the DFG-funded collaborative research centre SFB/TR 8 ‘Spatial Cognition’ and by the German Federal Ministry of Education and Research (Project 01 IW 07002 FormalSafe). This paper reflects the current state of the OntoIOp standardisation effort (Working Item 17347 in ISO/TC 37/SC 3), to which 13 experts appointed by 5 countries have actively contributed so far.

REFERENCES

- [1] [Gruber, Thomas R.](#) (June 1993). [A translation approach to portable ontology specifications](#) (PDF). *Knowledge Acquisition* **5** (2): 199–220. <http://tomgruber.org/writing/ontolingua-kaj-1993.pdf>.
- [2] MoU/MG N0221:2004 Semantic Interoperability and the need for a coherent policy for a framework of distributed, possibly federated repositories for all kinds of content items on a world-wide scale (adopted in 2005) http://isotc.iso.org/livelink/livelink/fetch/2000/2489/Ittf_Home/MoU-MG/Moumg221.pdf
- [3] MoU/MG N0222:2004 Statement on eBusiness Standards and Cultural Diversity (adopted in 2006) http://isotc.iso.org/livelink/livelink/fetch/2000/2489/Ittf_Home/MoU-MG/Moumg222.pdf
- [4] ICTSB (ed.). Design for All. ICTSB Project Team Final Report (2000) http://www.ictsb.org/Activities/Design_for_All/Documents/ICTSB%20Main%20Report%20.pdf
- [5] Bekiaris, E. D.; Bateman, J.A. Content Interoperability Standardisation. The open Reference Architecture of the OASIS Project with its open source Common Ontological Framework and related tools. China Standardisation, Special edition 2009. Beijing 2009, pp. 41–44
- [6] ISO/IEC 24752-4:2008 Information technology – User interfaces – Universal remote console – Part 4: Target description
- [7] Oliver Kutz, Till Mossakowski, Christian Galinski, and Christoph Lange. Towards a Standard for Heterogeneous Ontology Integration and Interoperability. International Conference on Terminology, Language and Content Resources (LaRC). 2011.

- [8] OWL 2 Web Ontology Language: Document Overview. Recommendation, World Wide Web Consortium (W3C), 2009.
- [9] Wang, X., and Chan, C. W. Ontology modeling using UML. In OOIS (2001), Y. Wang, S. Patel, and R. Johnston, Eds., Springer, p. 59.
- [10] Common Logic Working Group. Common Logic: Abstract syntax and semantics. Tech. rep., 2003.
- [11] Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., and Schneider, L. Sweetening Ontologies with DOLCE. In Proc. of EKAW 2002 (2002), vol. 2473 of LNCS, Springer, pp. 166–181.
- [12] Artale, A., Kontchakov, R., Lutz, C., Wolter, F., and Zakharyashev, M. Temporalising tractable description logics. In Proc. of the 14th Int. Symposium on Temporal Representation and Reasoning (TIME) (Washington, DC, USA, 2007), IEEE, pp. 11–22.
- [13] Euzenat, J., and Shvaiko, P. *Ontology Matching*. Springer, Heidelberg, 2007.
- [14] Mossakowski, T., and Kutz, O. The Onto-Logical Translation Graph. In Proc. of the 5th Int. Workshop on Modular Ontologies (WoMO-11) (2011), *Frontiers in Artificial Intelligence and Applications*, IOS Press.
- [15] Kutz, O., Mossakowski, T., and Lücke, D. Carnap, Goguen, and the Hyperontologies: Logical Pluralism and Heterogeneous Structuring in Ontology Design. *Logica Universalis* 4, 2 (2010), 255–333. Special Issue on ‘Is Logic Universal?’.
- [16] Lange, C., and Kohlhase, M. A mathematical approach to ontology authoring and documentation. In *MKM/Calculamus Proceedings (July 2009)*, J. Carette, L. Dixon, C. Sacerdoti Coen, and S. M. Watt, Eds., no. 5625 in LNAI, Springer Verlag, pp. 389–404.
- [17] RDFa Core 1.1. W3C Working Draft, World Wide Web Consortium (W3C).
- [18] Grosso, P., Maler, E., Marsh, J., and Walsh, N. W3C XPointer framework. W3C Recommendation, World Wide Web Consortium (W3C), Mar. 2003.
- [19] DCMI Usage Board, T. DCMI metadata terms. DCMI recommendation, Dublin Core Metadata Initiative, 2003.
- [20] Hartmann, J., Palma, R., Sure, Y., Suárez-Figueroa, M. C., Haase, P., Gómez-Pérez, A., and Studer, R. Ontology metadata vocabulary and applications. In *On the Move to Meaningful Internet Systems 2005: OTM Workshops (Oct. 2005)*, R. Meersman, Z. Tari, P. Herrero, et al., Eds., no. 3762 in LNCS, Springer, pp. 906–915.