Towards Interoperable Metadata Provenance

Kai Eckert, Magnus Pfeffer
University Library
University of Mannheim
Mannheim, Germany
Email: eckert/pfeffer@bib.uni-mannheim.de

Johanna Völker
KR&KM Research Group
University of Mannheim
Mannheim, Germany
Email: voelker@informatik.uni-mannheim.de

***Abstract*—Linked data has finally arrived. But with the availability and actual usage of linked data, data from different sources gets quickly mixed and merged. While there is a lot of fundamental work about the provenance of metadata and the commonly recognized demand for expressing provenance information, there still is no standard or at least best-practice recommendation. In this paper, we summarize our own requirements based on experiences at the Mannheim University Library for metadata provenance, examine the feasibility to implement these requirements with currently available (de-facto) standards, and propose a way to bridge the missing gaps. By this paper, we hope to obtain additional feedback, which we will feed back into ongoing discussions within the recently founded DCMI task-group on metadata provenance.**

I. Introduction

At the Mannheim University Library (MUL), we recently announced a Linked Data Service[1] (LDS). Our complete catalog with about 1.4 million is made available as RDF, with proper dereferenceable URIs and a human-readable presentation of the data as HTML pages. The title records are linked to classification systems, subject headings and to other title records. The Cologne University Library made its catalog data available under a creative commons CC-0 license, so we converted it to RDF and made it available along our own catalog.

The HTML view[2] provides browsable pages for all resources described in the RDF data. It fetches additional statements when users click on the URIs, provided that they are available by URI dereferencing. The resulting statements are presented to the user within the LDS layout and cannot be easily distinguished from the data that is made available by the Mannheim University Library itself. There is only a note about the "data space", basically indicating the domain where the dereferenced URI resides.

A good thing is that the service is totally *source-agnostic* and fetches and presents everything that is available. With two clicks, the user gets subject data from the library of congress (LoC), just because we use the German subject headings and the German National Library (Deutsche Nationalbibliothek, DNB) provides skos:match statements to LoC subject headings (LCSH).

A bad thing is that the service is totally *source-agnostic* (apart from the data-space notion). For example, the DNB states on its website that the data is provided only as a prototype, should only be used after a consultation and not for commercial applications. The LCSH data is public domain and freely available. But also within our triple store, there are different datasets. The MUL catalog is currently provided without a specific license, as questions about the proper licensing still are discussed. The data from the Cologne University Library has been processed by us and the processed data is provided by a the creative commons CC-0 license, too.

A. Motivation

Our predicament: We do want the LDS to be source-agnostic. But at the same time we want to know about the license of the data that is displayed to the user, and we want to present him with this information. Moreover, besides license and source information, we also have other information that we would like to make available to the user or other applications in a reusable way. But the current state of the art is that this information is either not made available within the RDF datasets yet – the case for DNB, LoC and our own data – or not in a consistent way. For example, the data from the OCLC service dewey.info[3] contains licensing statements as part of the RDF statements about a given resource (Ex. 1).

```
<http://dewey.info/class/641/2009/08/about.en>
 a <http://www.w3.org/2004/02/skos/core#Concept>;
 xhv:license
   <http://creativecommons.org/licenses/by-nc-nd/3.0/>;
 cc:attributionName
   "OCLC Online Computer Library Center, Inc.";
 cc:attributionURL <http://www.oclc.org/dewey/>;
 ...
 skos:prefLabel "Food & drink";
 skos:broader
   <http://dewey.info/class/64/2009/08/about.en>;
 cc:morePermissions
   <http://www.oclc.org/dewey/about/licensing/>.
```

Example 1: Provenance in dewey.info dataset

[1] http://data.bib.uni-mannheim.de
[2] currently implemented with Virtuoso RDF-Mapper
[3] http://dewey.info

As another example, the New York Times expresses provenance outside the actual data record, more precisely by means of statements about the data record (Ex. 2).

```
<http://data.nytimes.com/46234942819259373803.rdf>
  foaf:primaryTopic
    <http://data.nytimes.com/46234942819259373803>
  dcterms:rightsHolder
    "The New York Times Company"
  ...
  cc:license
    <http://creativecommons.org/licenses/by/3.0/us/>
  cc:attributionURL
    <http://data.nytimes.com/46234942819259373803>
  cc:attributionName
    "The New York Times Company"

<http://data.nytimes.com/46234942819259373803>
  ...
  a <http://www.w3.org/2004/02/skos/core#Concept>
  ...
  skos:prefLabel "Faircloth, Lauch"
  ...
```

Example 2: Provenance in New York Times dataset

Our goal is to make this kind of information available to the user in a *consistent* way. We respect all the different licenses and do not want to make users believe that all this data is provided by ourselves, without any licensing information.

Besides provenance information, we also need to provide other information that further qualifies single statements of the datasets. For example, in a past project we automatically created classifications and subject headings for bibliographic resources. We provide this data also via the LDS which is very convenient and greatly facilitates the reuse of the data. But automatically created results often lack the desired quality, moreover the processes usually provide further information, like a weight, rank or other measures of confidence [1]. All this information should also be provided to the user in a well-defined way.

B. Data, Metadata, Metametadata, ...

Data provided as RDF is not necessarily metadata in a strict sense; in general it is data about resources. But in many cases – and especially in the context of this paper – the resources are data themselves, like books, articles, websites or databases. In the library domain, the term **"metadata"** is thus established for all the data about the resources a librarian is concerned with – including, but not restricted to bibliographic resources, persons and subjects. This is the reason, why one cannot distinguish easily between data and metadata in the context of RDF. We therefore regard them as synonyms.

Metadata is itself data and there are a lot of use-cases where one wants to make further statements about metadata, just as well as metadata provides statements about data: who created the metadata, how was the metadata created, ... – in general additional statements to further qualify and describe the metadata. Thus we will refer to this kind of additional information unambiguously as **"metametadata"**.

C. Metametadata Principles

To achieve interoperability for accessing metametadata, choosing a representation of the metametadata is only the first, merely technical step. In our opinion, the following principles and requirements have to be met to achieve this type of interoperability:

1) Arbitrary metametadata statements about a set of statements.
2) Arbitrary metametadata statements about single statements.
3) Metametadata on different levels for each statement or sets of statements.
4) Applications to retrieve, maintain and republish the metametadata without data loss or corruption.
5) Data processing applications to store the metametadata about the original RDF data.

Requirements 1 - 3 address the technical requirements that have to be met by the metadata format(s) in use. They are met by RDF, but in RDF there are two distinct approaches that can be used to represent metametadata:

Reification: RDF provides a means for the formulation of statements about statements, called *reification*. In the RDF model, this means that a complete statement consisting of subject, predicate and object becomes the subject of a new statement that adds the desired information.[4]

Named Graphs: Another technique that can be used to provide statements about statements are the "Named Graphs", introduced by Carroll et al. [2]. The Named Graphs are not yet officially standardized and part of RDF. They have to be considered work in progress, but are already widely used by the community and can already be considered as a kind of de-facto standard that is likely to have a big impact on future developments in

[4] As a statement cannot be identified uniquely in RDF beside the notion of S, P and O, a reification statement refers to *all* triples with the given S, P and O. In our context, this ambiguity has no substantial effects, as identical triples are semantically equivalent to duplicated metadata that can be safely discarded as redundant information.

the RDF community.[5] Named Graphs are an extension of RDF, both on the model and syntax level. They allow the grouping of RDF statements into a graph. The graph is a resource on its own and can thus be further described by RDF statements, just like any other resource. There are extensions for SPARQL and N3 to represent and query Named Graphs, but they are for example not representable in RDF-XML.[6]

To meet requirements 4 and 5, further conventions among interoperable applications are needed that have to be negotiated on a higher level and are (currently) beyond the scope of RDF. By virtue of the following use-cases, we demonstrate that the technical requirements are already met and that we only need some conventions to represent such information in an consistent way – at least as long as the official RDF standard does not address the metametadata issue.

II. Example Use-Cases

The following use cases[7] are meant to be illustrating examples, especially to emphasize the need for the representation of arbitrary information – not only provenance – about data on various levels, from whole datasets over records to single statements or arbitrary groups of statements.

In this section, we develop a scenario where such metametadata can be used to prevent information loss while merging subject annotations from different sources. We show that this is the key to make transparent use of different annotation sources without compromises regarding the quality of your metadata. In line with our argumentation in this paper, we propose the storage of metametadata to mitigate any information loss and allow the usage of this information to achieve a better retrieval experience for the users. With various queries, we show that we can access and use the additional pieces of information to regain a specific set of annotations that fulfills our specific needs.

This scenario focuses on the merging of manually assigned subject headings with automatically assigned ones. Example 3 shows a DC metadata record with subject annotations from different sources and additional information about the assignments via RDF reification. Note that we present the triples in a table and give them numbers that are then used to reference them.

	Subject	*Predicate*	*Object*
1	ex:docbase/doc1	dc:subject	ex:thes/sub20
2	#1	ex:source	ex:sources/autoindex1
3	#1	ex:rank	0.55
4	ex:docbase/doc1	dc:subject	ex:thes/sub30
5	#4	ex:source	ex:sources/autoindex1
6	#4	ex:rank	0.8
7	ex:docbase/doc1	dc:subject	ex:thes/sub30
8	#7	ex:source	ex:sources/pfeffer
9	#7	ex:rank	1.0
10	ex:docbase/doc1	dc:subject	ex:thes/sub40
11	#10	ex:source	ex:sources/pfeffer
12	#10	ex:rank	1.0
13	ex:sources/autoindex1	ex:type	ex:types/auto
14	ex:sources/pfeffer	ex:type	ex:types/manual

Example 3: Subject assignments by different sources

There is one document (*ex:docbase/doc1*) with assigned subject headings from two different sources. For each subject assignment, we see that a source is specified via a URI. Additionally, a rank for every assignment is provided, as automatic indexers usually provide such a rank. For example, a document retrieval system can make direct use of it for the ranking of retrieval results. For manual assignments, where usual no rank is given, this could be used to distinguish between high quality subject assignments from a library and, for example, assignments from a user community via tagging.

The statements #13 and #14 are used to further qualify the source, more precisely, to indicate, if the assignments were performed manually (*ex:types/manual*) or automatically (*ex:types/auto*).

A. Use-case 1: Merging annotation sets

Usually, the statements from Example 3 are available from different sources (as indicated) and might also belong to different shells in the shell model. The integration requires to merge them in a single store. An interesting side-effect of the use of RDF and reification is that the merged data is still accessible from every application that is able to use RDF data, even if it is not possible to make reasonable use of our metametadata. This is demonstrated by the first query in Example 4, which retrieves all subject headings that are assigned to a document. As in RDF all statements are considered identical that have the same subject, predicate and object, every subject heading is returned that is assigned by at least one source. In most cases, these completely merged statements are not wanted. As promised, we show with the second query in Example 4 that we are able to regain all annotations that were assigned by a specific source (here *ex:sources/pfeffer*).

[5] See http://www.w3.org/2004/03/trix/ for a summary. There are already further extensions or generalizations of Named Graphs, like Networked Graphs [3] that allow the expression of views in RDF graphs in a declarative way. Flouris et al. propose a generalization to maintain the information associated with graphs, when different graphs are mixed [4]: Here, colors are used to identify the origin of a triple, instead of names. A notion of “Color1+Color2” is possible and the paper demonstrates, how reasoning can be used together with these colored triples. Gandon and Corby published a position paper [5] about the need for a mechanism like Named Graphs and a proper standardization as part of RDF.

[6] You can see the grouping of statements in a single RDF-XML file as the notion of an implicit graph and use the URI of the RDF-XML file to specify further statements about this graph, just like Ex. 2

[7] First published at the DC 2009 conference [6].

```
SELECT ?document ?value WHERE {
   ?t rdf:subject ?document .
 ?t rdf:predicate dc:subject .
 ?t rdf:object ?value .
}
```

document	*subject*
ex:docbase/doc1	ex:thes/sub40
ex:docbase/doc1	ex:thes/sub30
ex:docbase/doc1	ex:thes/sub20

```
SELECT ?document ?value WHERE {
   ?t rdf:subject ?document .
 ?t rdf:predicate dc:subject .
 ?t rdf:object ?value .
 ?t ex:source <ex:sources/pfeffer> .
}
```

document	*subject*	*source*
ex:docbase/doc1	ex:thes/sub40	ex:sources/pfeffer
ex:docbase/doc1	ex:thes/sub30	ex:sources/pfeffer

Example 4: Querying the merged statements

B. Use-case 2: Extended queries on the merged annotations

In the following we show two extended queries that make use of the metametadata provided in our data store. Usually, one does not simply want to separate annotation sets that have been merged, but instead wants to make further use of these merged annotations. For example, we can provide data for different retrieval needs.

The first query in Example 5 restricts the subject headings to manually assigned ones, but they still can originate from different sources. This would be useful if we are interested in a high retrieval precision and assume that the results of the automatic indexers decrease the precision too much.

The second query, on the other hand, takes automatic assignments into account, but makes use of the rank that is provided with every subject heading. This way, we can decide to which degree the retrieval result should be extended by lower ranked subject headings, be they assigned by untrained people (tagging) or some automatic indexer.

```
SELECT DISTINCT ?document ?subject WHERE {
 ?t rdf:subject ?document .
 ?t rdf:predicate dc:subject .
 ?t rdf:object ?subject .
 ?t ex:source ?source .
 ?source ex:type ?type .
 FILTER ( ?type = <ex:types/manual>  )
}
```

document	*subject*	*type*
ex:docbase/doc1	ex:thes/sub40	http://example.org/types/manual
ex:docbase/doc1	ex:thes/sub30	http://example.org/types/manual

```
SELECT DISTINCT ?document ?subject WHERE {
 ?t rdf:subject ?document .
 ?t rdf:predicate dc:subject .
 ?t rdf:object ?subject .
 ?t ex:source ?source .
 ?source ex:type ?type .
 ?t ex:rank ?rank .
 FILTER ( ?type = <ex:types/manual> || ?rank > 0.7 )
}
```

document	*subject*	*rank*
ex:docbase/doc1	ex:thes/sub40	1.0
ex:docbase/doc1	ex:thes/sub30	1.0
ex:docbase/doc1	ex:thes/sub30	0.8

Example 5: Ranked assignments and additional source information

III. Related Work

Early initiatives to define a vocabulary and usage-guidelines for the provenance of metadata was the A-Core [7] and based on it the proposal [8] for the DCMI Administrative Metadata Working Group (http://dublincore.org/groups/admin/). The working group finished its work in 2003 and presented the Administrative Components (AC) in [9], that addressed metadata for the entire record, for update and change and for batch interchange of records. Both initiatives focused more on the definition of specific vocabularies to describe the provenance of metadata. There was not yet a concise model to relate the metametadata with the metadata. For example, there was only an example given, hot to use the AC in an XML representation. This is not enough to enable applications the automatic integration of these information without proper knowledge, how the information is actually represented from a data model perspective.

An implementation with a clear semantic of metadata provenance statements is included in the protocol for metadata harvesting by the The authors (Rephrase with cite) in [10] (OAI-PMH). But the provenance information can only be provided for a whole set of metadata and there is no easy way to extend it with other additional information. The Open Archives Initiative provides with Object Reuse and Exchange (ORE) another, more abstract approach that addresses the requirement of provenance information for aggregations of metadata [11]. ORE particularly introduces and motivates the idea to give metadata aggregations specific URIs to identify them as independent resources. Essentially, ORE postulates the clear distinction between URIs identifying resources and URIs identifying the description of the resources. This is in line with the general postulation of “Cool URIs”[12] and the proposed solution to the so called httpRange-14 issue[8].

Hillmann et al. [13] considered the problem of metadata quality in the context of metadata aggregation. While mainly focused on the practical problems of

[8]httpRange-14 (http://www.w3.org/2001/tag/issues.html#httpRange-14) was one (the 14th) of several issues that the Technical Architecture Group (TAG) of the W3C had to deal with: “What is the range of the HTTP dereference function?”Basically, the problem is that if a URI identifies a resource other than a webpage (non-information resource), then under this URI, no information about the resource can be provided, because in this case, the URI would also be the identifier for this information. The solution is to use HTTP redirects in this case, as described in this mail: http://lists.w3.org/Archives/Public/www-tag/2005Jun/0039.html

aggregation, the paper addresses the aspect of subsequent augmentation with subject headings and changes the emphasis from the record to the individual statement. Noting provenance and means of creation on this level of detail is considered necessary by the authors. They proposed an extension of OAI-PMH to implement their solution. [14] further expands on quality issues and note inconsistent use of metadata fields and the lack of bibliographic control among the major problems. Preserving provenance information at the repository, record or statement level is one of the proposed methods to ensure consistent metadata quality.

Currently, the W3C Provenance Incubator Group (Prov-XG, http://www.w3.org/2005/Incubator/prov/) addresses the general issue of provenance on the web. The requirements abstracted from various use-cases are summarized and further explained in by Zhao et al. [15]. The conclusion of this paper is basically ours: We need further standardization for the representation of provenance information for interoperable provenance-aware applications. They recommend that a possible next RDF standard should address the provenance issue.

Lopes et al. [16] emphasize the need for additional information as well, they refer to them as annotations and examine the need for annotations without consideration of the actual implementation - be it reification or named graphs. They come up with five types of annotations – time, spatial, provenance, fuzzy and trust – that can be seen as the most obvious use-cases for additional information.

A general model for the representation of provenance information as well as a review of provenance-related vocabularies is provided by The authors (Rephrase with cite) in [17]. The model aims to represent the whole process of data creation and access, as well as the publishing and obtaining of the associated provenance information.

With the Open Provenance Model (OPM, http://openprovenance.org/) exists a specification for a provenance model that meets the following requirements [18]: Exchange of provenance information, building of applications on top of OPM, definition of provenance independent from a technology, general applicability, multiple levels of descriptions. Additionally, a core set of rules is defined that allow to identify valid inferences that can be made on the provenance representation.

Finally, a comprehensive survey about publications on provenance on the web was created by The authors (Rephrase with cite) in [19], who also mentions approaches to modeling provenance in OWL ontologies.

The most powerful means to dealing with metametadata in OWL is the use of higher-order logics, which is supported, e.g., by OWL Full. However, as this type of metamodeling comes at the expense of decidability [20], weaker forms of metamodeling such as *punning*, a restricted way of using identical names for different types of entities (e.g. classes and individuals), have been proposed by the OWL community. In OWL 2, annotation properties can be used to make statements about entities, axioms and even annotations, but as annotation properties do not have a defined semantics in OWL, integrated reasoning over the various layers of metadata requires additional implementation effort [21]. Vrandecic et al. [22] discuss different metamodeling options by virtue of several use cases, including the representation of uncertainty in ontology learning [1], as well as ontology evaluation based on OntoClean (see also [23]). In addition to these application scenarios, weak forms of metamodeling in OWL are used, e.g., for including linguistic information in ontologies [24], but only few of these approaches are able to leverage the full power of logical inference over both metadata and metametadata [25].

IV. Conclusion

This paper is meant as a discussion paper. We have proposed five principles for the proper representation of metametadata which, in our opinion, have to be met by all source-agnostic, yet provenance-aware, linked data applications.

We have demonstrated that the technical requirements can already been met, and that the remaining problem is concerned with the establishments of conventions which define best-practice recommendations. In particular, these conventions should clarify how the metametadata is actually represented – so that an application can become aware of this metametadata, retrieve, maintain and republish it in a proper way. Currently, there is no accepted best-practice that follows our principles. We are involved in the Metadata Provenance Taskgroup of the Dublin Core Metadata Initiative[9] which aims to develop such best-practice recommendations in an as-open-as-possible way. This is why we are seeking for feedback, ideas and contributions to the ongoing discussions and the outcomes of this task group – because we want metadata provenance. Now!

Acknowledgements. Johanna Völker is financed by a Margarete-von-Wrangell scholarship of the European Social Fund (ESF) and the Ministry of Science, Research and the Arts Baden-Württemberg.

[9] http://dublincore.org/groups/provenance/)

References

[1] P. Haase and J. Völker, "Ontology learning and reasoning – dealing with uncertainty and inconsistency," in *Uncertainty Reasoning for the Semantic Web I*, ser. Lecture Notes in Computer Science. Springer, 2008, vol. 5327, pp. 366–384.

[2] J. J. Carroll, C. Bizer, P. Hayes, and P. Stickler, "Named Graphs, Provenance and Trust," in *Proceedings of the 14th International Conference on World Wide Web (WWW) 2005, May 10-14, 2005, Chiba, Japan*, 2005, pp. 613–622.

[3] S. Schenk and S. Staab, "Networked Graphs: A Declarative Mechanism for SPARQL Rules, SPARQL Views and RDF Data Integration on the Web," in *Proceedings of the 17th International Conference on World Wide Web (WWW) 2008, April 21-25, 2008, Beijing, China*, 2008.

[4] G. Flouris, I. Fundulaki, P. Pediaditis, Y. Theoharis, and V. Christophides, "Coloring rdf triples to capture provenance," in *ISWC '09: Proceedings of the 8th International Semantic Web Conference*. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 196–212.

[5] F. Gandon and O. Corby, "Name That Graph - or the need to provide a model and syntax extension to specify the provenance of RDF graphs," in *Proceedings of the W3C Workshop - RDF Next Steps, June 26-27 2010, hosted by the National Center for Biomedical Ontology (NCBO), Stanford, Palo Alto, CA, USA*, 2010. [Online]. Available: http://www.w3.org/2009/12/rdf-ws/papers/ws06/

[6] K. Eckert, M. Pfeffer, and H. Stuckenschmidt, "A Unified Approach For Representing Metametadata," in *DC-2009 International Conference on Dublin Core and Metadata Applications*, 2009.

[7] R. Iannella and D. Campbell, "The A-Core: Metadata about Content Metadata," 1999, internet-Draft Document. [Online]. Available: http://metadata.net/admin/draft-iannella-admin-01.txt

[8] J. Hansen and L. Andresen, "Administrative Dublin Core (A-Core) Element," 2001.

[9] ——, "AC - Administrative Components: Dublin Core DCMI Administrative Metadata," 2003, final release of the Dublin Core Metadata Initiative Administrative Metadata Working Group. [Online]. Available: http://www.bs.dk/standards/AdministrativeComponents.htm

[10] Open Archives Initiative, "The Open Archives Initiative Protocol for Metadata Harvesting," 2008, protocol Version 2.0 of 2002-06-14, Edited by Carl Lagoze, Herbert Van de Sompel, Michael Nelson and Simeon Warner. [Online]. Available: http://www.openarchives.org/OAI/openarchivesprotocol.html

[11] ——, "Open Archives Initiative - Object Reuse and Exchange: ORE User Guide - Primer," Open Archives Initiative, 2008, edited by: Carl Lagoze, Herbert van de Sompel, Pete Johnston, Michael Nelson, Robert Sanderson and Simeon Warner. [Online]. Available: http://www.openarchives.org/ore/1.0/primer

[12] W3C SWEO Interest Group, "Cool URIs for the Semantic Web: W3C Interest Group Note 03 December 2008," 2008, edited by: Leo Sauermann and Richard Cyganiak. [Online]. Available: http://www.w3.org/TR/cooluris/

[13] D. I. Hillmann, N. Dushay, and J. Phipps, "Improving Metadata Quality: Augmentation and Recombination," in *Proceedings of the International Conference on Dublin Core and Metadata Applications*. Dublin Core Metadata Initiative, 2004. [Online]. Available: http://hdl.handle.net/1813/7897

[14] D. I. Hillmann, "Metadata Quality: From Evaluation to Augmentation," *Cataloging & Classification Quarterly*, vol. 46, no. 1, 2008. [Online]. Available: http://ecommons.library.cornell.edu/bitstream/1813/7899/1/Metadata_Quality_rev.pdf

[15] J. Zhao, C. Bizer, Y. Gil, P. Missier, and S. Sahoo, "Provenance Requirements for the Next Version of RDF," in *Proceedings of the W3C Workshop - RDF Next Steps, June 26-27 2010, hosted by the National Center for Biomedical Ontology (NCBO), Stanford, Palo Alto, CA, USA*, 2010.

[16] N. Lopes, A. Zimmermann, A. Hogan, G. Lukacsy, A. Polleres, U. Straccia, and S. Decker, "RDF Needs Annotations," in *Proceedings of the W3C Workshop - RDF Next Steps, June 26-27 2010, hosted by the National Center for Biomedical Ontology (NCBO), Stanford, Palo Alto, CA, USA*, 2010.

[17] O. Hartig, "Provenance Information in the Web of Data," in *Proceedings of the Workshop on Linked Data on the Web (LDOW) 2009, April 20, 2009, Madrid, Spain*. CEUR-WS, 2009, pp. 1–9. [Online]. Available: http://www.dbis.informatik.hu-berlin.de/fileadmin/research/papers/conferences/2009-ldow-hartig.pdf

[18] L. Moreau, B. Clifford, J. Freire, Y. Gil, P. Groth, J. Futrelle, N. Kwasnikowska, S. Miles, P. Missier, J. Myers, Y. Simmhan, E. Stephan, and J. V. den Bussche, "The Open Provenance Model Core Specification (v1.1)," 2009. [Online]. Available: http://openprovenance.org/

[19] L. Moreau, "The Foundations for Provenance on the Web," *Foundations and Trends in Web Science*, November 2009. [Online]. Available: http://eprints.ecs.soton.ac.uk/18176/

[20] B. Motik, "On the properties of metamodeling in owl," in *International Semantic Web Conference*, ser. Lecture Notes in Computer Science, Y. Gil, E. Motta, V. R. Benjamins, and M. A. Musen, Eds., vol. 3729. Springer, 2005, pp. 548–562.

[21] D. T. Tran, P. Haase, B. Motik, B. C. Grau, and I. Horrocks, "Metalevel information in ontology-based applications," in *Proceedings of the 23th AAAI Conference on Artificial Intelligence (AAAI)*, Chicago, USA, July 2008.

[22] D. Vrandecic, J. Völker, P. Haase, D. T. Tran, and P. Cimiano, "A metamodel for annotations of ontology elements in owl dl," in *Proceedings of the 2nd Workshop on Ontologies and Meta-Modeling*. Karlsruhe, Germany: GI Gesellschaft für Informatik, Oktober 2006.

[23] C. Welty, "Ontowlclean: Cleaning owl ontologies with owl," in *Proceeding of the 2006 conference on Formal Ontology in Information Systems*. Amsterdam, The Netherlands, The Netherlands: IOS Press, 2006, pp. 347–359.

[24] P. Buitelaar, P. Cimiano, P. Haase, and M. Sintek, "Towards linguistically grounded ontologies," in *ESWC*, ser. Lecture Notes in Computer Science, L. Aroyo, P. Traverso, F. Ciravegna, P. Cimiano, T. Heath, E. Hyvönen, R. Mizoguchi, E. Oren, M. Sabou, and E. P. B. Simperl, Eds., vol. 5554. Springer, 2009, pp. 111–125.

[25] B. Glimm, S. Rudolph, and J. Völker, "Integrated metamodeling and diagnosis in owl 2," in *Proceedings of the International Semantic Web Conference (ISWC)*, November 2006, to appear.