OM2R

Semantic Documentation of Ontology Mapping Lifecycle to Support Retrieval and Reuse of Ontology Mappings

A thesis submitted to the

University of Dublin, Trinity College

for the degree of

Doctor of Philosophy

Hendrik Thomas,

Knowledge and Data Engineering Group,

Trinity College,

Dublin

2014

DECLARATION

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Signed:

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Hendrik Thomas

May 2014

Permission to lend or copy

I, the undersigned, agree that the Trinity College Library may lend or copy this thesis upon request.

____

Hendrik Thomas

May 2014

Acknowledgements

I would like to thank Declan O'Sullivan and Rob Brennan for supervising my research, for their never-ending support, feedback and guidance from the other side of the world. Without their great persistence and confidence in my research this thesis never would have been possible.

My thanks also to everyone in the KDEG research group and the FAME project for the great community, collaboration and support they offered me.

I would like to give a special word of thanks to Bernd Markscheffel for introducing me to the world of research and the Semantic Web community.

In addition, thanks to Carola Winges and Sharron Kraakman for the proof reading and helpful feedback. My thanks to Rami Ghorab for helping me submit the thesis.

My utmost gratefulness and sincerest thanks to my mother and father, for all their love, devotion, care, encouragement and help throughout my PhD-journey.

I would like to especially thank my partner Christiane Winges for her absolute belief, unwavering support and never-ending love.

Finally, thanks are due to my daughter Emily, for reminding me every day how wonderful the world is and that we all need to do our best to make it a better place.

This thesis is dedicated to the memory of my grandparents, who in a way made all this possible.

Abstract

Ontology mappings are of critical importance for the Linked Data and the Semantic Web communities as they can help to mitigate the effects of heterogeneities, which are a major obstacle to the promise of interoperability of knowledge. To reduce creation costs and enable automated runtime integration, the description, discovery and most of all re-use of existing ontology mappings are needed. Meta-data can help to retrieve ontology mappings, to apply them and to manage them over time.

The research question posed in this thesis is: To what extent can a meta-data model aid an ontology engineer in the creation of consistent and relevant documentation of ontology mappings, for applications such as those supporting ontology mapping reuse? To address this question the ontology based OM2R model for documenting ontology mappings was designed. Experiments show that the OM2R model can help to improve mapping discovery and that the model is relevant for mapping reuse. Furthermore, the evaluation showed that the OM2R model can support the creation of consistent meta-data.

In summary, the OM2R model developed supports ontology engineers in the creation of relevant and consistent ontology mapping documentation which will be vital to address the ongoing need for mapping retrieval and reuse in the Semantic Web and Linked Data community.

Table of Contents

Figures

Tables

Abbreviation

FDM	Federal Relationship Manager
HTML	Hypertext Markup Language
IR	Information Retrieval
ISO	International Organization for Standardization
OM2R	Semantic Meta-Data Model of documenting ontology mapping
OMG	Open Management Group
OWL	Web Ontology Language
OWL DL	Web Ontology Language Description Logic
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
SQL	Structured Query Language
TM	Topic Maps
UML	Unified Modelling Language
URI	Uniform Resource Identifier
W3C	World Wide Web Consortium
WSDL	Web Services Description Language
XML	eXtensible Markup Language

1 Introduction

1.1 Motivation

Ontology mappings[1] are of critical importance for both the dynamic web of Linked Data [He11, Pr10] and the Semantic Web [Be11, Sh12]. The importance stems from the fact that mappings can help to mitigate the effects of heterogeneities[2] [Eu07] in ontologies. These heterogeneities represent a major obstacle to the promise of interoperability [Gr94] of knowledge [La08] within Linked Data datasets and within the Semantic Web.

The database community has long experience with mappings[3] [Be11c]. Their work has shown that creation of mappings is a complex and time-consuming task [Ka03, Do05]. To reduce creation costs and enable automated runtime integration, the description [Ro06, Je09], discovery and, most of all, re-use of existing ontology mappings is needed [We10, Sh12].

The database and library research communities have shown that retrieval and reuse of information in general requires suitable meta-data about the information itself [Be05, Ed07, No08]. The ISO standard ISO/IEC 11179-1 defines **meta-data** as "data that defines and describes other data" [IS04], commonly summarized as "data about data" [Ni04]. The generation and application of meta-data poses significant challenges. For example, a key challenge is that meta-data creation is time-consuming [Ed07, Te11]. This requires meta-data to be relevant for its application purpose, as users are more likely to document information which they deem relevant [Fu93, Du02]. For example the name of the author can be vital to judge the impact and validity of claims made in a research paper. Another challenge is that meta-data can only be used to its full potential if it is created consistently [Du10, IS04], in other words that it is complete and free of

[1] Ontology mappings are defined as a process of relating the vocabulary of two ontologies sharing a domain in such a way that the structure of ontological signatures and their intended interpretations are respected [Ka03].

[2] Heterogeneities between ontologies can be found on the syntactic, semantic and pragmatic levels [Pe97].

[3] The database community refers not specifically to mapping but rather to so called schema matchings. This is defined as the problem of generating correspondences between elements of two schemas. The involved challenges are very similar to those of ontology mappings [Mo09].

contradiction. For example, retrieval could fail to return suitable mappings because the purpose of a meta-data field has been interpreted differently by users.

Meta-data can help to retrieve ontology mappings, to apply them and to manage them over time. Jerome Euzenat, one of the leading researchers in the area of ontology interoperability, stated that mappings "must be complemented with rich metadata" [Sh08] to support their retrieval and management [Te11, Sh12].

The ISO standard ISO/IEC 11179-1 defines a **meta-data model** as a "representation of data, specifying their properties, structure and inter-relationships" [IS04], for example the Learning Object Model (LOM) [Lo09]. Meta-data models can help users create consistent documentation as they provide a template with ready-to-use model content [Ni04, Te11]. Models also can assist users in understanding the intended meaning [Fu93] and reduce creation efforts by guiding to application relevant aspects.

However, the current meta-data used to document ontology mappings is heterogeneous[4] [Eu11a], scarcely documented and where meta-data models do exist they focus primarily on representing specific alignments[5] [Sh12].[6]

In the opinion of the author of this thesis, a consequence of this situation is that requirements of applications, such as mapping retrieval and mapping reuse, are not addressed sufficiently. This may result for example, in insufficient meta-data for unambiguous identification of addressed ontologies and no guidelines exist as to how to create consistent documentation for mappings. These limitations present a major obstacle for the further development of applications that utilise mappings, such as those in the Linked Data area concerned with interlinking [Bi09, Ch13] or those in the integration area concerned with information uplift[7].

---

[4] In terms of expressiveness, language and scope [Th09]

[5] "Alignment is a set of correspondences between two or more (in case of multiple matching) ontologies (by analogy with molecular sequence alignment). The alignment is the output of the matching process." [Eu07]

[6] Please see chapter 2 for further details on the state of the art study on ontology mapping documentation.

[7] Semantic uplift is defined as the process of converting non-RDF data into an RDF based knowledge representation. [Br13]

1.2 Research Question

The research question posed in this thesis is:

To what extent can a meta-data model aid an ontology engineer in the creation of consistent and relevant documentation of ontology mappings for applications such as those supporting ontology mapping reuse?

The task of ontology mapping is defined as a process of relating the vocabulary of two ontologies sharing a domain in such a way that the structure of ontological signatures and their intended interpretations are respected [Ka03], resulting in a set of ontology mappings. An ontology engineer is defined as human with some experience in Semantic Web technologies. Ontology mapping reuse is defined as an activity where an existing ontology mapping is utilised to relate ontologies in another application context.

The meta-data model proposed in this thesis itself is designed as an ontology [An04]. This allows expressive representation to reduce ambiguity of the intended meaning of the meta-data fields [Gr94, Ma11]. Furthermore, it supports a flexible extension of the model to accommodate new developments in the dynamic field of mapping tools [Eu11a]. The meta-data model is designed as an additional documentation layer so it can be used independently from any specific mapping representation [Eu06] (see section 3.3).[8]

This thesis will focus on the activities that can be found in the ontology mapping lifecycle[9] as defined in [Su07] in order to help identify the meta-data aspects. The application areas of mapping reuse and retrieval have been chosen as the driving force to identify the relevance of these meta-data fields. Mapping retrieval is the key activity in mapping reuse and will therefore be a focus of this thesis (see section 3.4.4).[10]

---

[8] An example for representation of mappings are OWL [An04] or the Ontology Alignment Format [Eu07].

[9] The ontology mapping lifecycle describes the activities involved in creating a ontology mapping including identification of the address ontology, matching of ontologies, generating mappings from matches, leading to deployment of mappings and subsequent management activities including the application of and the reuse of mappings.

[10] The rationale is that information is required for an activity in the ontology mapping lifecycle or which are a result of such an activity are potentially relevant for an application such as mapping reuse and needs to be documented in the meta-data model.

The aspect that is of interest in the research question is that the meta-data model will aid creation of documentation that is **consistent**. In the American heritage dictionary consistency is defined as: "Agreement or logical coherence among things or parts [...], correspondence among related aspects." [Di00]. As the meta-data model proposed in this thesis is designed as an ontology, in this thesis consistency of the meta-data model is designed as a combination of two dimensions: logical and application consistency [Ha05] [11]. The first dimension is *logical consistency* where the meta-data model is viewed as a logical theory, and so is considered consistent if it does not contain contradictory information. In this context the model contains explicit relations to highlight contradiction-free meta-data field combinations. For example a child might be described by meta-data fields for age and height. However, it would be contradictory if the age was set to 3 years and the height to 2 meters as this is impossible.[12] The second dimension is *application consistency,* which addresses the specific conditions defined by the application [Ha05]. In this thesis, the meta-data will aid in the creation of application consistent documentation, if it contains the relevant meta-data fields. The activities in the ontology mapping lifecycle (see section 2.2.2 for details) provide the basis for the identification of relevant meta-data fields. Through reference to the lifecycle it will be determined if an individual mapping is documented correctly and completely.

The second aspect in the research question that is of interest is that the meta-data model will aid creation of documentation that is **relevant**. Relevance is a generic attribute and to justify the state of relevance depends on the individual domain and the object of interest. For example in the American heritage dictionary, the term relevance is defined as: "Having a bearing on or connection with the matter at hand." [Di00]. Thus, the meta-data model is considered relevant if it increases the likelihood of accomplishing the goals,

---

[11] Please note Haase defines *structural* consistency as a third dimension which ensures that the ontology "obeys the constraints of the ontology language with respect to how the constructs of the ontology language are used." [Ha05]. This aspect is tightly coupled to the toolset question, as a tool can help to enforce structural consistency. In this thesis the users always experience the OM2R meta-data model through a tool (see section 4.4.2 OM2R Editor) and as such the structural consistency can be guaranteed and will not be analysed further.

[12] In the open world scenario it is not possible to document every single possible compatible and incompatible relation between each meta-data field. As a result in this thesis two documentation instances are considered logically consistent, if they both contain only meta-data relations, where a compatible relation is expressed in the OM2R model.

which are implied by the task[13] [Ho02], which in this thesis is documentation that will support applications such as mapping reuse.

1.3 Research Objectives

To address the research question, the following research objectives have been derived:

Research Objective 1: Identify the requirements for meta-data documentation for applications such as mapping reuse. Additionally, identify the characteristics of meta-data support in current mapping tools and mapping representations (notations and formal languages).

Research Objective 2: Design a meta-data model for ontology mapping documentation that can support the creation of consistent and relevant ontology mapping documentation.

Research Objective 3: Evaluate if the meta-data model is able to support the creation of consistent and relevant documentation for ontology mappings in support of applications such as mapping reuse.

1.4 Technical Approach

This section outlines the activities undertaken in this thesis to answer the stated research question, including a state of the art study, three experiments and two case-studies.

A study of 22 common mapping representations used in 13 mapping and matching tools was carried out [Th09]. The objective was to identify the supported meta-data fields and their characteristics. More specifically, the extent was analysed to which ontology mapping lifecycle aspects can be documented with meta-data available in each mapping representation. Overall the study showed that the currently supported meta-data for mappings is heterogeneous and poorly documented. The focus instead was found to be primarily on design of the representations themselves. Furthermore, in 2013 a further evaluation of five tools and four common matching languages were conducted (see section 2.3). The result shows that the coverage of ontology mapping lifecycle aspects is

[13] Please note more specific definitions are available for various applications, e.g. in information retrieval relevance is the result of a match between the subject descriptors of a query and the descriptors assigned to a document. [Co97]

still limited; with for example, often no details about the mapping process being documented.

Based on the study, it was confirmed that the current approaches to meta-data for mappings do not support the creation of consistent and relevant documentation for ontology mappings. In addition, the result of this study provided the basis to identify the key meta-data model requirements that would aid the creation of consistent and relevant documentation for ontology mappings.

The design of the OM2R model is based on a requirement analysis that includes a literature review of common meta-data challenges (see section 3.2). Based on the identified challenges the library domain was evaluated as a common domain for meta-data use. The objective was to identify how commonly relevance and consistency is supported in meta-data models. In addition, the insights gained in the state of the art study (see section 2.7 and 3.3) were used to identify requirements for the OM2R model, particularly in relation to the current manner in which meta-data for mappings are supported. Furthermore, a domain analysis (see section 3.4) was conducted. More specifically, the impact of the extended ontology mapping lifecycle on mapping retrieval and as such on the requirements of the model to support mapping reuse is discussed. These evaluations helped to identify of the following key requirements for the OM2R model: an explicit semantic representation, explicit documentation of the model, design of meaningful relations between the meta-data fields and the ability to document the ontology mapping lifecycle to support retrieval and as such mapping reuse.

These identified requirements are addressed with six design features that constitute the design of the OM2R model. The implementation of the OM2R model as an OWL DL based ontology [Br05] provides a ready-to use model for meta-data documentation of ontology mappings (see section 4.1).

The OM2R model was evaluated in a series of three experiments. **The first experiment evaluated the benefits of an <u>ontology-based model structure</u> for documenting ontology mappings.** This experiment was designed as an automated retrieval test in a lab environment. It was based on a set of common mapping discovery tasks which were applied to sample mappings in two custom-built mapping retrieval tools.[14] The experiment showed that the ontology-based OM2R model approach has benefits for mapping discovery in terms of better retrieval efficiency and effectiveness.

Mapping reuse depends on many external factors which cannot be quantified fully for a valid automated evaluation. Feedback from a diverse group of participants is needed to mitigate the impact of the heterogeneity of the mapping reuse-cases [Eu06] which exist. **The second experiment was designed to create a <u>relevance</u> ranking of the proposed meta-data fields in the OM2R model by ontology engineers.** The aim was to rank an identified list of meta-data fields based on the lifecycle analysis. These were ranked by their relevance for mapping reuse as application, in order to reduce the effort needed to populate the meta-data model. Mapping reuse was chosen as the application as it is a high level task and incorporates more specific applications such as mapping retrieval.

The experiment was conducted in a lab environment. The participants were recruited online from the W3C Linked Open Data Project Mailing List[15] and universities with a known publication record on ontology matching, mapping experience and the Semantic Web. 49 participants were given two specific mapping reuse scenarios. The users were asked to rate the relevance for each of the OM2R meta-data fields, this is considered a task-oriented relevant rating in an online tool. In addition, the users had to select the most useful and least useful meta-data field for mapping reuse. This is considered a subjective rating as it is independent from a specific application task. These relevance ratings provided evidence for the creation of a relevance ranking among the proposed OM2R meta-data fields.[16] The specific metrics used are defined in table 1.

---

[14] One tool was based on the OM2R model version 1 and the other on the Ontology Alignment Format [Eu12].

[15] The public-lod@w3.org mailing list provides a discussion forum for members of the Linking Open Data project and the broader Linked Data community. The Linking Open Data project is a grassroots community effort founded in February 2007 as a W3C Semantic Web Education and Outreach Interest Group Community Project.

[16] Please note the relevance rating can be different depending if the user is asked to rate the relevance of a field in general (e.g. if presented only with the field name and definition) or if he/she is given the field and a specific field content for a given specific scenario.

Table 1 Metrics for relevance

Metric	Formula
Task-oriented - average relevance rating per meta-data field	Sum of the individual relevance rating entered by each participants / total number of participants
Subjective - average relevance rating per meta-data field	+1 for every time a user's rated this field as most relevant; -1 for every time a user rated this field as least relevant / total number of participants who rated this field

The results showed that fields related to the identification of the addressed ontologies, and those related to the matching and mapping process was considered most relevant. However, all proposed fields were considered relevant for mapping reuse to some degree by the participating ontology engineers.

The third experiment focussed on the support that the OM2R model can offer a user for the creation of consistent documentation for ontology mappings. More specifically, the experiment investigated if the OM2R model can be more supportive towards users if the meta-data model was expressed in an ontology rather than as an index-based structure. The index-based structure is used as a proxy for the current manner in which mapping meta-data is represented. For example, the Ontology Alignment Format, which is expressed as an RDF(S) file, only provides key-value meta-data fields which are typical for index-based meta-data structures. This provided a baseline to evaluate the contribution of the ontology-based OM2R model.

In this experiment, participants with a known track record in Semantic Web technologies were recruited from three international universities[17] and W3C Linked Open Data Project Mailing List[18]. The participants were asked to document two given mapping lifecycle steps based on textual instructions in an online lab environment in two stages. 48 participants completed the first stage of the experiment based on the ontology-based

[17] More specifically participant were invited from the Knowledge and Data Engineering group Trinity College (Ireland) , Information and Knowledge group, Technical University of Ilmenau (Germany) , Computer science group, Massey University of Auckland (New Zealand).

[18] The public-lod@w3.org mailing list provides a discussion forum for members of the Linking Open Data project and the broader Linked Data community. The Linking Open Data project is a grassroots community effort founded in February 2007 as a W3C Semantic Web Education and Outreach Interest Group Community Project.

OM2R editor.[19] 24 participants completed the second stage of the experiment with index-based OM2R editor.[20]

To measure the consistency of ontology mapping documentation (see section 1.4 for a detailed definition of consistency in this context) the common information retrieval metrics precision[21] and recall[22] were used [Eu07]. Generally these metrics are based on the comparison of an expected result with the effective result of the evaluated system. These results are considered as a set of items and in this case either the presence of a selected meta-data field or meta-data field combination in a gold standard[23]. Such measures are commonly used, well understood [Be11] and have been used for similar applications in ontology mapping evaluations in the past [Do02, Eu07]. The specific metrics for the logical and application dimension of consistency are defined in table 2.

Table 2 Metrics for consistency

Metric	Formula
Recall (application[24])	= number of expected meta-data field sections in the documentation instance / total number of expected meta-data field section in gold standard
Recall (logical[25])	= number of expected meta-data field combinations in the documentation instance / total number of expected meta-data field combinations in the gold standard
precision (application)	= number of expected meta-data field sections in the documentation instance / total number of meta-data field sections in the documentation instance
precision (logical)	= number of expected meta-data field combinations in the documentation instance / total number of meta-data field combinations in the documentation instance

19 According to log files the experiment system was accessed 61 times. Five of these accesses are linked to a test user to check if the experiment was accessible. These entries have been excluded from the result set.

20 Please note 26 users accessed the experiment web page and approved the experiment participation form. However, they did not proceed beyond the tutorial page and provided no relevant experiment data. This means 92.3% of the users who accessed the experiment page actually documented some or all of the experiment tasks. According to log files one additional access was registered which is linked to a test account used to check if the experiment was accessible. This entry is excluded from the result set.

21 Recall: Generally this metric is a measure of the ability of system to present all relevant items [Sa83].

22 Precision: A measure of the ability of a system to present only relevant items [Sa83].

23 In experiment three, the users are presented with a textual description of a specific mapping lifecycle. They are asked to document this information by using the OM2R model. Based on the provided instructions the author of the thesis defined the relevant information and translated them into an instance of the OM2R model which is complete, correct and contains no inconsistent statements. This OM2R instance is 100 % application and logical consistent and is used as a gold standard.

24 A meta-data model for ontology mapping documentation is application consistent if it contains the relevant meta-data fields and content for a given lifecycle scenario as defined in a gold standard.

25 In this context this relates to the fact that compatible relations exist between individual meta-data content options in the OM2R e.g. an ontology was expressed in the notation RDF/XML and in the formal language RDF(S) which reflect a compatible relation between this notion option and formal language.

The potential users of the OM2R model will most likely have diverse backgrounds. Combined with the fact that Ontology mapping is still a highly specialised topic [Eu11a] it is likely that some users will not be very familiar with mapping related concepts or the Semantic Web in general. Users with less experience in these two topics potentially create less consistent and relevant documentation. To investigate if the OM2R model can provide a similar level of support, to both those without experience and those with experience, the participants in stage 1 of experiment 3 were split into two groups[26] based on their ontology mapping experience level.

The results showed that the ontology-based representation for the OM2R model can offer a higher level of support compared to an index-based representation of the meta-data. Also, the results indicate that the OM2R model can offer a similar level of support for experienced and novice users.

Additionally, to investigate if the OM2R model can be applied beneficially for mapping reuse in real life scenarios, two case studies have been undertaken. **The first case-study focusses on the application of the OM2R model to support the documentation of alignment challenges for the "Ontology Alignment Evaluation Initiative" (OAEI)** [Eu11a]. This initiative conducts annual campaigns to assess ontology matching techniques. It is one of the few public matching repositories, and so requires support for management of reference alignments and interpretation support for users and researchers over time. This case-study was chosen as this community focusses on the interlinking challenge of heterogeneous sources, such as the Linked Data community [Ch13]. The case study shows how the OM2R model can help to create more consistent documentation in terms of higher correctness and less inconsistency for the OAEI. Furthermore, the OM2R can improve the relevance of mappings for the OAEI stakeholders as it provides more detailed lifecycle focus documentation, which is more useful for analysis of ontology matching trends and results over time. The case study also showed how the OM2R model can support third party researchers with more explicit, detailed, predictable and easy to interpret documentation.

[26] One group contained users with experience in ontologies and mappings (so called experts) and the other those without such experience (so called novices).

The second case study focusses on a federation system which aims to support data consumption from heterogeneous sources with a minimum of re-integration effort. The case study evaluates how the OM2R model can support activities of the Federal Relationship Manager (FRM) [Fe10, Ro12]. This includes mapping discovery, mapping evaluation, and change notification which contribute to a better support for mapping reuse. This case-study was chosen as this community addresses the issues of integration of information, application and services on a semantic level and as such goes beyond the matching focus of the OAEI. The case study shows that the OM2R model can satisfy the meta-data requirements of the FRM which has a focus on flexible extension and explicit documentation. In particular the ability to support consistent documentation is vital for the dynamic and reuse focus of such federations over time.

Table 3 summarises the conducted evaluations and their individual analysis focus.

Table 3 Summary of all conducted evaluations in this thesis

Activity	Focus
Experiment 1 Evaluation of the OM2R for ontology mapping retrieval	Focus on ontology mapping discovery and benefits for retrieval efficiency and effectiveness
Experiment 2 User based relevance ranking of the OM2R	Focus on relevance for mapping reuse Relevance ranking of proposed meta-data fields
Experiment 3 - User based evaluation of the OM2R model to support application and logical consistency	Focus on application and logical consistency Difference of the consistency support for novice and experts
Use-Case – Ontology alignment evaluation Initiative	Focus on application of the OM2R in a real life scenario for documentation of ontology matching
Use-Case – Federation	Focus on application of the OM2R in a real life scenario for or documentation of ontology mappings

1.5 Contribution

The major contribution of this thesis is the development of a meta-data model, OM2R model, for documentation of ontology mappings for applications such as ontology mapping reuse. Evidence indicates that the OM2R model will support the creation of both consistent and relevant ontology mapping documentation. There are also indications that the meta-data model proposed will support users who intend to reuse mappings with more explicit, detailed, predictable and easy to interpret meta-data about the mapping creation lifecycle.

The OM2R model is designed as an independent meta-data layer and can therefore be used in parallel with existing tools and mapping representations. The OM2R model can simplify the documentation process as it provides users with meta-data fields, field contents and structures relevant for mapping reuse. Thanks to the ontology-based structure of the OM2R model, the model can be extended flexibly to address new developments and individual mapping reuse scenarios. The OM2R model can support third party researchers and integration specialists that engage in mapping reuse in understanding the intended meaning of a mapping through its documentation. Experiments have shown the OM2R model can support domain experts and novice users on a similar level, which again will contribute to more consistent documentation creation.

This research will be of benefit for the fast growing Linked Data community [Um10]. The current growth of Linked Datasets increases the need for better interoperability support to allow easy and flexible consumption of data from heterogeneous sources [He11]. The OM2R model can assist data providers to document mappings more consistently and in more detail. This will assist mapping reuse and maximize network effects by making data consumption and linkages to other repositories easier. In addition, this work will also benefit the Semantic Web community efforts as the OM2R model can assist the reuse of mapping needed to bridge the gap between unstructured and semantically uplifted knowledge sources.

A minor contribution of this thesis is the practical instantiation of the meta-data model that can be used to improve the documentation for the "Ontology Alignment Evaluation Initiative" (OAEI) [Eu11a].[27] The OM2R model, first presented in [Th12], can help participants to document the community's matching submissions by providing a standard template. It will also help third party researchers to understand the meaning of the documentation fields due to its more rigorous approach. Most of all, it will benefit the OAEI organisers to manage changes to reference alignments and to track submissions over time to identify trends. The author of this thesis argues that an improved meta-data model can help to leverage the experience gained in the OAEI to extend its focus from a

[27] OAEI is the world leading initiative for evaluation of alignment/matching systems [Eu06].

pure test platform [Eu11a] to an alignments repository [Eu11b]. Thus the OAEI can showcase how alignments can be managed and reused on a large scale over time.

Table 4 provides an overview of the publications created by the author of this thesis in relation to the developed OM2R model. The listing is by publication year.

Table 4 Literature Overview of Hendrik Thomas in relation to the OM2R model

2008	Hendrik Thomas, Bernd Markscheffel, Tobias Redmann, From Subjects to Concept Clouds - Why semantic mapping is necessary, **1st International Conference on Knowledge Federation**, Inter University Centre Dubrovnik, Croatia, October 20-22, 2008, 2008.
2009	Bernd Markscheffel, Hendrik Thomas, Tobias Redmann, Developing Topic Maps Applications: Lessons Learned from a Digital Library Project, IADIS **International Conference e-Society** 2009, Barcelona, Spain, February 25-28, 2009, 2009, pp. 51 – 59.
2009	Thomas, H., O'Sullivan, D., Brennan, R.: Evaluation of Ontology Mapping Representations. In: Workshop on **Matching and Meaning**, Part of the as part of the AISB 2009 Convention, April 9th 2009. Edinburgh, Scotland 2009.
2009	Thomas, H., O'Sullivan, D., Brennan, R.: Ontology Mapping Representations: a Pragmatic Evaluation. In: **21st International Conference on Software Engineering and Knowledge Engineering.** SEKE 2009, 1 - 3 July 3, Boston, 2009, pp. 228 - 232.
2009	Thomas, Hendrik: Ontology Mapping Management, **European Summer School on Ontological Engineering and the Semantic Web** 2009, Cercedilla, Spain, 5th - 12th June 2009.
2010	Kevin Feeney, Rob Brennan, John Keeney, Hendrik Thomas, Dave Lewis, Aidan Boran, Declan O'Sullivan, Enabling decentralised management through federation, **Computer Networks Journal** (Elsevier), 54, (16), 2010, pp. 2825-2839.
2011	Thomas, H., Brennan, R. O'Sullivan, D.: MooM - a Prototype Framework for Management of Ontology Mappings. In: Proceedings of the **25th IEEE International Conference on Advanced Information Networking and Applications**, Singapore, 22-25 March, 2011, IEEE, pp. 548 – 555.
2011	Brennan, Feeney, K., Walshe, B., Thomas, H., O'Sullivan, D.: Explicit Federal Relationship Management to Support Semantic Integration, **1st IFIP/IEEE Workshop on Managing Federations and Cooperative Management**, Dublin, May 2011, IEEE, 2011, pp.1148-1156.
2012	Thomas, H., Brennan, R., O'Sullivan, D.: Using the OM2R meta-data model for ontology mapping reuse for the ontology alignment challenge - a case study. In proceedings of the **7th International Workshop on Ontology Matching (OM-2012)** collocated with the 11th **International Semantic Web Conference** (ISWC-2012) Volume 946, http://ceur-ws.org/Vol-946/ Boston, MA, USA, November 11, 2012.

1.6 Thesis overview

This thesis is organised as follows:

Chapter 2 State of the Art: defines the meta-data model and ontology mappings which are the core subject of interest in this thesis. In addition, the ontology mapping lifecycle is introduced which will be used as a basis for the OM2R model design. Furthermore, the chapter provides a state of the art study of mapping and matching tools and representations (notations and formal languages) to identify the characteristics of current meta-data support for ontology mapping documentation.

Chapter 3 Design of the OM2R model: Based on an analysis of common meta-data challenges, the result of the state of the art study and a domain analysis, the requirements for a meta-data model for documenting ontology mappings are derived. Furthermore, the extended ontology mapping lifecycle is discussed in detail as it provides the basis for the meta-data in the OM2R model. More specifically, an overview is given as to how the four identified requirements are addressed with the six design features of the OM2R model. The chapter includes a full example of the model and an overview of the evolution of the OM2R model.

Chapter 4 Implementation of the OM2R model: The first section describes how the OM2R model was implemented in a specific OWL-based ontology. Afterwards, the OM2R evaluation framework is discussed, starting with the research on model editors. This includes the prototype of a web based editor generated from the model and MooM as a proof of concept prototype for a UI based model editor.

In the following three chapters, the experiments that evaluate the OM2R model are presented.

Chapter 5 Evaluation of the OM2R model for ontology mapping retrieval: The first experiment evaluates the benefits of an ontology-based meta-data model structure based on an automated mapping retrieval hypothesis. The hypothesis for this experiment is: *The retrieval effectiveness and efficiency of ontology mapping discovery can be improved by using an ontology-based meta-data model describing the ontology, the mapping features and life-cycle information.*

Chapter 6 Participant-based Relevance Ranking of OM2R: The second experiment is a user-based relevance rating of the OM2R meta-data fields. The hypothesis for this experiment is: *The proposed meta-data fields and their structure are considered relevant for a mapping reuse decision.*

Chapter 7 User-based evaluation of the OM2R model to support application and logical consistency. The third experiment evaluates the support the OM2R model can provide to users in to the production of consistent documentation of ontology mappings. For this purpose the first hypothesis is defined as: *The OM2R model can support users in the creation of more application and logically consistent documentation for ontology mapping if the model is structured as an ontology rather than an index-based structure.* In addition, a second hypothesis is defined as: *The OM2R model can offer similar support in the creation of application and logically consistent documentation for ontology mappings for users with high ontology mapping experience, as for those with limited ontology mapping experience.*

The following two chapters presents two case studies on the application of the OM2R model in real life domains and were chosen based on an explicit need for investigating OM2R's support for mapping reuse.

Chapter 8 Case Study – OM2R for OAEI: The first study shows how the OM2R model can be used to document ontology matching for the "Ontology Alignment Evaluation Initiative" (OAEI) [Eu11a].

Chapter 9 Case Study – OM2R for Federation: The second case study focusses on application of the OM2R model to document ontology mappings in the Federal Relationship Manager (FRM) [Br06].

Chapter 10 Conclusion: This chapter reflects upon: to what extent the research objectives have been met, discusses the contributions made, presents suggestions for future work, and concludes with some final remarks.

2 State of the Art

This chapter provides an introduction to meta-data and meta-data models in general. Furthermore, the ontology mappings are defined and the ontology mapping lifecycle is presented. The core of this chapter is a state of the art study on meta-data which is used to document ontology mappings. The emphasis is placed on the nature of the meta-data and the support they offer for the creation of relevant and consistent documentation of ontology mappings. For this purpose, selected ontology mapping representations and match candidate generation tools are evaluated.[28]

2.1 Meta-Data and Meta-Data Models

Commonly, *meta-data* is characterised as "data about data" [Ni04]. In this thesis a more specific definition is used which is provided by the ISO standard ISO/IEC 11179: "data that defines and describes other data" [IS04]. This definition is more appropriate for this thesis as it highlights the descriptive nature of meta-data.

The use of the term meta-data differs between research communities. For example in computer science meta-data refers commonly to information which can be processed automatically [Sa83, Ga06]. In the librarian community it refers to any formal scheme for resource description [Ku98, Ra04]. Meta-data can facilitate the discovery and exchange of relevant information and is vital for interoperability[29] challenges [Be05,We10]. Meta-data can be classified according to National Information Standards Organization (NISO) [Ni04] using three categories: descriptive, administrative and structural.

Meta-data can be organized in *meta-data models* which are defined by ISO/IEC 11179-1 as "representations of data, specifying their properties, structure and inter-relationships" [IS04]. This definition will be used for this thesis as it focusses on properties but also on the relationships between the properties. The focus on relationship corresponds to the

---

[28] It is worth noting that no dedicated mapping tools or infrastructure systems are included in this analysis, as no dedicated comprehensive mapping tools or infrastructure currently exist. Instead the focus is placed on matching generation, which are a process step towards ontology mappings creation according to chapter 2.2.2 of the ontology mapping lifecycle. An analysis of matching tools can therefore provide insight regarding the current meta-data support.

[29] Interoperability can be defined as the ability of multiple systems with different characteristics such as software platforms, data structures, and interfaces to exchange data with minimal loss of content and functionality [Ni04].

network focus of the Semantic Web. Meta-data models are designed for a specific purpose, such as describing a particular type of information resource [Ni04]. A meta-data model specifies how a resource can be enriched with the provided collection of meta-data fields and semantics.

Meta-data models have been developed to describe a wide range of resources (textual and non-textual) and can be encoded in various syntaxes [IS04][30]. Common meta-data models are for example: the Dublin Core Meta-Data Set to describe Web-based documents [Du09, Th12], OMV an Ontology Meta-Data Vocabulary [OM12] for ontology development details such as language[31], Learning Object Meta-Data (LOM) a meta-data scheme for technology-supported learning resources [Lo09], MARC 21 Format for Bibliographic Data used for encoding electronic bibliographic catalogue records [Ma12] or Friend of a Friend (FOAF) for annotation of contact details for people [Fr12].

The importance of meta-data to describe ontology mappings has been recognised in [Sh12]. Shvaiko states that the key challenges for meta-data in the context of ontology mappings "is to provide convenient and interoperable support, on which tools and, more importantly, on which applications, can rely in order to store and share alignments. This involves using standard ways to communicate alignments and retrieve them. Hence, alignment metadata and annotations should be properly taken into account" [Sh12].

The aim of this thesis is the creation of a meta-data model for documenting ontology mappings. The design of the meta-data model is guided by the objective stated in the research question of this thesis, i.e. the meta-data model needs to support the creation of consistent and relevant documentation of ontology mappings.

The definition of a meta-data model provided by ISO/IEC 11179-1 [IS04] assists the actual design of the OM2R model by providing a clear list of all aspects of the model which need to be considered and can be used to offer support for consistency and

[30] Current meta-data models are often expressed in the Extensible Mark-up Language (XML) or the Resource Description Framework (RDF) [Be04].

[31] For example the OMV framework is used to document large-scale semantic applications in the distributed organizations in the NeOn project. Please find more information here: http://www.neon-project.org/nw/Welcome_to_the_NeOn_Project

relevance. More specifically, it relates to the representation structure which in this case is defined as an ontology based approach. Besides the selection of meta-data fields and their properties, the ISO definition is valuable as it highlights the importance of inter-relationships between the model components. In order to identify the needed meta-data to document ontology mappings it is necessary to understand the nature of ontology mappings and the activities involved in their lifecycle.

2.2 Ontology Mapping

In this section ontology mapping is defined and the ontology mapping lifecycle is presented.

2.2.1 Definition of ontology mapping

Different representations of similar information or different notations can lead to a mapping problem if such information is to be searched, exchanged or integrated across information systems. This is a well-known problem for the Linked Data community [He11, Pr10] and the Semantic Web [Be11] whose main purpose is to exchange explicitly represented knowledge across diverse environments. A common approach to mitigate the negative effects of potential heterogeneity of information models is to identify the specific correspondences between the different ontologies and to document these correspondences using an appropriate ontology mapping expression [Sh08].

Euzenat describes the ontology mapping task as follows: "Given two ontologies each describing a set of discrete entities (which can be classes, properties, rules, predicates, etc.), find the relationships (e.g., equivalence or consumption) holding between these entities."[Eu07]. In a more recent publication he defines ontology mapping as a task that "finds correspondences between semantically related entities of ontologies" [Sh12]. Bouquet provided in 2003 a more formal definition for ontology mapping [Bo03]. Given two ontologies A and B, ontology mappings are composed of a set of correspondences between pairs of simple or complex entities (e.g., terms, classes, individuals) Xa and Yb, belonging to A and B respectively. A mapping is described as the quadruple: [Xa, Yb, R, n]. The assigned correspondence R can be any suitable relation, such as a simple equivalence relation, or a complex set of relations. The degree of confidence in that correspondence is represented by n and represents an abstract level of trust in the fact that the correspondence is appropriate.

In this thesis, the following definition for *ontology mapping* is used which is derived from a definition provided by Kalfoglou: Ontology mapping is defined as the task of relating the vocabulary of two ontologies sharing a domain in such a way that the structure of ontological signatures and their intended interpretations are respected [Ka03].[32] The definition focusses on the activities involved and highlights the objective of ontology mapping. This definition is more suitable for this thesis as the meta-data model which is developed in this thesis will support the mapping activity rather than a specific mapping representation.

In the Semantic Web literature, it can be observed that different terms are used within the problem domain, for example ontology mapping vs. ontology matching. Both terms relate to the problem described by Y. Kalfoglou but focus on different stages of the process. To provide a consistent terminology within this thesis, ontology matching is defined as the process of identifying equivalence candidate correspondences between *ontology elements* (such as class, properties) based on an automated algorithm or manual evaluation as defined in [Eu04]. In contrast *ontology mapping* builds upon matching and covers the identification of correspondences using any correspondences types and that is manually confirmed [Su07]. Thus, ontology matching can be seen as a sub task of the ontology mapping process. In other words ontology matching identifies mapping candidates which consists of a number of individual alignments between ontology elements. Those alignments are later confirmed or used in the ontology mappings creation process.[33]

---

[32] Note that in the original wording of the definition he defined "ontology mapping as the task of relating the vocabulary of two ontologies that share the same domain of discourse in such a way that the mathematical structure of ontological signatures and their intended interpretations, as specified by the ontological axioms, are respected" [Ka03].

[33] This process can result in a full or complete mapping of ontologies, where for every ontology element in a source and in a target ontology an appropriate mapping correspondence is defined [Su07]. Alternatively partial mapping can contain any subset of mapping correspondence between two or more ontologies.

2.2.2 Ontology Mapping Lifecycle

Ontology mapping is a complex and time-consuming process [Ka03, Eu07]. The individual stages of the creation of an ontology mapping are described in the ontology mapping lifecycle. The lifecycle provides details about the involved activities. These are vital for the design of the ontology mapping meta-data model, as the activities indicate what information and decisions need to be documented in the model. So far no common agreement has emerged on the lifecycle phases [Su07, Sh12].

In the following, a lifecycle is presented based on that proposed by O'Sullivan [Su07]. This process was chosen as it focusses explicitly on ontology mapping creation. According to literature research no other ontology mapping lifecycle exist or is limited to ontology matching only.[34]

1. *Identification phase*: The first task is the discovery of the ontologies which will be the subject of the mappings in terms of the identification of the target and source ontologies. A second task is the characterisation of those ontologies with respect to their amenability for being mapped. The goal of this task is to identify difficulties that may be involved in undertaking a mapping, e.g. based on the form and quality of the ontology.

2. *Matching phase*: The objective of this phase is the identification of mapping candidates, either identified by manual selection or by automated matching processes [Eu04, Eu06]. This phase covers the planning of the matching process which involves making the decision as to whether a match should be attempted between two ontologies, e.g. based on individual matching policies [Be04a]. Also relevant is the selection and execution of matching algorithms because there is a wide range of matching algorithms that can be applied depending on the requirements of the individual use-case [Eu04].

[34] Note that Pavel Shvaiko and Jerome Euzenat discuss lifecycle activities that lead to the creation of ontology mappings. However, these are not explicitly structured as a lifecycle and are limited to matchings and as such are too narrow for the scope of this work [Sh12].

3. *Mapping phase*: The third stage involves the generation of information necessary for the execution of mappings as well as the creation of mappings that are relevant for the context of usage. Particularly relevant is the planning of the mapping approach and the selection of an appropriate mapping correspondence from the generated list of possible mapping candidates. The identified committed and approved mappings can then be rendered into different mapping formats in order to enable processing and sharing. This phase also includes the actual interpretation of the ontology mappings in order to enable semantic interoperability between the addressed ontology in the individual application scenario, e.g. merging or alignment.

4. *Management phase*: Ontology mappings generated in the previous phases need to be managed and maintained until their withdrawal. This includes the sharing of mapping information with third parties, the integration of mapping into other mapping applications and can include a conflict or consistency check as well as the altering of mappings if basic preconditions have changed.

This lifecycle proposal focusses only on the initial creation matching and mapping phases. This thesis uses ontology mapping reuse as the application domain for the meta-data model. To address this scope the author of this thesis has extended the lifecycle proposed by O'Sullivan [Su07]. More specifically, the management phase has been extended with activities related to mapping reuse and meta-data creation. These activities are described in more detail in chapter 3.4.2 but in the meantime figure 1 provides an overview of this ontology mapping lifecycle including the added mapping reuse related activities which are displayed as grey ellipses.

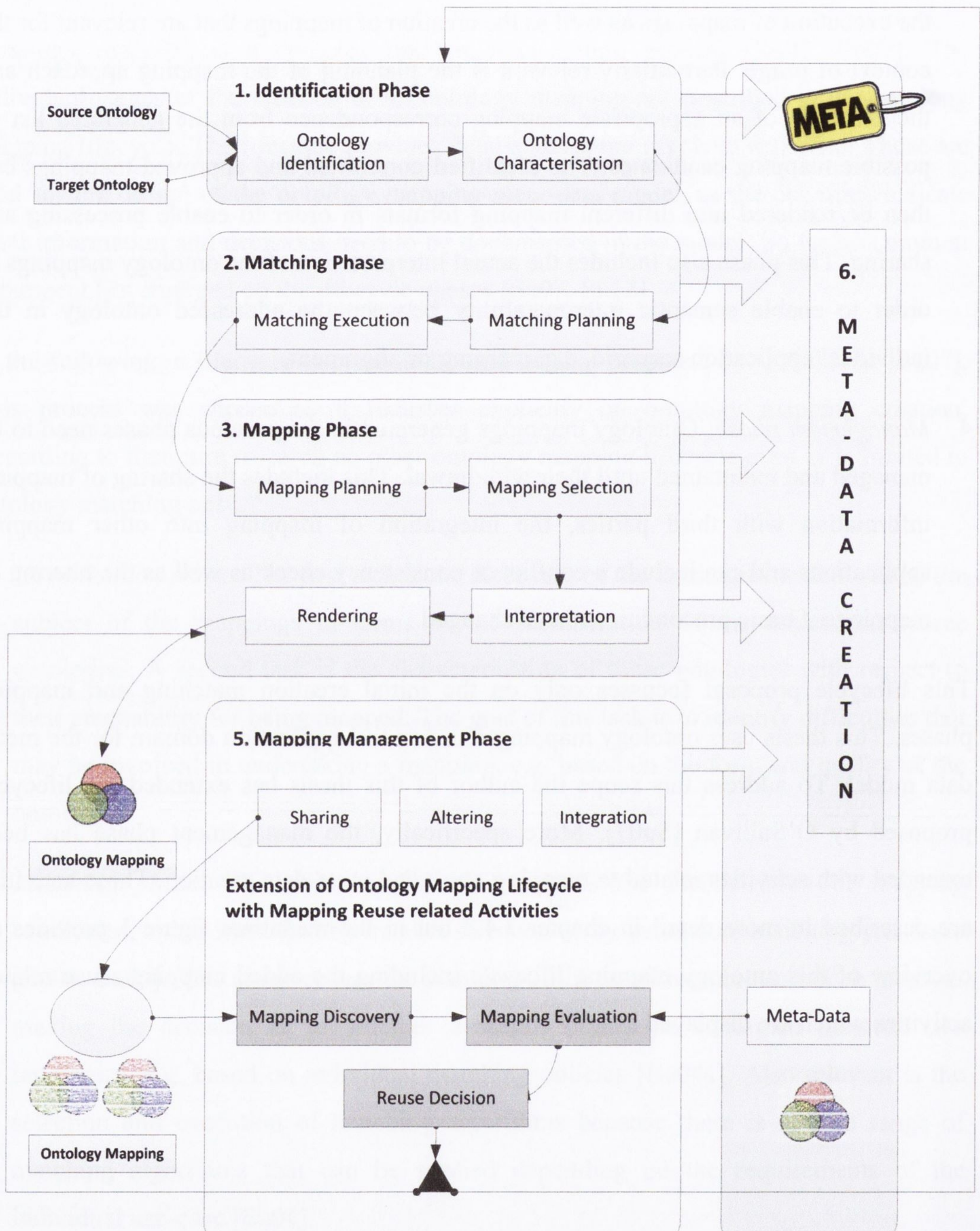

Figure 1 Extended ontology mapping lifecycle based on [Su07]

2.3 *Methodology for evaluation of current ontology mapping meta-data*

Pavel Shvaiko and Jerome Euzenat stated that “A first step in promoting sharing, manipulating and reusing alignments is to be able to use a standard for [expressing] alignments. No such standard exists at the moment.” [Sh12]. This deficiency creates the need for a review of the diverse meta-data currently used to document ontology mappings. In this thesis an emphasis is placed on reuse of ontology mappings. As support for consistent mapping documentation is vital for reuse, particular emphasis on the ability to support consistency was placed in the review.

Publications which address meta-data issues for ontology mappings are rare[35] compared to the high number related to ontology matching, e.g. matching algorithms to identify mapping candidates [Sh04]. Previous studies on ontology matching and mapping systems have focussed on technical capabilities and matching effectiveness [Ka03, Eu04, Mo09, Sh12]. A related evaluation of ontology mapping representation information was published by the author of this thesis in 2009 [Th09]. In that study 13 different mapping and matching applications were evaluated. This state of the art study is based on the previous study but was updated and extended with additional aspects focusing on consistency.

The first objective of the review is to analyse the current use of meta-data for ontology mapping. In this thesis the following metrics are applied, which are based on a proposed metric collection by Zaveri and Maurino in their systematic review of data quality aspects of Linked Open Data in [Za12]. All metrics are highlighted in italics.

The metric *number of supported meta-data fields* provides an indication of the complexity of the model and the amount of detail that can be documented for ontology mapping. Furthermore, the metric *level of relationship between the meta-data elements*

[35] This statement is based on the evaluation of the publication in the International Workshop on Ontology Matching one of the most renowned community events for ontology mapping / matching researchers. If the workshops in 2013, 2012 and 2011 are considered in total 76 OAEI papersrs and technical papers were accepted. Of these only one paper addresses the ontology mapping meta-data question [Th12]. Please note this paper presented a case study for the OAEI and was written by the author of this thesis.

provides an indication of the expressiveness of the meta-data model. The following categories are used: index relations[36], thesaurus relations[37] and ontology relations. This range of categories was chosen to cover the most common spectrum of relationships in the Semantic Web [Te11] starting from no explicit relations between fields (index) to meaningful typed relations (ontologies). In addition, the metric *ontology mapping lifecycle completeness* offers details on the kind of information that the meta-data fields are designed to document, in the specific context of ontology mapping. The completeness is defined as "the degree to which information is not missing" [Za12]. More specifically, the metric indicates which ontology mapping lifecycle stages are documented by the meta-data model: identification, characterisation, matching, mapping, management, reuse, meta-data documentation.

The second objective is to provide an overview of the current consistency support that the meta-data model can offer. According to the definition of consistency the first aspect is support for *application consistency* that relates to the correctness and completeness of the documentation. The actual offered support of a model is difficult to measure and to compare as it can be conceptual, tool or process based. In order to provide an objective metric, the focus is given to the *number of satisfied application consistency aspects* supported per evaluation. Based on a review of literature from the linked data community and librarian community, the author of this thesis has selected the following consistency aspects listed in figure 2 that can contribute to support application consistency.

[36] Traditionally used in libraries where the meta-data field consists of key value pairs, e.g. back of book indexes. No explicit relation types can be found between the fields. Even if the model uses a semantic language such as RDF(S) the information stored in key value structures. In other words the meta-data information are contained in single string values per field rather than explicit relations between objects. For example the Ontology Alignment Format can be classified as this type as it is commonly expressed in RDF(S) but contains only key-value pairs, e.g. <method>org.StringDistAlignment</method>

[37] For this level the meta-data model needs to support hierarchical relationships consisting of the relations such as broader term (hypernym) or narrower term (hyponym). These can be interpreted as class relationships or part-whole relationships.

1. *Explicit documentation of the meta-data model* which help users to understand the intended meaning, and therefore can support a homogenous application of the model in the community.[38]
2. *Documentation is embedded in the mapping meta-data model* so as to make access easier, it is available at all times during the documentation task.
3. *Meta-data fields have unique identifier* in terms of a URI which is essential during the evaluation of the model and helps with documentation exchange
4. *Meta-data model offers versioning information* which again supports the evaluation of the model.[39]
5. *Availability of examples of annotated mapping representations.* An example which focusses on the meta-data can help explain the intended application of the meta-data fields.[40]
6. *Availability of a specialised meta-data editor tool* can assist users in the creation of consistent documentation in order to help to avoid inconsistent field use and to be more convenient then a text based notation.[41]
7. *Support offered for automated population of model.* Manual meta-data creation is time-consuming and automated processing of existing meta-data fields can speed up the process and avoid inconsistencies between the meta-data and the mapping representation [De12].
8. *Model provides meta-data content options for their fields.* A controlled list of options per field or a list of suggestions can support the consistent application of the model in the community.

Figure 2 Application consistency aspects

The other part of consistency is the support for *logical consistency* that focusses on the avoidance of inconsistent statements within documentation. In a similar manner as above, the *number of satisfied logical consistency aspects* per item is evaluated. Based on the same literature review, the author of this thesis has selected the following aspects listed in figure 3 of a model in order to support logical consistency.

1. *Meta-data model contains embedded information on compatible field content combinations.* If the model provides compatible combinations then these can be used by the users and for consistency checks.
2. *Documentation of the meta-data model does provide information on compatible field content combinations.* If not embedded in the model such information can be provided in the documentation and be helpful as well.[42]
3. *Support of automated consistency verification services.* This can be a great help for users to check their own documentation but also the creation of other documentation for logical consistency.
4. *Support of reasoning* which can help to identify inconsistencies in an ontology-based documentation [Xi08]

Figure 3 Logical consistency aspects

[38] An example for such an extensive documentation of a meta-data model can be found in the library of congress classification system on http://www.loc.gov/catdir/cpso/lcco/

[39] For example the version and recent changes to Marc21 can be found on http://www.loc.gov/marc/bibliographic/.

[40] For example the library of congress offers an online tutorial on how to use and interpret their classification system on http://www.loc.gov/catdir/cpso/classwebtutorial/1intro.html

[41] For example http://www.w3.org/2004/02/skos/tools offers an overview to dedicated tools to create meta-data in the SKOS format to support consistency.

[42] An example is the multi-facet classification system proposed by Fugman [Fu89] in which index terms are related to specific activities, e.g. Knife (Tool) vs. Knife (Weapon).

2.4 Selection and overview of evaluated systems

To get a better understanding of current meta-data for ontology mappings it is helpful to analyse: (i) existing representation for ontology mappings and (ii) match candidate generation tools in the way that they support specific meta-data fields. This classification is based on a proposal provided in [Eh07].

One of the most popular tools for match candidate generation is the Ontology Alignment API [Eu11a]. This tool supports the export of alignments into the following representations: Ontology Alignment Format (INRIA), Expressive and Declarative Ontology Alignment Language (EDOAL), OWL (Version 1 and 2) [Br05] and Simple Knowledge Organisation System (SKOS) [Mi09]. Other tools support those export formats as well, for example [Ha08]. Thus those representation vocabularies will be considered as popular and be evaluated here.

The selection of match candidate generation tools is based on the Ontology Alignment Evaluation Initiative (OAEI) [Eu11a, Eu11b] that organizes annual campaigns to evaluate matching methods. Similar to the selection process used in [Sh12], the author of this thesis selected three tools that have repeatedly participated in the last three OAEI campaigns and have archival publications. In addition, two well-known matching systems are evaluated which focus more on matching mapping infrastructure [Eh07]: Ontology Alignment API [Je11] and NeOn [Pe08]. Please note these tools were selected as they provide sophisticated matching functions, have a considerable publication history and are considered important matching tools in current publications [Sh12]. These two tools focus more on infrastructure functions rather than a particular matching algorithm and can as such provide a view on meta-data support outside the limited scope of the OAEI.

Table 5 lists the ontology mapping representations and tools, the five evaluation metrics used (described in section 2.3) and the evaluation results. Please note the details and rationale for the individual scores listed are discussed for the individual languages or systems in the following sections 2.5 to 2.6.

Table 5 Meta-data model evaluation overview

Target	Number of supported meta-data fields	Level of relationship between meta-data elements	Ontology mapping lifecycle completeness	Number of satisfied application consistency aspects	Number of satisfied logical consistency aspects
Ontology Mapping Representation Languages					
Ontology Alignment Format (INRIA) [Eu04].	23	Index	Identification, Matching	3 of 8	0 of 4
Expressive and Declarative Ontology Alignment Language (EDOAL) [Eu07]	23	Index Relations	Identification, Matching	3 of 8	0 of 4
OWL Version 1 [Br05].	8	Index	Identification, Management	5 of 8	2 of 4
OWL Version 2 [OW12]	8	Index Relations	Identification, Management	5 of 8	2 of 4
Simple Knowledge Organisation System (SKOS) [As06].	7	Index Relations	Management	4 of 8	0 of 4
Match Candidate Generation Tools					
Falcon 4.2 [Hu08]	4	Index Relations	Identification	3 of 8	0 of 4
Anchor-Flood [Ha09]	4	Index Relations	Identification	3 of 8	0 of 4
AgreementMaker [Cr09]	6	Index Relations	Identification Matching	3 of 8	0 of 4
Ontology Alignment API [Je11]	23	Index Relations	Identification, Matching	3 of 8	0 of 4
NeOn toolkit [Pe08].	5	Index Relations	Identification, Matching	3 of 8	0 of 4

Each evaluated tool or representation is discussed in more detail in the following sections. This table shows that INRIA and EDOAL enables the most detailed documentation of ontology mappings. Both offer the highest number of meta-data fields (23) and cover two stages of the lifecycle. OWL offers the most support for application consistency and it is the only item which offers some support for logical consistency.

2.5 Meta-data support in Ontology Mapping Representation Languages

In this section the representation languages are evaluated with regards to the meta-data fields that they offer to document ontology mappings. Firstly SKOS [Mi09] and OWL1 are presented which are common Semantic Web representations and can be used to represent mappings. Secondly those representations which were explicitly designed to document ontology matching and ontology mappings are presented, that is the Ontology

Alignment Format (INRIA) [Eh04] and EDOAL [Eh07]. Note that non-standard extensions of OWL are not considered here.[43]

2.5.1 Simple Knowledge Organisation System (SKOS) [Mi09]

The Simple Knowledge Organization System (SKOS) is a common meta-data model for sharing and linking knowledge organization systems in the Semantic Web. It refers to a formal language developed within the World Wide Web Consortium (W3C) designed for representation of thesauri, classification schemes, taxonomies and other controlled vocabularies. SKOS defines semantic relationships that can be used to express a variety of mapping relations between concepts, in particular equal, broader or narrower. [As06]. The availability of this range of relation concepts has the advantage of being very general but offer only limited formal semantics [As06].

SKOS offers seven dedicated meta-data concepts for documentation which can be used to provide details about the mapping relations, for example *skos:editorialNote*. The meta-data coverage of the life-cycle is low as only the ontology mapping management stage is covered. More specifically, SKOS is able to document changes to a mapping over time, for example *skos:changeNote* or *skos:historyNote*. There are no typed relationships defined in the SKOS specification between these fields and therefore the level of relationship is considered as index-based. An explicit documentation of each meta-data field is available.[44] Meta-data fields have a unique identifier using the namespace in the RDF(S) version. The W3C offers an overview of the various editing tools available for SKOS.[45] So far no automated tool for population of SKOS meta-data for ontology mapping is available. SKOS is expressed in RDF(S) and does not contain any relations that would support reasoning such as OWL DL. W3C offers a consistency check for SKOS files[46] but it checks only for compliance with the SKOS technical specification and does not support any logical consistency checks between field content selections. Overall the meta-data model with regard to ontology mapping documentation can be

[43] The reason for this decision is that the focus is placed on well-known and standardized meta-data support, which is currently available. If the meta-data support is provided in the standard it can be assumed it is well known and supported by most tools.
[44] For further details see http://www.w3.org/TR/skos-reference/#notes.
[45] An overview of SKOS tools can be found here http://www.w3.org/2001/sw/wiki/SKOS/LegacyTools
[46] See http://www.w3.org/2001/sw/wiki/SKOS_Consistency_Checker for more details.

considered basic. The overall reason for this naturally, is that SKOS is not designed to represent ontology mappings explicitly and therefore lacks specialised fields for this scope.

2.5.2 Ontology Web Language Version 1 (OWL) [Mc04]

OWL refers to a set of knowledge representation languages that are standardized by the W3C. OWL is available in two versions. Version 1 of OWL provides an extended vocabulary based on RDF and is composed of three sub languages: OWL Light, OWL DL and OWL Full [Br05]. Each sub-/language provides a different level of semantic expressiveness. The semantics of OWL Light and OWL DL are based on Description Logic [Br05] which provides automated reasoning support. In contrast OWL Full is less restricted and provides a flexible semantic model intended to provide compatibility with RDF Schema. Commonly OWL ontologies are exchanged using the RDF/XML syntax but other formats are available, for example, Turtle or N3. OWL can be used to express correspondences between ontologies on a basic level. In OWL the relations *owl:equivalentClass*, *owl:equivalentProperty* and *owl:sameAs* can be used to express that different elements refer to the same concept [Sh06a]. Based on mappings expressed in OWL co-references between different ontologies can be found in the web of linked data, for example [Gl09] provides an example of such a service.[47]

According to the standard, OWL defines four properties related to versioning information[48]. In addition, the four generic annotation properties of RDF(S) are supported, that is *rdfs:label*, *rdfs:comment*, *rdfs:seeAlso*, *rdfs:isDefinedBy*. These fields can be used to document the identification phase and is also relevant information for the management phase. It is worth mentioning that other meta-data models such as Dublin Core properties can be used to annotate properties in OWL. As those are non-standard extensions of OWL, they are not considered here.

The meta-data fields supported in OWL are documented and examples are available. Each element has its unique identifier and the OWL model has a version. Various editors

[47] A novel implementation of this approach can be found on http://www.sameAs.org.

[48] The concepts are: versionInfo, priorVersion, backwardCompatibleWith, incompatibleWith, DeprecatedClass, DeprecatedProperty

for OWL and the related meta-data fields are available, for example Protégé[49]. However, as the fields are generic, no content options are given. The sub language OWL DL supports automated reasoning which can be useful for logical consistency checks. In addition, the specification defines the standard property *owl:incompatibleWith* which can be used to document incompatible content options. However, the specification does not provide any example of incompatible relations as it provides only the vocabulary not the content.

2.5.3 Web Ontology Language Version 2 (OWL2) [Wa12]

The OWL2 Web Ontology Language[50] inherits the language features, design decisions, and use-cases of OWL 1, but adds several new features including increased expressive power for properties, and extended support for data types.

In relation to meta-data, OWL2 [OW12] offers the same meta-data properties related to versioning information as OWL1. However, OWL2 extends the ability of the language to model meta-data. The first improvement relates to relationships. OWL1 DL requires a separation between classes and individuals. OWL2 has removed this restriction and so allows different uses of the same term. This process is called "punning" and as a result the same term can denote a class and individual.[51] This allows an ontology engineer to assign the meta-data properties both to instances and classes. The second improvement is that OWL2 allows annotations, such as a label, to be given to axioms. As such OWL2 allows for annotations upon ontologies, entities, anonymous individuals, axioms, and annotations themselves.

With these improvements in OWL2 although it is possible to document relations between meta-data fields more flexibly, no additional meta-data fields are defined. Therefore the suitability of OWL1 and OWL2 for documenting ontology mapping is similar, in that both languages lack specific meta-data fields designed to document the matching and mapping lifecycle related activities. Please note OWL1 and OWL2 satisfies the same aspects for application and consistency support. As such OWL is the only analysed

---

[49] The latest version of this editor can be found on http://protege.stanford.edu/

[50] The latest version of the format can be found on: http://www.w3.org/TR/owl2-overview/

[51] See http://www.w3.org/2007/OWL/wiki/Punning for more details.

language that provides any help regarding logical consistency. More specifically, OWL1 DL and OWL2 provide the ability to document incompatible relations and support reasoning, both vital to avoid inconsistent documentation.

2.5.4 Ontology Alignment Format (INRIA) [Eu04, Eh07]

The Ontology Alignment Format[52] (also referred to as INRIA format) was developed by Euzenat and his research group. This representation was designed specifically to express ontology candidate matches to support their transformations, sharing, retrieval and comparisons. The format is based on RDF(S) [Eu04, Eu06] and a range of different mapping correspondences can be expressed, for example subsumed relation, equivalence, subclass etc. [Eu04].

Since 2004 the Alignment format [EU04] is the recommended format to be used in the Ontology Alignment Evaluation Initiative (OAEI) [Eu11a]. In addition, this format is used by many ontology candidate match generation systems [Me10, Sh12].[53] It can be noted that currently no standard for the representation of ontology candidate matches or mappings has evolved. However, the Ontology Alignment Format is used widespread and as Euzenat pointed out it can be seen as "a natural starting point for standardization" [Sh12].

Overall the format supports 23 meta-data fields. In annex A a full list of all standard meta-data fields is shown. Please find below in figure 4 an example of the Ontology Alignment Format as an RDF document.

[52] The latest version of the format can be found on: http://alignapi.gforge.inria.fr/format.html
[53] More system who support INRIA can be found on http://alignapi.gforge.inria.fr/impl.html

```xml
<?xml version='1.0' encoding='utf-8' standalone='no'?>
<rdf:RDF
xmlns='http://knowledgeweb.semanticweb.org/heterogeneity/alignment#'
xmlns:rdf='http://www.w3.org/1999/02/22-rdf-syntax-ns#'
xmlns:xsd='http://www.w3.org/2001/XMLSchema#'
xmlns:align='http://knowledgeweb.semanticweb.org/heterogeneity/alignment#'>
<Alignment>
<xml>yes</xml>
<level>0</level>
<type>**</type>
<align:method>fr.inrialpes.exmo.align.impl.method.StringDistAlignment</align:method>
<align:time>7</align:time>
<onto1>... </onto1> <onto2>... </onto2> ...
</Alignment> </rdf:RDF>
```

Figure 4 Ontology Alignment Format example

The available meta-data fields focus on the ontology identification, for example indicating the file location of the target ontology using *align:URI1* and *align:onto1*. In addition, some basic descriptive information is provided as well, for example *align:language* of the addressed ontologies. Furthermore, details about the applied automated matching algorithm can be documented, for example using *align:method* or *align:time*. The relationship type is index-based as each meta-data field contains a simple string value. Please note the format allows other non-standard annotations to be included if they are identified by a unique URI, for example mapping purpose or processing time. An overview of all the meta-data fields, including a short bullet point definition, is available online.[54] Not one of the other application consistency aspects is satisfied with regards to the offered meta-data support. With regards to logical consistency aspects neither the meta-data model nor the documentation offers any details on compatible field combinations. As the format is based on RDF(S) no reasoning is supported.

2.5.5 Expressive and Declarative Ontology Alignment Language (EDOAL) [Br04, Br12]

The Expressive and Declarative Ontology Alignment Language (EDOAL)[55] extends the Ontology Alignment Format. This format is suitable to capture equivalence or subsumption correspondences. EDOAL was created to allow more precise but also more

[54] See http://alignapi.gforge.inria.fr/labels.html for details

[55] The language was originally created by François Scharffe as the SEKT Mapping language (http://www.sekt-project.com/). It later been published as OMWG Ontology Mapping language (see http://sourceforge.net/projects/mediation/) before it was integrated in the Ontology alignment API project.

complex alignments to be expressed between ontologies [Eu07], using a minimal vocabulary on top of the original Ontology Alignment Format.[56]

The main strength of EDOAL in comparison to the INRIA format is the ability to group named entities found in ontologies (e.g. classes and properties) with Boolean operators (e.g. disjunction) or construction operators (e.g. reflexive) and constraints (e.g. range, cardinality). In addition, in EDOAL transformation of property values can be specified.

EDOAL is based on RDF(S) but was inspired by OWL as similar features are used [Da10]. However, EDOAL is not created as an OWL ontology[57] but it can be exported into an OWL language by simplifying correspondence relations, for example using *owl:subClassOf.* Thus EDOAL does not support reasoning. EDOAL does not define any additional meta-data elements for documentation. However, it supports the full list of meta-data fields available for the INRIA alignment format. Therefore EDOAL satisfies the same application and logical consistency aspects such as the alignment format.

2.5.6 Summary of evaluation of ontology mapping representations

In summary, from table 5 in section 2.4 it can be seen that of all the representations, the INRIA format and EDOAL enable the most detailed documentation of ontology mappings in that both offer the highest number of fields (23) and they cover two stages of the lifecycle. However, both representation languages do not offer any support for creating consistent meta-data.

OWL 1 and 2 offers only 8 meta-data fields which are focussed on the identification and the management phases of the lifecycle but no specific matching or mapping details are supported. However, OWL offers the most support for application consistency and is the only representation that offers some support for logical consistency.

---

[56] The latest version of the language is available on http://alignapi.gforge.inria.fr/edoal.html

[57] "EDOAL is less than OWL: it does not allow for defining new named entities, which is arguably one of the main features of OWL." [Da10]

It can be concluded that none of the analysed representation languages can offer the documentation characteristics needed to support consistent creation and reuse of mapping meta-data in support of mapping reuse.

2.6 Meta-data support in Match Candidate Generation Tools

In this section five ontology matching related tools are presented with a focus on the individual meta-data they support to document generated ontology mappings.

2.6.1 Falcon-AO

Falcon-AO[58] is an automatic ontology matching system which supports matching of ontologies expressed in RDF(S) and OWL [Fe07, Hu08]. Falcon-AO is implemented in Java as an open source project. Falcon-AO provides a simple graphical interface but no graph based visualization of matching results [Hu08]. Falcon supports several matching algorithms including a linguistic and a graph based approach [Hu08].

The matching results can be exported in the INRIA alignment format. However, only 4 standard Ontology Alignment API fields[59] for the Ontology Alignment Format are supported. Please note there are no limits of meta-data fields for the OAEI challenge, so the choice to support only 4 meta-data fields was made by the tool developer. This relates to a source of the ontology and a URI identifier. However, no other lifecycle steps are documented. The tool itself does not offer any further meta-data or functions to support application and logical consistency. As the alignment format is limited to the used meta-data fields, the level of relationship is considered as index-based.

2.6.2 Anchor-Flood

The Anchor-Flood[60] system was developed by Hanif Seddiqui [Ha09]. It is a matching creation tool designed to handle particularly large ontologies efficiently. It allows processing of an ontology in RDF(S) and OWL to generate one to one alignments. The

[58] The latest version of the tool can be found on http://ws.nju.edu.cn/falcon-ao/index.jsp, 2012.
[59] See http://alignapi.gforge.inria.fr/labels.html for a full list
[60] The latest version of the tool can be found on http://www.kde.cs.tut.ac.jp/~hanif/

system starts with exactly matched concepts (so called anchors) and then analysing their neighbours to build segments out of the ontologies to be matched. A key differentiator is that the correspondences are based on terminological (WordNet and Winkler-based string metrics) and structural similarity measures. The meta-data support is identical to Falcon-AO as only 4 of the standard Ontology Alignment API fields[61] are supported, for example ontology label and a URI identifier. No other meta-data is available. No further support for application and logical consistency is provided.

2.6.3 AgreementMaker

AgreementMaker[62] is a matching system developed at the Advances in Information Systems Laboratory at the University of Illinois [Cr09]. The aim of this system is to provide an easy to use graphical interface to allow a domain expert to create alignment of ontologies manually and efficiently. In addition, it allows the execution of a number of automatic matching algorithms which are syntactic, structure and instance based. The system handles ontologies in XML, RDFS, OWL, N3 and outputs 1:1, 1:m, n:1.

The point of difference to Falcon-AO and Anchor-Flood is the graphical interface. Each ontology class is shown as a node with a different colour. This provides users with information about activities (e.g. selection) and differentiation between the alignments. Figure 5 shows a screen shot of the tool interface:

[61] See http://alignapi.gforge.inria.fr/labels.html for a full list
[62] The latest version can be found on http://agreementmaker.org/

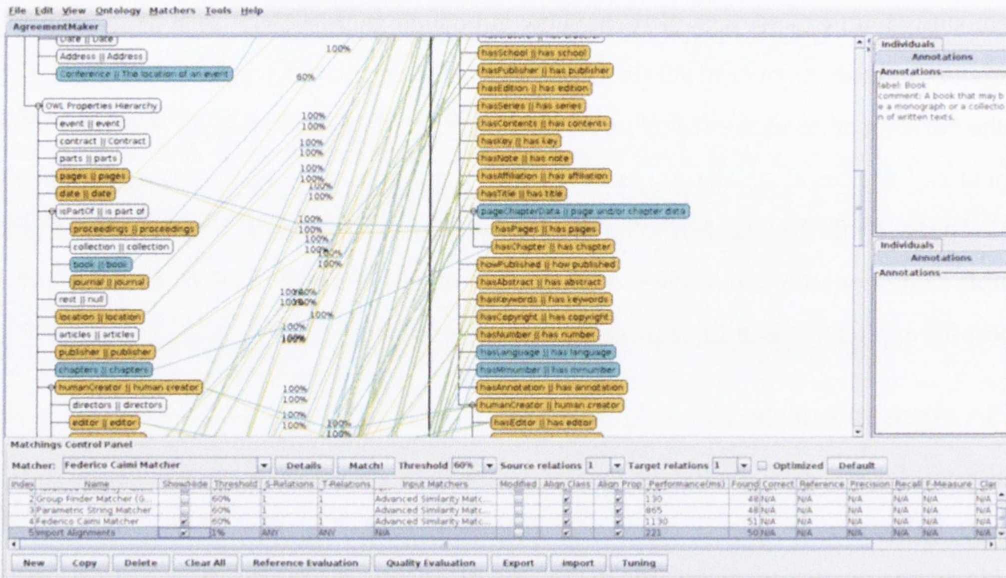

Figure 5 Screenshot of the Matching Tool AgreementMaker

This interface makes it easy for the ontology engineer to annotate mappings with meta-data to store the similarity value and a short name of the matching algorithm. In addition, a field for a comment and a label are supported. The resulting alignments can be exported in the INRIA alignment format but only 6 basic meta-data fields are supported, that is ontology label and a URI identifier for target and source ontology. In other words the interface fields for comment and label of the algorithm can be entered in the tool but not exported anywhere. Please note the meta-data fields which can be entered in the interface can only be exported in a tool specific proprietary XML notation. Thus only four meta-data fields can be exchanged in a notation to describe the mapping and no support for application and logical consistency is provided. In short the tool is focussed on the actual matching generation but not on the documentation process.

2.6.4 INRIA Ontology Alignment API

The INRIA Ontology Alignment API[63] is a collection of Java libraries and scripts for expressing and sharing ontology alignments [Je11,Eu05]. It offers the following services: storage, retrieval and sharing of alignments; manipulation of alignments; generation of

[63] The latest version 4.5 of the API can be found on: http://alignapi.gforge.inria.fr/

alignments. It is open source software and can be extended freely.[64] The API enables the creation of alignments based on different matching algorithms including terminological and structural version [Eu08]. Overall, the Ontology Alignment API is one of the most sophisticated toolsets for management of alignments and matching algorithms. It is well documented[65] and together with the Ontology Alignment Format it provides an excellent infrastructure for users. This is confirmed by the high number of tools who use or integrate with this API.[66]

With regards to meta-data the API supports all fields of the Ontology Alignment Format. Thus the same application and logical consistency aspects are satisfied. According to the developer, the API allows other applications to attach information that are not part of the supplied meta-data model to the alignments. Those are rendered and stored in the Alignment Server database but only if those annotations contain string values and so in accordance with our analysis it can be considered that an index-based relationship approach is supported. A list of all currently known meta-data extension fields can be found here.[67] Please note the API supports the creation of scripts for consistency tests but those are currently limited to alignment comparisons and structural verification. However, no check for logical inconsistencies is possible as such information is not stored in the API or the alignment format.

2.6.5 NeOn toolkit

The NeOn toolkit[68] is an ontology engineering environment developed in the NeOn Project [Pe08]. The toolkit offers an open source platform with the objective to support the completed ontology engineering life-cycle. It is based on the Eclipse platform and provides an extensive set of plug-ins covering various activities (currently 45).[69] In this context the Alignment plug-in[70] is most relevant as it allows the automatic generation and

[64] An overview of system who use or extend this API can be found here: http://alignapi.gforge.inria.fr/impl.html
[65] Please see http://alignapi.gforge.inria.fr/tutorial/ for an overview of the documentation.
[66] The following page provides an overview of all systems who use the API http://alignapi.gforge.inria.fr/impl.html
[67] Please see http://alignapi.gforge.inria.fr/impl.html for details.
[68] The latest version of NeOn can be found here: http://neon-toolkit.org/wiki/Main_Page
[69] An overview of all plug-ins can be found here http://neon-toolkit.org/wiki/Neon_Plugins
[70] The plug-ins can be found here http://neon-toolkit.org/wiki/2.3.1/Alignment

management of ontology alignments. For this purpose the INRIA Ontology Alignment API is used.

The alignment plugin allows matching of ontologies, trim alignments by applying thresholds to existing alignments retrieve and render alignment in a particular format and store an alignment permanently on the server. The main point of difference to the INRIA alignment API is that all these activities can be performed in a graphical interface.[71] The Ontology Alignment API offers only a command interface.

With regards to meta-data, the NeOn toolkit allows the export of alignments into the alignment format. Only the standard API fields to identify the ontology path and URI are supported as well as the used algorithm identified by a label. However, the tool does not allow these fields or the format to be edited or to be extended.

2.6.6 Summary

These example matching tools show that the support of mapping meta-data fields is very low compared to the actual meta-data fields supported by the mapping representation language that they state support for, for example Falcon-AO supports only 4 out of 23 meta-data fields available in the INRIA alignment format. This highlights that the tools are designed for match candidate creation but not for mapping documentation. Only the AgreementMaker offers an interface to edit meta-data but is again unable to export all fields in a common mapping representation language such as the INRIA alignment format. The NeOn toolkit provides a similar limited meta-data support as only the standard INRIA fields are supported. The Ontology Alignment API is created by the same people such as the alignment notation and naturally all proposed meta-data fields are supported. However, the overall meta-data support is still limited as meta-data fields contain string values and the support for application (3 out of 8) and logical consistency (0 out of 4) is still very low.

[71] Please refer to the following link for an interactive webcast demo http://www.neon-project.org/nw/Movie:_Alignment

2.7 Conclusions

This study shows that the majority of tools and languages for matching and mappings allows for only an incomplete and limited documentation of the ontology mapping lifecycle. More specifically, only the identification and matching phases are addressed, e.g. URI and location of the addressed ontology. The Ontology Alignment Format offers the highest amount of meta-data fields (23) but is limited to the matching and identification stage of the ontology mapping lifecycle. Only the SKOS system offers four meta-data fields focussed on versioning which is relevant for management aspects. This general lack of lifecycle related meta-data represents a challenge for the management and reuse of mappings.

The next focus was placed on expressiveness of the supported meta-data. It can be noted that all tools and languages use an index-based approach where meta-data is stored in single key value fields as strings but not as explicit entities with a defined relationship between the fields.[72] In the opinion of the author this is problematic as most meta-data elements have a complex knowledge structure and explicit relations can be helpful for retrieval, for example where an algorithm has a particular implementation that is linked to a specific matching tool.

Furthermore, the evaluation showed that only a small number of application consistency aspects were satisfied by the representations, matching tools and infrastructure tools, for example the INRIA representation satisfies 3 out of 8 aspects. In general it can be observed that there is a lack of detailed documentation and comprehensive examples of annotated representations. This is particularly challenging if one wants to enable correct and consistent use of meta-data and interpretation of the intended meaning of the meta-data. No versioning information for the meta-data models is offered and no documentation is embedded directly in the models. Only for OWL is an explicit editor available which can help maintain meta-data and support consistent documentation. However, even the benefit of this is limited as the provided meta-data fields in OWL are very general and not focussed on the specific details of matching and mapping process,

[72] Only OWL2 is able to express explicit relations between meta-data annotations.

for example no explicit fields to define the algorithm used to create a candidate match. Not one of the languages provides meta-data content options. Overall this supports the conclusion that meta-data model support for the generation of application consistent documentation is low currently. This makes retrieval and reuse of mappings difficult due to the likely inconsistent documentation.

Not one of the stated aspects for meta-data model support for generating logically consistent documentation is satisfied by the tools. Only the OWL language can offer a property relation for compatible meta-data fields. In addition, only OWL DL supports reasoning which can help to deduce inconsistent statements automatically. The downside is that OWL is a very generic language and not designed to address the specific needs of ontology mapping representations or documentation. Again the gathered information supports the conclusion that current tools and formats for ontology mappings offer very little meta-data model support in regards to logical consistency.

Overall, it can be noted that no standardised ontology mapping representation has yet emerged or is generally accepted [Sh12]. However, the INRIA Ontology Alignment Format is used as a de facto standard as it is recommended by the OAEI [Eu07]. Support for documentation of ontology mapping lifecycle is fragmentary. Support for the creation of consistent documentation is low. This makes it difficult for an ontology engineer to reconstruct how a particular mapping was created and/or to create consistent documentation himself. In other words the sharing of a common understanding of the ontology mapping creation and reuse process is not supported sufficiently by state of the art meta-data models.

In the following chapter the result of this study will be used to identify requirements for the OM2R model. Together with the result of a domain analysis they will guide the design of the OM2R model to enable the support of relevant and consistent meta-data documentation for ontology mappings.

3 Design of the OM2R model

This chapter presents how the requirements for the OM2R model were derived based on a literature review on the common challenges for meta-data models, the result of the state of the art study in chapter 2 and a domain analysis. Overall four requirements were identified which resulted in six design features of the OM2R model to support consistency and relevance. The chapter also includes an overview of the evolution of the OM2R model.

3.1 Methodology

The design of the OM2R model is based on a requirement analysis where a list of individual requirements is identified and then satisfied with a specific design feature.

The basis for the requirement analysis in this thesis is a literature review on common meta-data challenges (see section 3.2). In this section relevance and consistency are derived as key quality aspects for the OM2R model. Based on the identified challenges the library domain was examined as a common domain for meta-data use. The objective was to identify how relevance and consistency is commonly supported in meta-data models.

In addition, the insights gained from the state of the art study (see section 3.3) are used to identify requirements for the OM2R model, in particular in relation how meta-data for mappings are supported currently.

Furthermore, a domain analysis (see section 3.4) is conducted which highlights the ontology-mapping retrieval focussed use-case of the OM2R model. More specifically, the impact of the extended ontology mapping lifecycle on mapping retrieval and as such on the requirements of the model is discussed.

Figure 6 provides an overview of the activities involved in the requirement analysis and shows the identified requirements (R) and the corresponding design features (D) of the OM2R model.

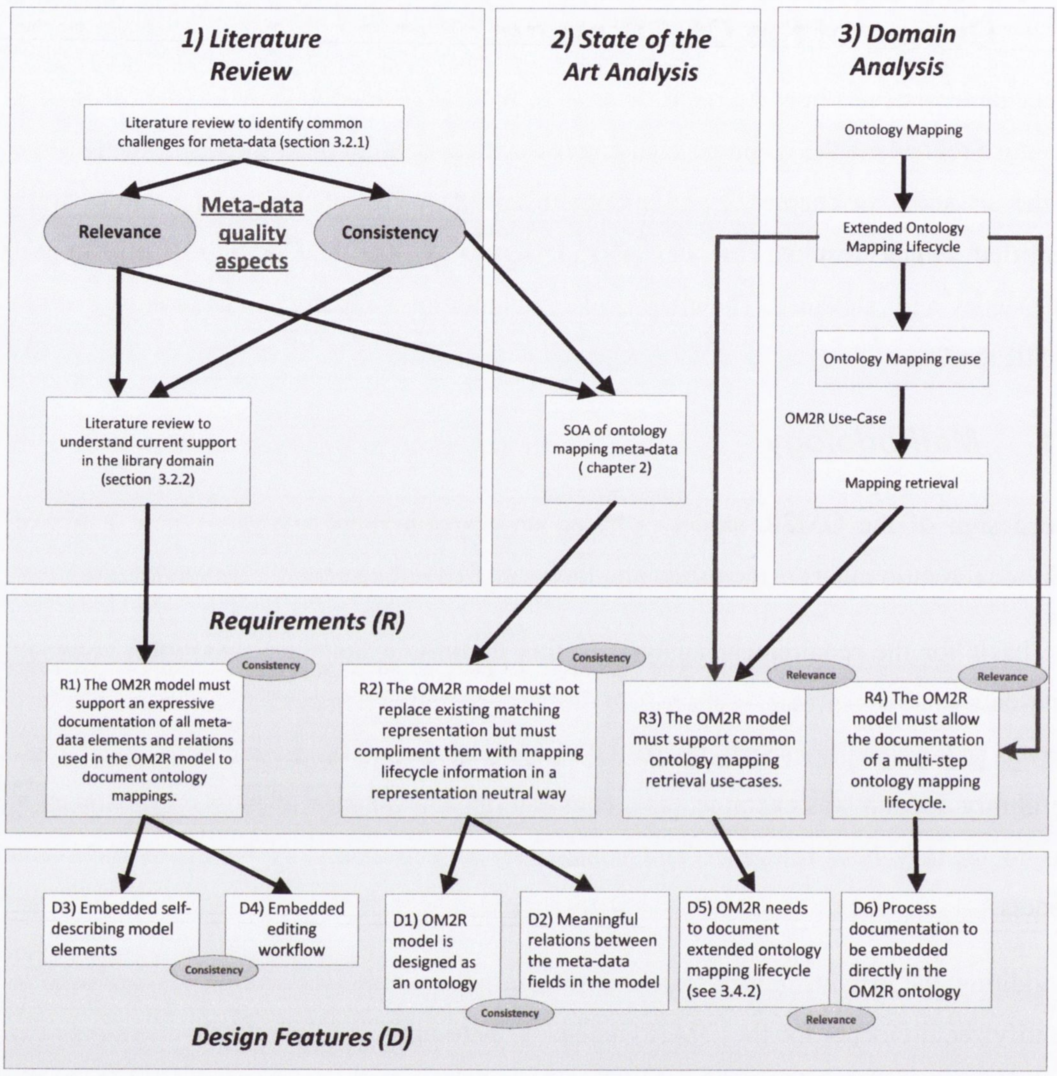

Figure 6 Overview of activities to identify design requirements and resulting features of the OM2R model

3.2 Common challenges for meta-data models

In this section, the library domain has been chosen in order to identify common challenges for meta-data model creation and usage. This is because the library domain has had the most extensive focus on meta-data design and usage over the years and it relies heavily on the creation of consistent and relevant meta-data. In other words, of all domains, the Library domain is where the most emphasis and effort has been placed on developing concepts and tools around meta-data.

3.2.1 Quality aspects for meta-data models

R. Niccole Westbrook has investigated meta-data model usage in the library domain and found that the challenges for meta-data models regarding quality aspects are quite diverse [We12].

A key challenge for any meta-data model is the time-consuming creation which depends on the complexity of the chosen model [Ed07, Te11]. This is closely related to the ability of the indexer to provide an accurate and complete meta-data representation of the information object [Fu89] by using the given model. For example, the Learning Object Model (LOM) allows a very detailed description of learning resources but has been criticised for its complexity and difficulty of use [Na04]. The **relevance** of the meta-data model for the application purpose is vital to mitigate these creation difficulties. The more relevant and therefore useful a model, the more likely it is that a user is willing to invest effort to document the meta-data [Fu93, Du02].

A meta-model should support interoperability and as such meta-data can only be used to its full potential if it is created **consistently** [Du10, IS04]. For example, data creation by users who are not familiar with the meta-data model can lead to quality problems, e.g. missing mandatory elements or incorrect use of fields [Pe06, Ma99]. Particularly problematic is the inconsistent use of a meta-model vocabulary. It can lead to incorrect or contradictory field combinations, which can make it difficult to find relevant information [Ni04]. To address the need for consistency, a meta-data model should include a "clear statement on the conditions and terms of use" [Ni09]. This can help reduce inconsistencies as it makes the intended meaning clear for the creator and future consumer [Pe06].

This brings up the question as to how relevance and consistency being vital quality aspects for meta-data models are supported in current systems [We12]? The following section presents an evaluation on how these two aspects are supported in the library domain. This will provide insights relevant for the design of the meta-data model for ontology mapping documentation.

3.2.2 Current approaches to support meta-data model quality

The quality of meta-data has been a vital topic [Ca03] for libraries for a long time and has become an important focus for digital libraries, too [Ya11]. The library domain was selected for this analysis in this section as it is one of the most common places where meta-data is widely used and created. Particularly attractive is the focus on long term use of the meta-data by different users which is similar to the context of ontology mapping reuse. The database community has also a strong focus on meta-data [Be11c] but they are more implementation driven and compared to libraries more short lived.

In libraries consistency and relevance are primarily addressed by the use of so-called controlled vocabularies [Ra04] where the indexers need to choose their documentation from a given meta-data vocabulary. In other words, controlled vocabularies define meta-data fields, structures and their content options. As such they can be seen as a meta-data model. One of the well-known examples of such a controlled vocabulary is the Library of Congress Classification [Bi00].[73] Combined with training and best practices the controlled vocabulary allows an indexer to create an accurate and consistent meta-data description. This is vital for users to anticipate meta-data fields and the specific keywords used to describe the relevant topics.

The following subsections discuss in more detail the methods which are used in the library domain to support consistency and relevance of meta-data.

3.2.2.1 Support for relevance of meta-data

Stvilia [St08] highlighted the need for maintenance of meta-data models. New developments need to be considered and controlled vocabularies altered or extended. For example the term AIDS for the "acquired immune deficiency syndrome" was first invented in the late 1970's. Previously no specific name was available, only a plurality of rather fuzzy descriptions in medical reports. The medical classification in libraries needed

[73] http://www.loc.gov/catdir/cpso/lcco/

to reflect these new developments and as such include the term in their classification and train indexers in their usage.

The Library of Congress Classification is a good example of how the relevance of the meta-data model can be improved with suitable management processes over time. More specifically, the classification "is the result of the combined efforts of cataloguers, editors, and policy specialists". Cataloguers propose new or changed class numbers when required by new material being catalogued, and formulate major developments as time permits. Editorial teams accumulate the new proposals, determine the exact wording of anchor points, annotation instructions, and index entries, and produce a list for review by policy specialists and other staff at the weekly editorial meeting."[74] The developments are published quarterly.[75]

Beside such process driven efforts, relevance is supported by the development of suitable standards that help to evolve and track change. For example PREMIS is an international standard that offers a data dictionary for preservation meta-data to support the preservation of digital objects and ensure their long-term usability.[76]

3.2.2.2 Support for consistency of meta-data

R. Niccole Westbrook in [We12] defined consistency in the library domain as two concepts: semantic consistency and structural consistency. Semantic consistency is the "extend to which the collections use the same values (vocabulary control) and elements for conveying the same concepts and meanings throughout" [Sh05]. In contrast structural consistency "is the extent to which similar elements of a meta-data record are represented with the same structure and format" [Sh05].

In the library community, semantic consistency is focussed around the indexer as meta-data creation is seen as a manual activity [Ra04]. The key for consistency is suitable training and the ability of the indexer to be familiar with the complete controlled vocabulary and its meaning. This is vital for him to be able to choose the most

[74] http://www.itsmarc.com/crs/mergedProjects/scmclass/scmclass/ historical_note_on_the_library_of_congress_classification_scm.htm
[75] http://www.loc.gov/cds/products/product.php?productID=39
[76] http://www.loc.gov/standards/premis/

appropriate descriptive meta-data for a particular resource [Fu89]. This depends on suitable documentation of the meta-data and easy ways to navigate the vocabulary for the indexer.

In addition, semantic consistency can be supported by promoting automatic extraction of meta-data from digital resources, automatic classification and entity identification [De12]. These need to be combined with the manual "professional" approach, e.g. author names extracted from resources are matched against authority knowledge bases.

Structural consistency is particularly vital for digital libraries as they often combine meta-data from different libraries which can use different formats and standards to represent their meta-data [Ra04]. To address this challenge languages and standards are proposed which help to provide a uniform representation of meta-data to make exchange and integration easier. Semantic Web based formats are popular as they provide an expressive and implementation independent representation, e.g. MARC (MAchine-Readable Cataloging)[77] which is used to align the U.S. and Canadian library meta-data formats.

3.2.3 Requirements for OM2R based on the challenges for meta-data models

Section 3.2.2 has shown that for the creation of consistent meta-data, a user needs to have a detailed understanding of the meta-data model. Thus, a detailed documentation of the meta-data model and its application is vital to allow the user to understand the intended meaning of the meta-data fields, of the model structure and their optimal usage. This is particularly essential for the creation of consistent models over time, which leads to the first requirement.

R1) The OM2R model must support an expressive documentation of all meta-data elements and relations used in the OM2R model to document ontology mappings.

In other words the OM2R model must support the sharing of a common understanding about the meta-data fields, their intended application and meaning.

[77] http://www.loc.gov/marc/bibliographic/

In the following section the results of the state of the art review (see chapter 2) on the current meta-data support in mapping language and application is discussed in order to identify further design requirements.

3.3 Requirements for the OM2R model based on the State of the Art Review

The state of the art in section 2.7 has revealed that no standard format for representing ontology matching and mapping has emerged yet [Be11]. One possible reason for this lies in the multitude of tools and approaches for mapping/matching available [Sh12]. Each one has unique features which a generic format cannot document in all cases.[78] However, if the documentation focus is shifted away from a specific implementation towards the needed activities in the ontology mapping lifecycle, then it is necessary for meta-data to be more uniform. For example, what are the addressed ontologies and what is the process used in the identification of match candidates? In addition, from a practical perspective it is difficult to establish a new mapping format standard as a wide range of tools would need to be altered in order to generate and process this new format. This leads to the second requirement.

R2) The OM2R model must not replace existing matching representation formats but must compliment them with mapping lifecycle information in a mapping representation neutral way.

The author of this thesis argues that it is better to support flexible enrichment of existing and future ontology mapping representations in order to augment their usage, reuse and management [Th12]. This approach ensures relevance as the model can address new or extend existing relevant aspects covered in the current mapping representation.

Furthermore, the state of the art study has revealed a common lack of meta-data fields for mappings. This means that there is currently very little support for the creation of

[78] For example the matching tool AgreementMaker [Cr09] allows users to define an explicit label for the applied matching algorithm. However, the tool supports the popular INRIA format but this format is not able to document the custom algorithm label.

complex and consistent meta-data due to the lack of guidance for the user. The meta-data is intended to be shared between different users and be used in different application contexts, e.g. mapping reuse and retrieval. It is also likely that the editor and user of the meta-data will be different [Pe06]. For example in the Ontology Alignment Evaluation Initiative OAEI [Eu11a], meta-data that documents matches are created by participants in the competition but will be reviewed by external researchers. In addition, the use of the meta-data will change over time as more and more mapping applications and reuse scenarios are supported [St08]. Section 3.2.2.1 has shown that relevance of models can be improved by regular reviews of the models by domain experts. This implies that ontology engineers need to be involved closely in the development and evolution of the OM2R model to ensure its relevance for applications such as mapping reuse. This provides additional support for the first requirement R1 that demands that the OM2 model must support an expressive documentation of all meta-data elements and relations used in the OM2R model.

The two requirements R1 and R2, guided the high level design of the OM2R model with regards to the documentation and representation approach. However, this leads to the design question as what specific aspects of ontology mapping should be documented in the meta-data model? This question is addressed with the help of a domain analysis in the following section.

3.4 Domain analysis

This section presents an analysis of the application domain of the OM2R model with a focus on the extended ontology mapping lifecycle. It provides the basis for further requirement identification.

3.4.1 OM2R application use-case

Ontology mapping creation is a complex and time-consuming process as shown in section 2.7. As stated in section 2.2.1, ontology mapping is defined as a process of relating the vocabulary of two ontologies sharing a domain [Ka03]. Typically this process is conducted by ontology engineers as they have the necessary understanding of ontologies and the needed domain knowledge.

In the design of the OM2R model, it was decided to initially scope the motivating application to that of ontology mapping reuse. Ontology mapping reuse narrows down this definition to cases where the ontology mapping is not created for a particular integration task at hand. Instead relevant existing ontology mappings that were created by other ontology engineers, perhaps in another context, are reused.

Evaluating the mapping reuse activity is difficult as it is often affected by various external factors. As a result, this thesis will place emphasis upon ontology mapping retrieval which is the first phase of any mapping reuse attempt [Br12] (see section 3.4.2 for an extended mapping lifecycle). Users or automated systems need to find available mappings first before they can make a reuse decision. The rationale is that support for ontology mapping retrieval will directly support ontology mapping reuse. In addition, retrieval is one of the most common application areas for meta-data in general [Du02, Be05]. Based on this insight the following use-case for the application of the OM2R model has been defined.

> *An ontology engineer is confronted with the need to integrate various ontologies of a common topic. To achieve this objective the engineer is asked to find relevant existing ontology mappings based on the provided meta-data documented in the OM2R model.*

This use-case is displayed in figure 7 as a UML use-case diagram which highlights the activity areas and how they are connected. This includes the creation of the ontologies, the mapping between the ontologies and their meta-data documentation in the OM2R model. The meta-data is then used to support the retrieval of the actual mappings.

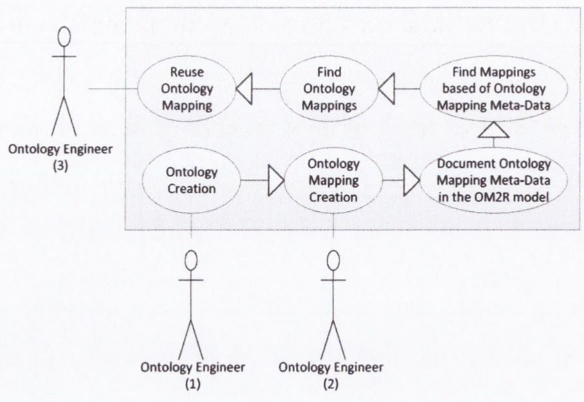

Figure 7 UML Use-Case Diagram for OM2R

The focus of the ontology engineer is on the integration of existing ontologies by reusing existing mapping. This use-case shows that the ontologies, mappings and meta-data were created by different stakeholders.

3.4.2 Extended Ontology Mapping Lifecycle for Ontology Mapping Reuse

Retrieval is based on meta-data that adequately represents the required information object [We12] which in this case is the ontology mapping and its creation lifecycle. The key to understand how a particular ontology mapping was created lies in the individual activities found in the ontology mapping lifecycle presented in section 2.2.2. In other words, the individual activities provide evidence of what details need to be represented in meta-data fields to allow a user to find mappings based on these details, e.g. what is the target ontology of the mapping.

This original lifecycle proposed by O'Sullivan [Su07] and shown in section 2.2.2 does not include the necessary activities for mapping retrieval, reuse and meta-data generation. To address this wider scope, the author of this thesis has extended the original lifecycle and more specifically, the management phase with activities relating to mapping reuse and meta-data creation (see figure 1 in section 2.2.2). These activities are described next.

5. *Mapping Reuse Phase:* If ontology mappings are to be shared between different parties or organisations, a reuse of a mapping in the same or a different context may be appropriate [Sh12]. These activities cover all the steps necessary to come to a

reasonable decision as to how to conduct reuse [Be08]. This involves first the *retrieval of existing ontology mappings*, e.g. between two specified ontologies or for a particular ontology element. In the next task each of the identified mappings needs to be analysed. This is based on the available meta-data which documents the individual creation lifecycle of the mapping and the requirements of the individual reuse scenario. Based on the results, a *reuse decision* can be made. Either a reuse is appropriate and the selected mapping can be executed directly in phase five or if not another iteration of the ontology mapping generation process needs to be started in phase one.

6. *Meta-Data Documentation Phase:* To support the reuse activities it is vital to document the relevant aspects of the ontology mapping. This is similar to the key function of traditional libraries to provide qualitative descriptive information towards information resources to assist the retrieval [Ab89, Ni09]. In this phase meta-data is collected and processed to document the individual aspects of the ontology mapping creation lifecycle. The meta-data can either be extracted automatically from the ontologies and mappings or manually entered by involved stakeholders. This task needs to be performed in parallel to all previously discussed phases. In addition, the collected meta-data can be exported in a suitable ontology-based representation in order to support the processing and interoperability of the ontology mappings meta-data.

3.4.3 Ontology Mapping Retrieval Use-Cases for the OM2R model

Ontology mapping retrieval is the essential activity for the mapping reuse. This leads to the question: *What are specific common ontology mapping retrieval use-cases?*

To answer this question the author of this thesis has identified 14 mapping discovery use-cases by analysing the activities found in the ontology mapping lifecycle (see section 3.4.2). The design of each use-case is based on two assumptions. First a mapping can be reached based on the characteristics of the addressed ontologies or the mapping itself, e.g. find a mapping between a specific source and target ontology. Second, a reuse of mappings can be influenced by the way the mapping was created. As a result, the individual activities involved in the ontology mapping lifecycle provide guidance to define common mapping retrieval criteria, e.g. an ontology engineer can be looking for mappings which were created by a specific algorithm. Based on these considerations 14 use-cases were derived and are shown in the table 6.

Table 6 Common Ontology Mapping Retrieval Use-Cases

#	Mapping Discovery Use-Case	Motivation
1. Characterization phase of the ontology mapping lifecycle		
1	Find mappings between specific source & target ontologies (1a) or ontology elements (1b)	To support information exchange between two specific knowledge models you need to find them first
2	Find mappings for a particular ontology (2a) or an ontology element (2b)	Identify interoperable ontologies which could be reused on a class level
3	Find mappings between a specific source ontology to any target ontology with specified characteristics, e.g. ontology language	Only specific ontologies can be reused, e.g. to support reasoning a formal language of OWL DL is required
4	Find mappings expressed in particular mapping format (notation)	Depending on individual application only a specific mapping notation maybe supported
5	Find ontology mappings for a specific correspondence types	Depending on reuse scenario simple equivalence mappings or more complex mappings (narrow, broader etc.) may be required
2. Matching and 3. Mapping phase of the ontology mapping lifecycle[79]		
6	Find mappings created either automated, manually or in a combination	Fundamental differences between automated and manual mapping creation exist
7	Find mappings created by a specific matching type	Many different matching approaches (algorithms) are available and this could be used as a search criteria
8	Find mappings created by a specific matching algorithm implementation	Helps to narrow search down to a specific implementation of a matching approach
9	Find mappings created by a given matching algorithm configuration	Depending on the applied parameters an algorithm can produce quite different matching results
10	Find automated created mappings based on matches with a high confidence level	Most automated matching applications provide a confidence level for filtering
11	Find manually created mappings depending on the involved users	Depending on skills and backgrounds different involved users can create different mapping, thus may not be suitable for reuse scenarios (e.g. product engineers vs. customers)
12	Find mappings created by a particular author	Trust only a specific author
13	Find mappings created for a particular context	The reason for mapping has a major impact on the creation process
4. Management phase of the ontology mapping lifecycle		
14	Find a specific version of a mapping	Mappings are evolving and need to be managed

[79] In this list no differences are made between matchings and mappings as both process steps are closely related. The main motivation was to keep the resulting validation questionnaire short and to avoid confusion on the side of the participants in regards to the used terminology.

The relevance of these proposed use-cases was verified with the help of an anonymous questionnaire [Sa83] in September 2009. The hypothesis for this mini experiment was defined as: *The proposed ontology mapping discovery use-cases are relevant for an ontology mapping reuse.*

For the methodology of the mini experiment, an approach was chosen where all participants were asked to rate each of the use-cases for their likely frequency of occurrence (possible answers: 2 - frequent, 1 - occasional, 0 - rare) and relevance to mapping applications (possible answers: 2 - highly relevant, 1 - medium relevant, 0 - irrelevant). Based on this data the average frequency of the likely occurrence and the average relevance was calculated. Please see annex C for a copy of the questionnaire form. The evaluation was conducted on paper and manually evaluated by using MS Excel.

Participants were invited from PhD students from the KDEG (Trinity College, Ireland) and industrial partner involved in the research project Federated Autonomic Management of End-to-End Communications Services" (FAME)[80]. These two target groups were chosen because they have a known background in ontology mapping applications. More specifically, KDEG has an established mapping related research focus and the objective of FAME is interoperability support in federations with the means of mappings. In total 7 participants completed the questionnaire. The full details of the results for this questionnaire are in annex D.

An analysis of the results shows that the use-case "Find mappings between specific source & target ontologies (1a) or ontology elements (1b)" has been rated as most relevant (mean relevance rating of 100 %) and most frequent (mean occurrence rating of 86 %). All other proposed mapping discovery tasks have been rated as relevant with a minimal mean of 50 %. The lowest mean for likely occurrence was 36 % for the task "Mappings created by a specific matching algorithm implementation?".

[80] For more information please see the project home page on http://kdeg.cs.tcd.ie/fame

3.4.4 Requirements based on the domain analysis

The use-case of the OM2R defined in section 3.4.3 highlights that ontology mapping reuse is based on the ability of the ontology mapping engineer to discover relevant ontology mappings. This leads to the following requirement for the OM2R model.

R3) The OM2R model must support common ontology mapping retrieval use-cases.

The ontology mapping lifecycle as shown in section 2.2.2 shows that the creation of mappings is a complex process with multiple phases [Be11]. For example, the matching phase could be conducted in two stages with an initial string based matching algorithm and a second manual matching step to improve the accuracy further. This leads to the next requirement for the OM2R model.

R4) The OM2R model must allow the documentation of a multi-step ontology mapping lifecycle.

In the following section the identified requirements are addressed with suitable design features of the OM2R model design.

3.5 Design of the OM2R model

In this section the design features of the OM2R model are presented which address the identified model requirements. The following table 7 provides an overview of these and indicates what meta-data documentation aspects they primarily support. Each design feature is discussed in detail in the following subsections 3.5.1 and 3.5.2.

Table 7 Overview of requirements and design features

Requirements	Design Features	Support
R1) The OM2R model must support an expressive documentation of all meta-data elements and relations used in the OM2R model to document ontology mappings.	D3) Self-describing model elements embedded directly in the OM2R model	Consistency
	D4) Embed an editing workflow	Consistency
R2) The OM2R model must not replace existing matching representation formats but must compliment them with mapping lifecycle information in a mapping representation neutral way.	D1) OM2R model is designed as an ontology	Consistency
	D2) Meaningful relations between the meta-data fields in the model	Consistency
R3) The OM2R model must support common ontology mapping retrieval use-cases.	D5) Meta-data fields in the OM2R need to document the process and details of the extended ontology mapping lifecycle (see 3.4.2)	Relevancy
R4) The OM2R model must allow the documentation of a multi-step ontology mapping lifecycle.	D6) Process documentation to be embedded directly in the OM2R ontology	Relevancy

3.5.1 Design features that support consistency

The core requirement for the model is that the OM2R must not replace existing matching representation formats but must compliment them with mapping lifecycle information in a mapping representation neutral way (R2). To address this requirement the **D1) OM2R model is designed as an ontology** which allows the documentation to be achieved with expressive semantic statements. This means in contrast to the common approach where simple key-value string pairs are used, the OM2R model will express statements using RDF triples, e.g. *om2r:Source_Ontology om2R:hasHumanReadableName* "BioTop Ontology". This allows the reuse of generic relations in the model to describe different objects of interest, e.g. target and source ontologies has a name but also a matching representation. The ontology-based approach makes the OM2R an open framework where users can express information flexibly and can extend the vocabulary if needed.

Furthermore, an ontology-based approach allows the expression of **D2) meaningful relations between the meta-data fields in the model** which is not possible in a flat index structure which is the most common representation form for mapping meta-data currently. For example the OM2R model defines the relation *om2r:compatible_to* which indicates compatible field content pairs. For example OWL as a formal language is compatible with RDF(S) as its notation. This linkage is vital for consistency as is also the explicit statement about logical consistent feature combination between different related fields. This is superior to a traditional controlled vocabulary where the field content options are defined but not how they relate to options of other fields.

The requirement R1 states that the OM2R model must support an expressive documentation of all meta-data model elements and relations used in the OM2R model to document ontology mappings. This means the OM2R model must support the sharing of a common understanding about the meta-data fields, their intended application and meaning. The ontology-based approach enables the OM2R to offer detailed model documentation in terms of **D3) self-describing model elements embedded directly in the OM2R model**. Each model element has been designed to link to a short descriptive name, a textual definition, acronyms and a unique identifier. It is argued that this will help users to understand the intended meaning of each field, as all required information are embedded directly in the model rather in an external source. This approach provides the user with a ready-to-use vocabulary for documenting mappings and which can also be processed by tools automatically (see section 4.3). The meta-data fields in the OM2R model are linked and built on each other. For example a user should start with the general matching method field (manual or automated matching process) and then add more and more details as far as available (e.g. the specific matching algorithm applied). Thus, knowledge of the order in which the meta-data fields should be edited is essential for sharing a common understanding of the OM2R model fields. In order to help users edit these fields in an appropriate order the model is designed to **D4) embed an editing workflow**. For this purpose each OM2R field is given an *om2r:editing_priority* in terms of a number. An editing tool is able to interpret these numbers to create an interface with the corresponding order of fields. Experiment 3 (section 8.4.3) provides an example implementation of an editor that uses this information. As this information is embedded directly in the model, it does not rely on a particular tool implementation which again helps satisfy the first requirement (R2) of the OM2R model. Such editors can be updated in a dynamic manner if new context specific meta-data fields are added. The self-describing fields combined with the clear editing workflow helps users to understand the intended meaning of the fields and show how to optimally fill the model with content. This supports consistency as it helps to avoid incorrect usage of the model and support the logical flow of content creation for users.

3.5.2 Model design to support relevance of the model

The use-case for the OM2R is mapping discovery and as such the core requirement to support relevance is that the model must support common ontology mapping retrieval use-cases (R3). The OM2R addresses this need with **D5) meta-data fields in the OM2R need to document extended ontology mapping lifecycle (see 3.4.2).** They allow a detailed description of the ontology mapping lifecycle to support common mapping discovery tasks (see section 3.4.2). The list of supported meta-data fields in the OM2R is presented in section 3.6 which provides a full example of the OM2R model in action. In addition, the meta-data is structured according to the ontology mapping lifecycle steps, e.g. identification phase, matching phase. Please note the relevance of the OM2R fields have been verified in experiment 2 (see chapter 6). The example of the OM2R in section 3.7 provides an overview of all meta-data fields and their relation to the lifecycle step.

The fourth requirement (R4) states that the OM2R model must allow the documentation of a multi-step lifecycle in order to be able to represent complex ontology mapping lifecycles adequately. This is also vital to satisfy R3 in regards to mapping retrieval. The OM2R model addressed this need with classes and properties that allow the **D6) process documentation to be embedded directly in the OM2R ontology**. More specifically, the model contains the relation *om2r:belongs_to_step* which links each meta-data field to an ontology mapping lifecycle step. Please note a meta-data field can belong to more than one stage. For example the field *om2r:matching_algorithms* is linked to the first and second step in the matching activity, relating to matching iterations.

Figure 8 shows an example of the modelling template used in the OM2R model.

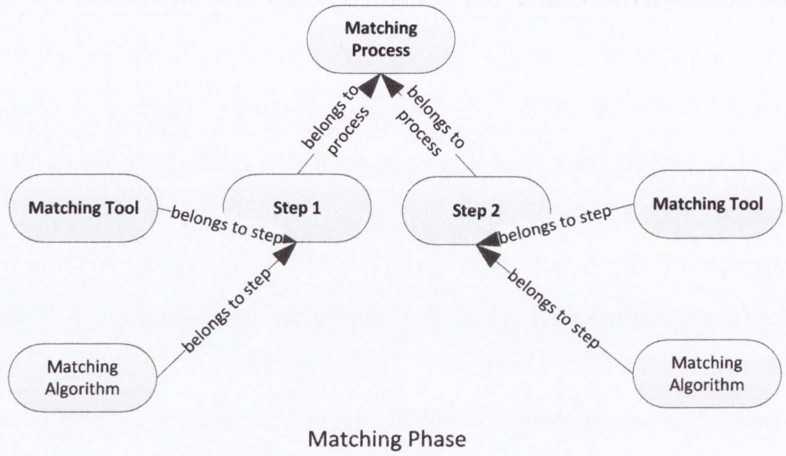

Figure 8 Multi-step process modelling in the OM2R model

The OM2R model offers a flexible representation of mapping meta-data and combined with the embedded documentation is designed to adapt to changing mapping domains and requirements over time. The development of the OM2R model mimics this expectation as the model underwent three improvement cycles which are discussed in the following section.

3.6 OM2R model overview

The design features presented in section 3.5 define the content and structure of the model. The driving factors of the OM2R model is the representation as an ontology which enables an expressive representation of the meta-data and addresses elements of interest in a language independent way. The following figure 9 provides an overview of the resulting high level layout of the OM2R model.

In the resulting layered approach the ontologies that are the subject of mappings are placed on the lowest layer. The middle layer contains the ontology mapping representations which specify the particular correspondences between the addressed ontologies elements, e.g. X maps to Y. The top layer contains the OM2R model where each entity of interest is represented in the model (light grey ellipses) and meta-data field provide details about them (white oval). As such the model provides meta-data to describe expressively the individual mapping representations and ontologies (arrow line). Please note the meaning of the individual elements in this graphic are explained in the legend at the bottom of the graphic:

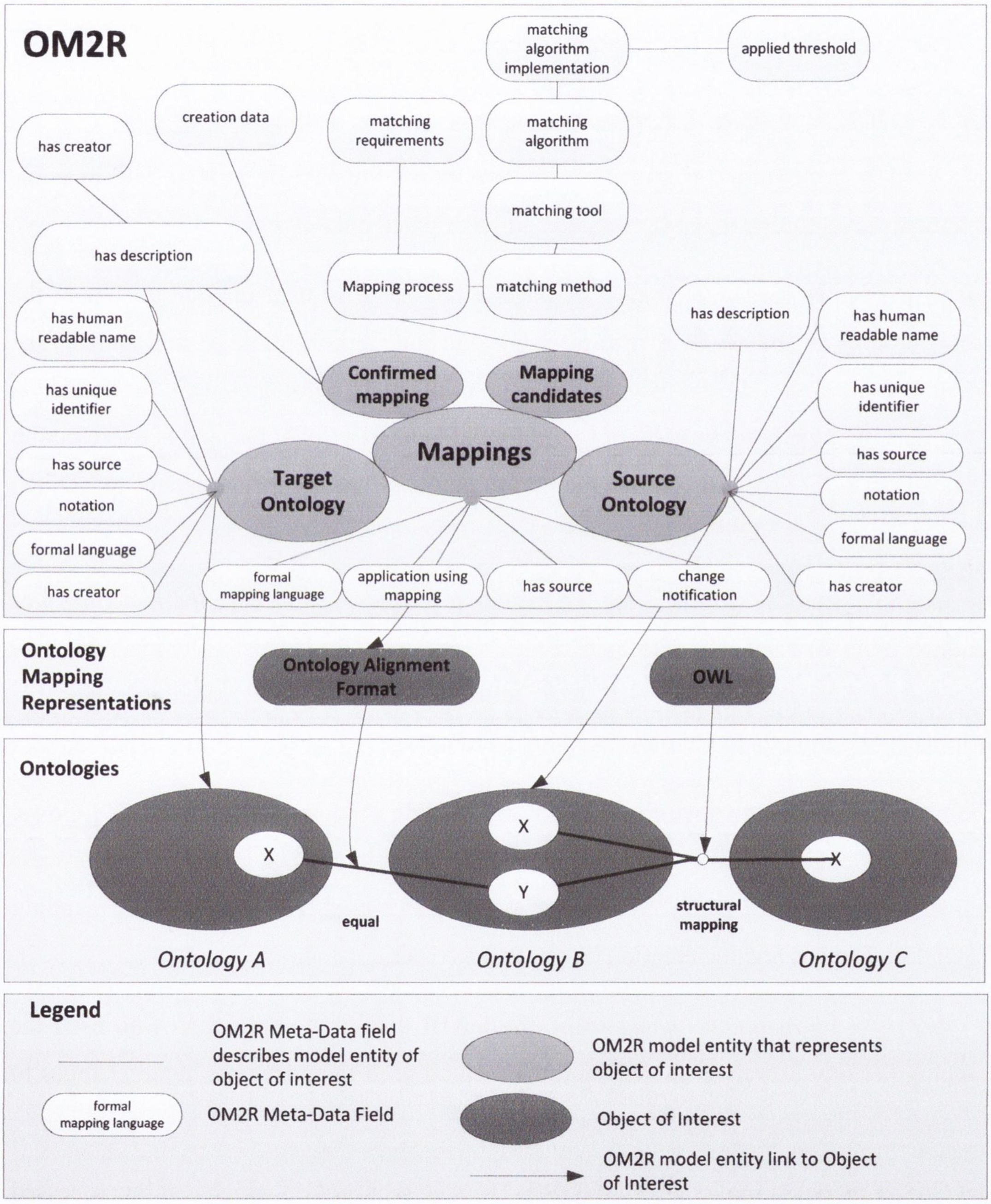

Figure 9 Overview the OM2R model

The model expressed as an ontology can be found on the attached DVD in the *folder D OM2R model*. Protégé project files are also provided which will enable a reader to apply this model for a given matching scenario but also to extend it for individual needs. The following section provides a full example of an instantiation of theOM2R model. The corresponding OWL DL representation of the model can be found in annex H.

3.7 OM2R model example

This section presents an example of an instance of the OM2R model. The example documents an ontology mapping between a book related ontology and a library ontology.

More specifically the OM2R model contains a number of statements that document the book ontology is the target of the mapping and the library ontology the source. The model provides the details where the source of both ontologies can be found. It determines that the book ontology has the formal language OWL DL and is expressed in the notation RDF/XML. In contrast the library ontology has the formal language RDF(S) and used the notion RDF/XML.

The example provides details about the matching process. In this case the matching was conducted in two steps. An automated process was applied to step 1. The matching tool Alignment API and Server 3.1 was utilized to apply the matching algorithm StringDistAlignment() to identify mapping candidates.

The mapping process is documented with the mapping requirement defined as follows: expert knowledge needed to ensure quality. The mapping is conducted in one process step where the mapping candidates are reviewed by an expert. The creator of the mapping is Hendrik Thomas and the creation date is 2014-05-02.

In regards to the mapping management, the model the know application who used the mapping (Trinity Library Searcher) and the method used to notify user about changes to the mapping (Trinity Library RSS new feed).

The OM2R model is implemented in OWL DL (see section 4.1) which can be processed and queried easily but is difficult to represent in a compact form. As a result the example shown in figure 10 shows the OM2R model in a simplified triple representation which highlights the subject and object of the discussion as well as the relevant predicates (relation types).

Please note not all OM2R fields are presented for demonstration purposes, for example all model elements have self-describing meta-data field like om2r:hasDescription which is not shown here.

The statements highlighted in green provide an example for the meta-data fields used for the self-describing element, e.g. the source ontology *om2r:hasHumanReadableName* "Book Ontology". The statements in orange demonstrate how a 2 step matching process can be documented in the OM2R model. The blue coloured statements show use of meaningful typed relations between field options. In this case, the notation XML/RDF is marked as compatible to OWL DL. A suitable editor is able to utilise this information to provide the users with active feedback of compatible selection options which can help to improve logical consistency (see section 4.4.2).

A full OWL DL export of this example can be found in annex H.

subject	predicate	object/subject	predicate	object
Book	is instance of class	om2r:Ontology	is sub class of	Lifecycle_Entities
	om2r:is_Target	om2r:Source_Ontology		
	om2r:has_Description	Classification of book types		
	om2r:has_Unique_Identifier	http://mybookontology		
	om2r:has_Human_Readable_Name	Book Ontology		
	om2r:has_Source	Ontology_Book_URL		
			o2mr:lastCheckedDate	5/4/2014
			o2mr:url_value	http://goo.gl/mRrlRJ
			om2r:belongs_to_phase	om2r:Identification_Phase
	om2r:has_Notation	om2r:RDF_XML	om2r:compartible_to	om2r:OWL_DL
				owl2r:RDF_S
			om2r:belongs_to_phase	om2r:Identification_Phase
	om2r:has_Formal_Language	o2mr:OWL_DL	om2r:compartible_to	om2r:RDF_XML
				om2r:Turtle
			om2r:belongs_to_phase	om2r:Identification_Phase
Library	is instance of class	Ontology	is sub class of	Lifecycle_Entities
	om2r:is_Target	om2r:Target_Ontology		
	om2r:has_Description	Classification of books in libraries		
	om2r:has_Unique_Identifier	http://trinity.org/lib/ontology		
	om2r:has_Human_Readable_Name	Library Ontology		
	om2r:has_Source	Library_Ontology_URL		
			o2mr:lastCheckedDate	3/4/2014
			o2mr:URI_value	http://goo.gl/GJTMGL
			om2r:belongs_to_phase	om2r:Identification_Phase
	om2r:has_Notation	om2r:RDF_XML	om2r:compartible_to	om2r:OWL_2
				om2r:OWL_DL
				owl2r:RDF_S
			om2r:belongs_to_phase	om2r:Identification_Phase
	om2r:has_Forma_lLanguage	owl2r:RDF_S	om2r:compartible_to	om2r:Turtle
				om2r:RDF_XML
				o2mr:N3
			om2r:belongs_to_phase	om2r:Identification_Phase
Mapping	is instance of class	Mapping	is sub class of	Lifecycle_Entities
	o2mr:is_Addressing	Book_Ontology		
		Library_Ontology		
	om2r:has_ Description	Mapping between books & library		
	om2r:has_Unique_Identifier	http://trinity.org/lib/map		
	om2r:has_Human_Readable_Name	Mapping between books & library		
	om2r:has_Matching_Requirements	Quality is most important		
	o2mr:has_Matching_Process	om2r:matching_proces_BookToLib	om2r:belongs_to_phase	om2r:Matching_Phase
			om2r:has_process_step	om2r:matching_step_1
				om2r:matching_step_2
om2r: matching _step_1	om2r:has_Unique_Identifier	1st step in this matching process		
	om2r:has_matching_Method	om2r:AutomatedMatching	om2r:compartible_to	om2r:Alignment API 3.1
				om2r:MAFRA
	om2r:has_Matching_Tool	om2r:Alignment API and Server 3.1	om2r:compartible_to	Automated Matching
	om2r:has_Matching_Algorithm	om2r:StringDistAlignment()	om2r:compartible_to	om2r:Alignment API 3.1
	o2mr:ElementTheAlgorymIsBasedOn	om2r:AnyRDFSLabel	om2r:compartible_to	om2r:StringDistAlignment()
	om2r:has_Matching_Scope	om2rMatchingOfAllElements	om2r:compartible_to	om2r:AnyRDFSLabel
	om2R:has_AppliedThreshold	om2r:50%SimilarityMeasure	om2r:compartible_to	om2r:AnyRDFSLabel
om2r: matching _step_2	om2r:hasUniqueIdentifier	2st step in this matching process		
	om2r:has_matchingMethod	om2r:AutomatedMatching	om2r:compartible_to	om2r:Alignment API 3.1
				om2r:MAFRA
	om2r:has_Matching_Tool	om2r:Alignment API and Server 3.1	om2r:compartible_to	Automated Matching
	om2r:ha_sMatching_Algorithm	om2r:StringDistAlignment()	om2r:compartible_to	om2r:Alignment API 3.1
	o2mr:Element_TheAlgoIsBasedOn	RDFS Labels for Classes	om2r:compartible_to	om2r:StringDistAlignment()

subject	predicate	object/subject	predicate	object
	om2r:has_Matching_Scope	om2rMatchingOfAllElements	om2r:compartible_to	om2r:AnyRDFSLabel
	om2R:AppliedThreshold	om2r:75%SimilarityMeasure	om2r:compartible_to	om2r:AnyRDFSLabel
Mapping	om2r:hasMappingRequirements	Expert knowledge needed to ensure quality		
	o2mr:hasMappingProcess	om2r:mapping_process_BookToLib	om2r:belongs_to_phase	om2r:Mapping_Phase
			om2r:hasProcessStep	om2r:mapping_step_1
om2r: mapping step_1	om2r:has_process_details	Candidates reviewed by experts		
	om2r:has_mapping_creator	Hendrik Thomas	dc:FirstName	Hendrik
			dc:Email	hendriktho@gmail.com
	om2r:has_date_Of_Mapping_Creation	om2r:2014-05-02	om2r:hasDay	2
			om2r:hasMonth	May
			om2r:hasYear	2014
			om2r:hasTimeZone	Auckland
Mapping	om2r: hasApplicationThatUsesTheMapping	Trinity Library Searcher	o2mr:lastCheckedDate	5/4/2014
			o2mr:URI_value	http://goo.gl/mRrsd
			om2r:belongs_to_phase	om2r:Management_Phase
	om2r:has_Change_Notification	Trinity Library RSS new feed	o2mr:lastCheckedDate	5/4/2014
			o2mr:URI_value	http://goo.gl/murlk
			om2r:belongs_to_phase	om2r:Management_Phase
	om2r:has_Source	Mapping_URL		
			o2mr:lastCheckedDate	3/4/2014
			o2mr:URI_value	http://goo.gl/GJTd
			om2r:belongs_to_phase	om2r:Identification_Phase
	om2r:has_Notation	om2r:RDF_XML	om2r:compartible_to	om2r:OWL_DL
				om2r:INRIA
			om2r:belongs_to_phase	om2r:Identification_Phase
	om2r:has_Mapping_Formal_Language	owl2r:INRIA	om2r:compartible_to	om2r:Turtle
				om2r:RDF_XML

Figure 10 Example of the OM2R model

3.8 Evolution of the OM2R model

In the previous sections the final version of the OM2R model has been described. However, the model has evolved in various aspects since its first version in 2009 (OM2R v1) which was created with a focus on mapping retrieval (see chapter 6 for details). The OM2R model was improved further for experiment 2 (OM2R v2) and finalised for experiment 3 (OM2R v3). In this section the changes to the individual meta-data fields are discussed.

An advantage of the ontology-based structure of the model is that such evolution details can be expressed explicitly in the model. In the OM2R each model element is linked to the OM2R version via the relation *om2r:has_version*. Thus the final version of the OM2R all elements are linked to the *om2r:LOM2R_v3_version* of the model. Further changes can then be linked to new version but the history version of the model retains if needed.

3.8.1 Changes from OM2R v1 to v2

OM2R v1 was designed for use in experiment one which investigated if retrieval of ontology mappings can be improved by using an ontology based OM2R model compared to using an existing ontology alignment representation. This was the reason why the available meta-data fields were focussed purely on the support for mapping retrieval.

In OM2R v2 the focus of the model shifted from retrieval to mapping reuse support. To address this issue the OM2R model was extended to provide more details for the actual matching and mapping process. For example the OM2R fields for matching algorithm and matching algorithm implementation were introduced. These added fields provide more details but are very specific to the actual process. To ensure that the user understood the intended meaning of each field the embedded meta-data documentation was introduced in OM2R v2. Furthermore, more high level meta-data fields for contexts, requirements and objectives of the actual mapping processed were added. At this stage these three fields are designed in the OM2R model as free text fields as no common values could be identified for a control list. The unique identifier field was also introduced in OM2R v2 to improve the support for automated processing tools and to enable the creation of the OM2R editor for experiment 3.

The OM2R v2 model was used in experiment two (see chapter 6) which was designed to derive a relevance ranking for the meta-data fields. The result shows that ontology name and process information are considered most relevant by the participants. Date and creator related details emerged at the bottom of the rating but were still considered relevant by a normalized number of 5.3 out of 49 participants. This ranking was used to improve the model further to version three (see chapter 6 for full details).

3.8.2 Changes from OM2R v2 to v3

The third version of the OM2R improved the model further to strengthen the support for application and logical consistency. For this purpose the model was extended to provide the compatibility relations between content options to enable the editing tool to recommend compatible content options. In addition, free text fields introduced in version 2 were redesigned, e.g. mapping requirements and goals. Free text fields are more likely to create inconsistencies as no automated checks can be applied to the content. In addition, from a user perspective a consistent differentiation between context, requirements and objectives is difficult. For example, the fact that a matching needs to be created fast and efficient can be seen by one user as an objective but for another as a requirement. Thus a design decision was made to focus on designing meta-data fields where a user can select from a given set.[81] Hence, the field contexts, requirements and objectives were combined to one meta-data field for requirements to support consistency.

In OM2R v1 only one text field was offered for the ontology name. In the library community names are used extensively to identify resources [Fu89]. It can be observed that names are used for different purposes, e.g. an abbreviation a file index, a short precise name to help users find the resources or a more detailed description to highlight the key topics. Consequently, a single field for ontology names is likely to be used by users for different purposes. This represents an obstacle to consistency. As a result in the OM2R v3 the field ontology name was extended to provide a more sophisticated solution with multiple names and definitions per element. This was motivated by the fact that the ontology name was identified as the most relevant field in the relevance rating experiment 2.

Furthermore, the field related to language aspects has evolved from a single field for ontology language in OM2R v1 to three fields describing the formal language, the notation and the formal language for mappings. This development was motivated by the OM2R design requirement R3 which demands the sharing of a common understanding of each meta-data field. A notation and a formal language are two fundamental different things. It was helpful to replace the generic term language with these two more specific terms to help users understand the intended meaning and avoid inconsistent usage.

In addition, the ability to document a multi-step matching and mapping processes was introduced. The OM2R v3 represents the final model version. Experiment 3 has provided

[81] This way these field options can be linked via the compatible relations to ensure consistency.

evidence to support the claim that the model can support application and logical consistency.

Overall this design for the OM2R model offers a way to provide a detailed and relevant documentation of ontology mapping in a semantic expressive model. The strength is the embedded meta-data documentation to support users in understanding the meaning of the model element. The next chapter discusses how this model design can be implemented in a OWL DL based ontology which can be managed and shared between users.

4 Implementation of the OM2R Model

This chapter presents how OWL DL was selected as the ontology language to implement a semantic representation of the OM2R model. Furthermore, an overview of the structure is given to show how the meta-data fields, the process and the embedded documentation of the elements are implemented in OWL DL. Additional tools are presented which were developed as part of this research to support the evaluation work for OM2R based mapping retrieval as well as OM2R editing and management.

4.1 Selection of the formal ontology language for the OM2R model

The first design requirement (R1) states that the OM2R model should provide a flexible complimentary extension of current mapping formats. The second requirement (R2) demands that an explicit documentation of the mapping and all model elements should be provided. To satisfy these requirements the OM2R model is designed as an ontology [An04]. This allows expressive representation to reduce ambiguity of the intended meaning of the meta-data fields [Gr94, Ma11]. Furthermore, it allows a flexible extension of the model to accommodate new developments in the dynamic field of mapping tools [Eu11a]. In addition, with the OM2R model it is not only intended to support ontology engineers but also automated processing. This is particularly essential to meet the interoperability requirements arising from large data sets, as found in the Linked Data community, or in the case of high number of ontology matchings such as found in the OAEI community. Automated reasoners can be used to reveal implicit knowledge but are also helpful in automated consistency checking. This leads to the question as to what specific ontology language the model should be implemented in?

The most common format for ontology mappings according to the state of the art analysis is the Ontology Alignment Format which is expressed in RDF(S). RDF(S) is able to offer an explicit representation but is not expressive enough to support automated reasoning. OWL DL adds more expressive structures to RDF(S), e.g. allowing predefined relations to be limited to particular class or instance. OWL DL also has a high number of reasoning tools available.[82] OWL-DL Version 1 was chosen as the formal language to implement the OM2R model in this thesis, as it provides the necessary expressivity and supports reasoning sufficiently in language and available tools. At the time of development of the OM2R, the newer OWL 2 edition was not fully developed and tool support was immature. It is worth mentioning that good progress has been achieved in the development of tools supporting OWL 2. For example widely used ontology editor Protégé has been extended to support constructs provided by OWL2. The commercial tool TopBraid composer also currently supports OWL2. Therefore, the impact of implementing the OM2R model in OWL 2 will be minimal. Please see annex G for a discussion of the differences.

4.2 Class structure and process representation in the OM2R model

The design presented in chapter 3 dictates the structures and meta-data fields which need to be presented in the model. This leads to the question as to how the meta-data fields of the OM2R should be represented in a specific OWL-DL ontology.

To find a suitable modelling approach or template the author of this thesis reviewed the following semantic languages / vocabularies. The focus was placed on the way meta-data information about the model or the domain of interest is stored. Languages /vocabularies were selected because they are either common Semantic Web languages or have an explicit focus on meta-data. As such their objective is similar to the OM2R model which documents ontology mappings. The following languages /vocabularies were examined.

- FOAF [Fr12] – Friend of a Friend format for documenting contact information
- Dublin Core [Du12] – common standard to document author information for information resources such as books
- RDF, RDF(S), OWL [An04]– common Semantic Web languages\Topic Maps [Ga06] – Semantic Web language based on back-of-book indexes

[82] See http://www.w3.org/2007/OWL/wiki/Implementations for an overview of OWL reasoning tools.

- SKOS [Mi09] – ontology vocabulary to represent index and thesauri in the library domain

Ontology Alignment Format [Eu06] – the most common vocabulary to express ontology matchings

Based on the review of these languages the ontology implementation for the OM2R model was designed. In detail, the OM2R model is structured into three main classes:

- *OM2R:Model_Elements* – this class contains all elements which are used to describe the mapping lifecycle, e.g. om2r:matching_algorithm.
- *OM2R:OM2R_Model* – this class contains one instance which represents the actual model and used to identify the version of the OM2R model and descriptive information about the model itself.
- *OM2R:Ontoloy-Mapping_Lifecycle_Entities* – contains the actual representation of the addressed ontologies and of the mapping representation. This OM2R element uses reification[83] to represents the objects which are described in the documentation and contains all the meta-data, e.g. om2r:creation_date or om2r:location.
- *OM2R*:Abstact – offers generic elements which are more high level and reused in more than one place in the class om2r:Model_elements like om2r:steps.

Figure 11 provides an overview of the individual classes. A more detailed description of the class and instances can be found in the model itself which is located in folder *D OM2R model* on this thesis DVD.

[83] This is closely related to the general question of reification, as defined in the topic maps standard. It refers to the activity "[…] to record information about when statements were made [...]". In short, this refers to the ability to make statement about model elements itself. See http://www.w3.org/TR/rdf-primer/#reification for more information

Figure 11 Class overview of the OM2R model

Another key design decision for the model is the ability to embed an editing workflow for each element. The editing workflow defines in which order the individual meta-data fields should be populated. This is implemented by assigning an Integer number in the data type *om2r_default_priority* to each instance and class reification instance. The OM2R editor (see section 7.4.4) is able to query these numbers and create a sort order for editing of the meta-data fields.

In addition, each element is linked to lifecycle steps. This is achieved by using the object property *om2r:belongs_To_Phase* which links each subclass of the class *om2r:Model_Element* to an instance of the class *om2r:Lifecycle_Process_Steps*. A similar approach is used to define process steps for the matching and mapping phase.

4.3 Embedded Model Documentation for self-describing fields in the OM2R model

A key feature of the model is its embedded meta-data documentation. In version 2 of the OM2R model the name field assigned to each addressed ontology was expressed as an individual instance of the class *om2r:ontology_name*. This approach is flexible as the name instance could be linked explicitly to other information. For example "car" is an instance of om2r:ontology name. The instance "car" can be linked to a language, e.g. "car" om2r:hasLanguage "English". The downside of this modelling approach is its complexity which makes practical documentation creation cumbersome. In particular the retrieval is more complex as a triple pattern is needed to be defined with the name element and linkage to the individual instance.

A meta-data model is more likely to be used if its fast handling and fit for purpose can be guaranteed. Therefore, a decision was made to implement the naming aspects of the OM2R model in an alternative more simple and straightforward modelling style. Instead of using class instances, the following data properties are used for each meta-data model element: *om2r:has_Definition*, *om2r: has_Unique_Identifier*, *om2r:has_Human_Readable_Name*.

The naming of these data properties have been inspired mainly by meta-data fields from the Dublin Core vocabulary, e.g. http://purl.org/dc/elements/1.1/identifier. However, a more explicit naming was used in order to satisfy the model requirement R2 which aims to support explicit documentation. For example the data property *human_readable_name* was chosen instead of the common rdfs:label data property[84]. To make the use and interpretation of the field explicit and easier to understand, the author of this thesis

[84] See http://www.w3.org/TR/rdf-schema for a detailed definition of this data property/.

decided to move the definition of rdfs:label "human-readable version of a resource's name" to the more explicit name of *om2r:has_Human_Readable_Name*. Please note an element can be assigned more than one name which allows the representation of alternative names or abbreviations.[85] This data property centric approach makes creation of the OM2R model easier and searches faster. Only property values need to be matched rather than complex triple patterns related to class instances.

A decision made in the design of the model demands that every element in the OM2R model should have some descriptive meta-data attached. However, the implementation choice of using OWL-DL leads to a problem as data properties can only be enumerated for instances but not classes. This is closely related to the general question of reification[86], as defined in the topic maps standard. It refers to the activity "[...] to record information about when statements were made [...]". In short, this refers to the ability to make statement about model elements itself. OWL DL has a strict separation between classes and instances. As a result in OWL DL it is not possible to assign meta-data data properties to a class. However, this is necessary to satisfy requirement 2 (R2) where an explicit documentation of each model element is needed.

In the OM2R model a work-a-round to this problem is applied, namely where for every class a specific instance is created. The name of this instance matches the one given to the class but with a lower capital and a leading "c_" character. In the OM2R model this is interpreted as if this instance is a reification of the actual class, thus all meta-data found for this instance describes the actual class. In addition, this c_ instance is an instance of the class om2r:Class_Representator. Figure 12 shows a screenshot to demonstrate this approach for the class *om2r:mapping* and the instance *om2r:c_mapping*.

[85] The default name is indicated in the string with a tailing * in the string value. This is considered a work-a-round but sufficient for the OM2R editors and common ontology editors such as Protégé.

[86] See http://www.w3.org/TR/rdf-primer/#reification for more information

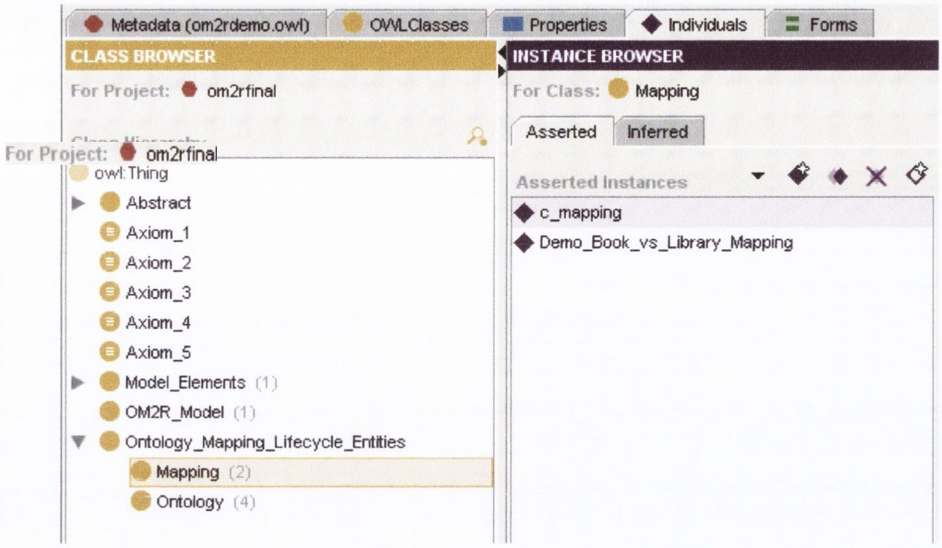

Figure 12 Reification for meta-data documentation

This is the only non-standard aspect of the OM2R implementation. To mitigate any misinterpretation these implementation decisions are explained in the instance of the class *om2r:OM2R_model* which acts as a main reference and introduction element for the OM2R model as a whole. Please note the element *rdfs:comment* field for the ontology itself refers to these fields for further details as this would be the common place for such explanations to appear.

4.4 Tools developed for the OM2R model

In this section various tools are described which were developed in this research to support the OM2R instance creation, and to demonstrate potential applications of the OM2R model. More specifically this section presents the OM2R Editor, the OM2R Finder and MooM.

4.4.1 Ontology Mapping Management – MooM Prototype

The primarily focus of the OM2R model is to support ontology engineers in documenting ontology mapping supported by suitable editing tool. However, often a more automated framework is needed to create and manage mappings on a larger scale [Sh12]. See chapter 9 for an example.

To investigate if the OM2R model can be used in this way the prototype system MooM has been developed as a proof of concept. It supports management of meta-data for ontology mappings. The following UML activity diagram in Figure 13 gives an overview of the architecture of the MooM prototype.

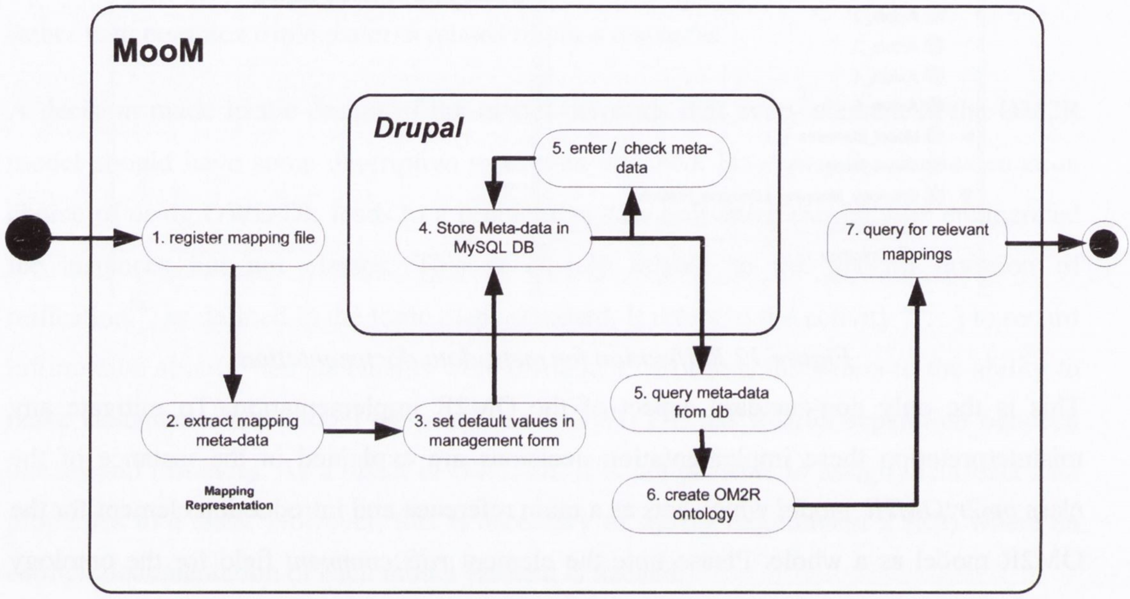

Figure 13 UML activity diagram of the MooM Prototype

MooM provides a graphical interface for uploading and maintaining of relevant meta-data for ontology mappings. MooM is based on the content management system Durpal 5.15 [Dr12]. MooM was developed as a custom Drupal module in PHP to support custom multi page forms. It enables users to register an ontology mapping into the system (1) and to enter all necessary meta-data (5) is stored in the MySQL DB backend.

Figure 14 shows a screenshot of the management form in Durpal with all current available mapping meta-data for a selected mapping instance.

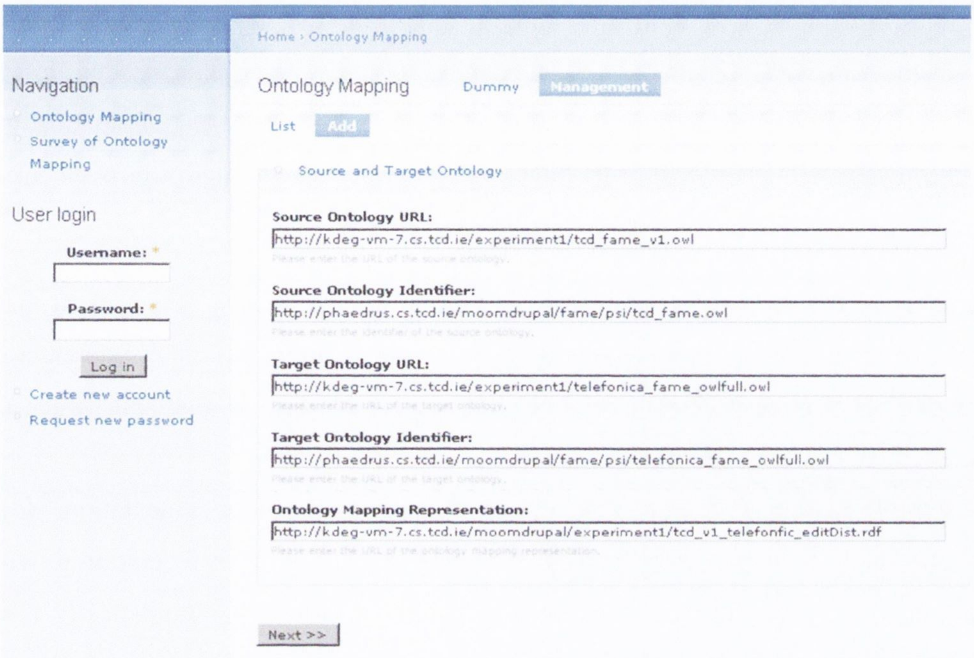

Figure 14 Screenshot of MooM

All relevant meta-data fields can be entered manually by the user. The key contribution of this tool is a set of custom web services which analyses a given mapping representation. The web services can extract relevant meta-data fields from the mapping representation automatically (2). The results are then shown in the editing form (3) for the appropriate meta-data and can be confirmed by the user. For example, the web service getOntologyLanguage() identifies the language of an ontology automatically. The web service getOntologySourcePath() returns the source path of the ontology. All web services are based on the JAVA web services framework AXIS 2[87]. In a final step the MooM Exporter connects to the database (5) and transforms the meta-data into the ontology-based OM2R model (6) expressed in the OWL DL ontology. The MooM tool was used to generate the OM2R model instance used for experiment 1.

Using Drupal as an open source content management system offers the advantage of a graphical interface and allows a flexible extension of the system. Drupal supports sophisticated functionality like full text search but on the other hand ties the OM2R model to a particular implementation. Thus, compared to the OM2R editor it is less

[87] More information can be found under http://axis.apache.org/axis2/java/core/

flexible with regards to dynamic changes, such that the interface would need to be adjusted manually if the OM2R model changes, e.g. if a new field is added. In contrast the interface for the OM2R editor used in experiment 3 is generated automatically based on the model itself and can reflect automatically any change and improvement of the model.

The state of the art study (see chapter 2) has shown that most of the current matching tools and environments export the ontology alignment format but only basic meta-data is provided. That means there is typically not much meta-data is available which can be extracted. It was thus decided to focus on a subsequent model and editor developments on how to best support a user to create relevant and consistent documentation. However, the case study in chapter 9 on ontology mapping management for federation provides an analysis of all meta-data information which can be extracted automatically from the ontology alignment format in more detail.

In the following section the OM2R editor is presented. This editor shows how a dynamic editing interface can be used to created OM2R model based documentation which can support the creation of consistent and relevant mapping documentation (see section 7.4.4).

4.4.2 OM2R Editor

The OM2R model offers the ability to document a multi-step mapping process to document ontology mapping lifecycle details. This motivates the need for a suitable interface which can support ontology engineers in the creation of meta-data which leads to the use-case for the OM2R editor:

The OM2R editor must allow an ontology engineer to view the available OM2R meta-data fields including the available embedded documentation. The editor must allow the editing of the meta-data fields based on the available content options defined in the model.

The OM2R editor implemented was employed in experiment 3 (see section 7.4.3) where users are asked to document mappings in a given scenario using the OM2R model.

The OM2R editor is based purely on the model itself. This means the editor processes the information contained in the model and based on the provided fields and relationships between the fields, the interface is generated automatically. As a result each change to the model results in an automatic update to the interface. This feature demonstrates how changes in requirements to document an ontology mapping can be handled on a model level, independent from a specific software implementation.

A screenshot of the OM2R tool can be found in figure 15. The online version of the tool can be accessed on:

http://ontologymappingdocumentation.com.escalade.mochahost.com/2012_Experiment_3_v10/1_intro.jsp

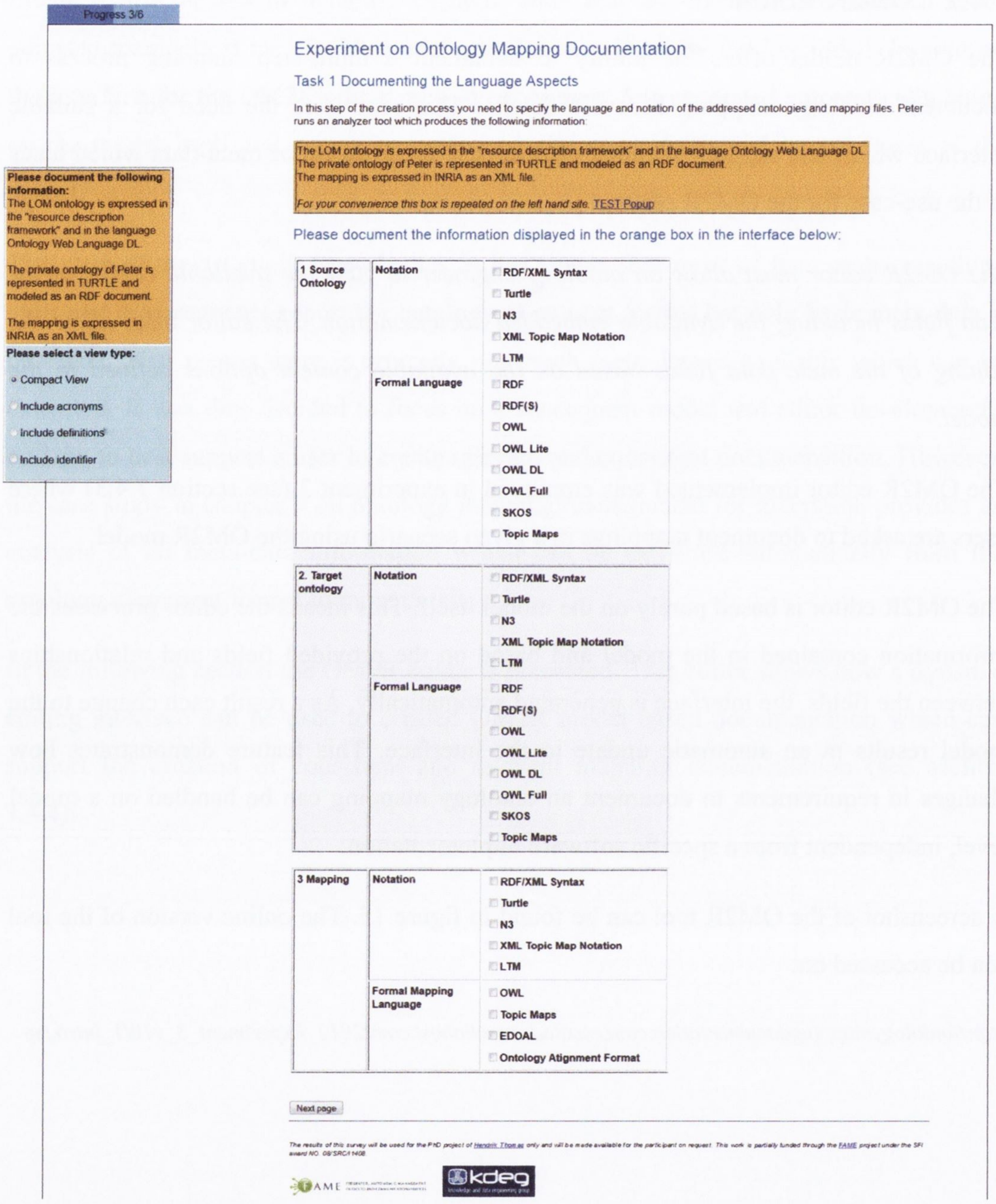

Figure 15 OM2R editor

The UML activity diagram in Figure 16 provides an overview of the activities supported by the editor. The activities are numbered and are related to the features of the OM2R editor in the following explanation text.

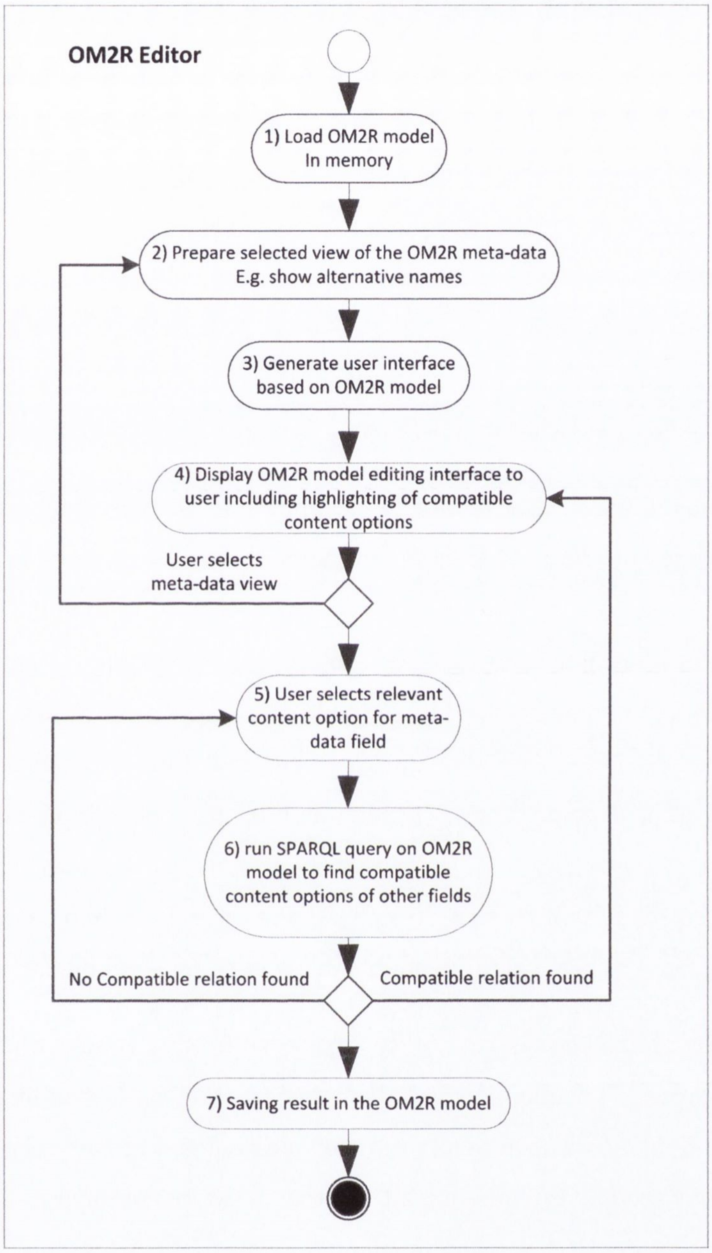

Figure 16 Activity Diagram of the OM2R Editor

The first feature of the editor is that the interface is generated automatically based on the OM2R model. For this purpose the OM2R model is loaded as an ontology into the system (1). The meta-data fields and content options are extracted from the OM2R model (2) and displayed to the user in a triple format based layout (3+4). The model is expressed as a number of semantic statements and the interface mimics this principle. For each meta-data field: the object of interest, the corresponding meta-data field and the options available to populate the field are shown. The user needs to tick the appropriate option to express that the meta-data field content is relevant for the given scenario (5). For example the first object of interest is the source ontology. The OM2R model describes that the source ontology can be described with the meta-data field notation. The model defined a number of notation formats which are displayed as option to choose from for the user, e.g. the source ontology is expressed in the notation XML/RDF. Please see figure 10 in section 3.7 for an example how such a comparability relation is expressed in the OM2R model. In the last step the meta-data information is stored as an OM2R model (7).

A second feature of the OM2R editor is its ability to visually highlight compatible content options based on the corresponding semantic relations expressed in the model (6). For example if the user selects the option for notation XML/RDF then the system detects automatically from the meta-ontology that the content option formal language RDF is compatible. Thus, both elements are highlighted blue. This demonstrates how a tool can utilise the rich expressive information to avoid contradicting statements and therefore help to achieve logical consistency. Highlighting is just one option and other more sophisticated uses are possible, e.g. pro-active suggestion system, presenting a dynamic view and so on. As default the model shows the default name for each element. The order of the fields is defined by the weighting numbers assigned for each meta-data field.

The third feature of the tool is its ability to use embedded model documentation to generate views types. The intention is to support users in understanding the intended meaning of each field. More specifically for each element and in particular the meta-data field instance, the model provides an all names, a list of acronyms, a short textual description as well as the URI as a unique identifier. The tool extracts this information for the generation of the view types (2+3). Each view type is focussed on particular meta-data information and the user can freely choose between those view types. The following view types are available: show only default names, show all acronyms, show all descriptions, show unique identifier of the relevant subjects.

In summary, the OM2R editor generates the interface as a special view of the OM2R model. Thus, each change in the model will result in an update to the editor automatically. The OM2R model not only consists of the actual documentation information but also provides information a tool can use to provide active support in creation of consistent documentation. The downside of this approach is that the OM2R is more complex than traditional models (e.g. LOM) as it not only contains meta-data fields but also workflow related information. The full potential of the OM2R model structuring of course can only be achieved if an editor can interpret the offered information appropriately, e.g. highlighting of compatible content options. In addition, the editor supports the full set of OM2R model but the available implementation is by intention rather simple and text-based. The reason for this design approach is that the tool is used in two experiments where the focus lies on the model itself (see experiment 3 and 4 for further details).

4.4.3 OM2R Finder

A typical application for meta-data is the support for information retrieval [Ch99]. The OM2R finder was developed in this thesis to investigate the benefits of the OM2R model for ontology mapping retrieval. OM2R Finder is designed as a JAVA based command line tool to discover existing ontology mappings based on the meta-data provided in an OM2R model. The UML activity diagram in figure 17 provides an overview of the involved functionality.

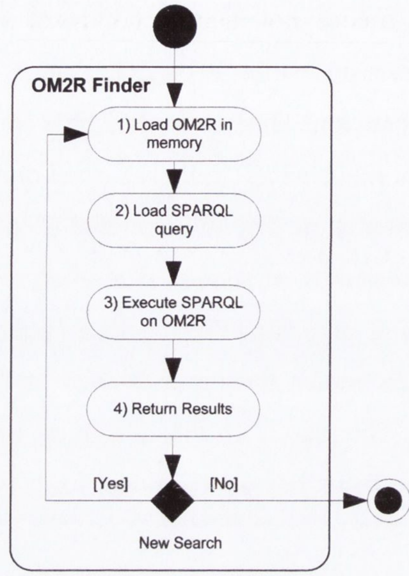

Figure 17 UML activity diagram of the OM2R Finder

The OM2R Finder loads (1) the latest version of the OM2R model into the application which describes the available ontology mappings. In a next step, a pre-defined SPARQL query is loaded (2) and executed on the OM2R model (3) to identify the relevant mappings. The SPARQL query defines a triple-pattern which expresses the required characteristics of the required mapping. In a final step the results are returned in the console.

OM2R Finder allows an evaluation of how ontology mappings can be discovered based on the OM2R model. The advantage of the tool is that semantic expressive search statements can be formulated using SPARQL. This makes the OM2R Finder a very powerful tool as all available information in the model can be used freely. The tool supports only OM2R version 1, as it was only designed to support an automated retrieval test in experiment one (see section 6.7 for details).

The first limitation of this tool is the lack of a user friendly interface. Queries can only be executed on a command level. In addition, although the use of SPARQL enables powerful querying over the model, the design of SPARQL queries turned out to be quite a complex and specialised task.

The next chapter presents the first experiment of this thesis which evaluates the OM2R model for ontology mapping retrieval. For this purpose the OM2R Finder is used in an automated retrieval experiment in a lab environment.

5 Experiment 1 - Evaluation of the OM2R for ontology mapping retrieval

This section presents an experiment where the OM2R model is used for an automated ontology mapping retrieval application. Evidence shows that the OM2R model can result in higher retrieval effectiveness and efficiency when compared to a mapping retrieval based on the common ontology alignment notation. The chapter contains the hypothesis, methodology, setup, data and conclusion.

5.1 Motivation

The extended ontology mapping lifecycle presented in section 3.4.2 shows that the first step in any ontology mapping reuse activity is the identification of existing relevant ontology mappings. The meta-data documentation of these ontology mappings provides the basis for the actual discovery retrieval process. Currently the majority of ontology matchings and mappings are expressed in the Ontology Alignment Format.

This motivates a need to evaluate the benefits of the OM2R model for ontology mapping retrieval compared to the current approach. Mapping retrieval was selected as the application area for this experiment as retrieval is one of the most common applications for meta-data in general [Du02, Be05]. The research described here was first published in the Workshop on Matching and Meaning (AISB 2009) [Th09a] and the European Summer School on Ontological Engineering and the Semantic Web 2009 [Th09c].

5.2 Hypothesis

The following hypothesis was defined for this experiment:

The retrieval effectiveness and efficiency of ontology mapping can be improved by using the OM2R model compared to ontology alignment notation.

Retrieval effectiveness and efficiency are common dimensions for the evaluation of retrieval systems [Ca11]. The term effectiveness refers to the level to which the system attains the stated objectives. In this context how much a mapping retrieval system can retrieve relevant mappings while withholding irrelevant mappings [Sa83]. The term efficiency refers to how economical the system is when achieving the objective that is time incurred in discovering relevant mappings.

Please note this experiment focusses only on the meta-data model and the impact of specific mapping retrieval, search engines are considered out of scope.

5.3 Methodology

This experiment is designed as an automated retrieval test in a lab environment [Sa83]. Such simulated studies are a common measurement for retrieval effectiveness and efficiency as those measurements in operational systems can be expensive in many aspects [Ch99, Ca11].

This experiment is based on a set of common mapping discovery tasks which were applied to a set of sample ontologies and corresponding mappings. The ontologies and mappings are based on data provided by the telecommunication company Telefónica as part of the research project FAME [Fa12, Br12]. They have been verified by a senior postdoctoral researcher involved in this project. Please see section 3.4.3 for details of how the mapping discovery tasks were identified. The author of this thesis has defined for each discovery task a list of relevant mappings. These are considered to be the gold standard and are used to determine retrieval effectiveness and efficiency.

The author is not aware of any retrieval tools which are designed explicitly for ontology mapping discovery. In order to evaluate the OM2R model the *OM2R Finder* (see section 4.4.3 for full details) was compared to the mapping retrieval tool *File Finder* which uses meta-data expressed in the Ontology Alignment Format. Both tools serve as infrastructure to enable SPARQL [Ar11] queries to retrieve mappings based on the provided meta-data. By comparing the retrieval effectiveness and efficiency between both tools, evidence can be gathered with regards to the benefits for mapping retrieval.

Salton and McGill [Sa04] proposed the following five evaluation criteria for information retrieval tests: recall and precision, response time, user effort, form of presentation, collection coverage. This experiment is designed as an automated test. Thus the focus will be on functional metrics which can be calculated automatically.

As a result the metrics recall and precision are used to measure retrieval effectiveness. Recall relates to the ability of the system to retrieve relevant documents. Precision relates to its ability not to retrieve non-relevant documents. The ideal system attempts to achieve 100 % recall and 100 % precision, but this is difficult to achieve because as the level of recall increases, precision tends to decrease.

It can be assumed that a system retrieves (a)+(b) number of documents, of which (a) document are relevant and (b) documents are non-relevant. Furthermore, (c)+(d) document are left in the collection after the search has been conducted. Thus (c) represents the documents which are relevant for the query but could not be retrieved and (d) documents are not relevant and thus have been correctly rejected. The metrics can be calculated as follows:

- **Recall**: Proportion of the relevant item received = a / (a + c)
- **Precision**: Proportion of retrieved items that are relevant = a / (a + b)

For retrieval efficiency the metrics query time and query complexity are used as they provide insights about the costs involved in identifying mappings.

- **Query time**: This is the average time needed to obtain a response from the system. In this experiment, this related to query time needed for each conducted query in terms of how many milliseconds are required to return the identified relevant mappings. The query time is an established and often used metric for evaluating retrieval systems [Ca11].
- **Query complexity**: This relates to the SPARQL query in terms of how many RDF triple patterns are used in the applied query. The amount of graph pattern involved in a query is an important component to evaluate the complexity of a query [Pe09].

The author of this thesis acknowledges that the controlled environment and experiment scenario can create a bias. To minimize the bias ontologies and mappings were chosen based on a real life scenario from the research project FAME. This project targets mapping retrieval explicitly and can help to create a realistic test setup. The mapping discovery tasks were deduced based on a questionnaire where participants have a known experience in mapping creation and application. Critical points in regards to potential bias are the two custom-built retrieval tools. They act merely as supportive infrastructure to enable the execution of SPARQL queries on the OM2R model and existing mapping formats. SPARQL is a well-established and common method to retrieve information from a semantic model. This makes SPARQL a valid and likely technology for mapping retrieval. [Ar11]. Furthermore, this experiment will focus only on the relative performance differences between the File Finder and the OM2R Finder. Absolute performance differences are impacted too heavily by the specific design of the retrieval tools and mappings. However, the general trend of the performance can provide evidence with regards to the hypothesis, for example, if the file finder is not able to identify a mapping based on author details.

5.4 Setup

This section describes the retrieval tasks the experiment is based on, the used ontologies and mappings and the experiment retrieval systems.

5.4.1 Experiment tasks

The set of common ontology mapping retrieval use-cases defined in section 3.4.3 are used as the basis for this experiment. For each use-case a specific mapping retrieval task was formulated by the author of this thesis. All tasks are listed in the table 8:

Table 8 Common Ontology Mapping Retrieval Use-Cases

#	**Mapping Discovery Use-Cases**
1	Find mappings between specific source & target ontologies (1a) or ontology elements (1b)
2	Find mappings for a particular ontology (2a) or an ontology element (2b)
3	Find mappings between a specific source ontology to any target ontology with specified characteristics, e.g. ontology language
4	Find mappings expressed in particular mapping format
5	Find ontology mappings for specific correspondence types
6	Find mappings created either automated, manually or in a combination
7	Find mappings created by a specific matching type
8	Find mappings created by a specific matching algorithm implementation
9	Find mappings created by a given matching algorithm configuration
10	Find automated created mappings based on matches with a high confidence level
11	Find manually created mappings depending on the involved users
12	Find mappings created by a particular author
13	Find a specific version of a mapping
14	Find mappings created for a particular context

5.4.2 Experiment Scenario and Ontologies

The following scenario was used in this experiment and applied to the identified mapping discovery tasks:

An ontology engineer is confronted with the need to create an ontology based application to integrate FAME-related information from the research partner Trinity College Dublin (TCD) and an industrial partner. For this purpose both groups have created ontologies representing relevant entities such as services, service providers and devices.

For this scenario three sample ontologies have been created by the author of this thesis based on data provided by the telecommunication company Telefónica as part of the research project FAME [Fa12, Br12]. All three ontologies represent a common knowledge domain with relevant concepts from the "Home Area Network Devices (HAN)" domain [Br12].

All three ontologies reflect different conceptualizations of the chosen concepts. A first ontology represents a potential view from the perspective of Trinity College

(TCD_v1.owl) and models relevant entities in a simple flat structure. A second ontology represents a newer version of this ontology with an improved structure (TCD_v2.owl), for example, the introduction of a subclasses hierarchy. A third ontology was created representing a potential view of the industrial partner upon this domain (industry.owl). This ontology has a slightly different structure, identifiers and labels for the same concepts.

All ontologies were expressed in OWL as a common ontology language. The source code for ontologies can be found on the attached DVD in the folder *A Experiment 1 Automated retrieval experiment*. The following table 9 shows the class structure of all three sample ontologies used in this experiment.

Table 9 Sample ontologies used for experiment 1

owl:Thing access_network_provider access_point communication_services gateway han_owners messaging_services multi_media_services router service_provider	owl:Thing ▼ parties access_network_provider service_provider han_owners ▼ services messaging_services multi_media_services ▼ devices router access_point gateway	owl:Thing ▼ devices access_points gateways routers ▼ involved_members wireless_network_provider wired_network_provider end-user service_provider ▼ services communication_services messaging_services energy_management
FAME TCD - version 1	**Fame TCD - version 2**	**FAME Industrial partner**

5.4.3 Experiment Mappings

To apply the selected discovery tasks, a set of sample ontology mappings was developed. Those mappings cover correspondences between the ontology TCD_v1.owl and industry.owl as well as between TCD_a2.owl and industry.owl. To reflect the variety of mapping creation techniques, the mappings between these ontologies were created based on the algorithm *EditDistNameAlignment* and *StringDistAlignment* of the Ontology Alignment API [Eu07]. These algorithms can only identify simple equivalence mappings. To include more complex relations (e.g. narrow or broader terms), mappings were also created manually by a senior postdoctoral researcher and a PhD student of the Computer Science Department, Trinity College Dublin.

A common representation of ontology mappings has not yet emerged (see section 2.7). To reflect the current state of the art, different mapping representations are used for this experiment. More specifically, the mappings are expressed in the common Ontology Alignment Format [Eu07] which is used as standard for the annual ontology alignment challenge OAEI [Eu11a]. The mappings are expressed in the Simple Knowledge Organisation System (SKOS) schema [Sk09] commonly used in the libraries community. Overall, this results in a total of 16 unique mapping files which can be found on the attached DVD in the folder *A Experiment 1 Automated retrieval experiment.*

For each of the 14 common mapping tasks, a specific discovery scenario was defined by the author. For example the objective of the task 1a is to identify all available mappings between the newest version of the TCD ontology (tcd_v2.owl) and the ontology created by the industrial partner (industry.owl).

For each tool and each task, a SPARQL query was defined by the author of this thesis, to identify relevant mappings. For example the SPARQL queries shown in table 10 were used for task 1a.

Table 10 SPARQL Example queries used in the File Finder and the OM2R Finder for Mapping Discovery Task 1

File Finder	OM2R Finder
PREFIX k: <http://knowledgeweb.semanticweb.org/heterogeneity/alignment#> SELECT DISTINCT $a $b WHERE { $a k:onto1 <http://phaedrus.cs.tcd.ie/OM2Rdrupal/fame/psi/tcd_fame.owl>. $b k:onto2 <http://phaedrus.cs.tcd.ie/OM2Rdrupal/fame/psi/telefonica_fame_owlfull.owl>}	PREFIX OM2R: <http://fame.org/OM2R#> SELECT DISTINCT $r WHERE { $a OM2R:mapping $r. $a OM2R:source_ontology_version '2'. $a OM2R:source_ontology 'http://phaedrus.cs.tcd.ie/OM2Rdrupal/fame/psi/tcd_fame.owl' $a OM2R:target_ontology 'http://phaedrus.cs.tcd.ie/OM2Rdrupal/fame/psi/telefonica_fame_owlfull.owl'}

5.4.4 Experiment System

For this experiment two mapping retrieval tools were developed. The UML activity diagram shown in figure 18 provides an overview of both systems. Please note the individual activities are explained below in more detail:

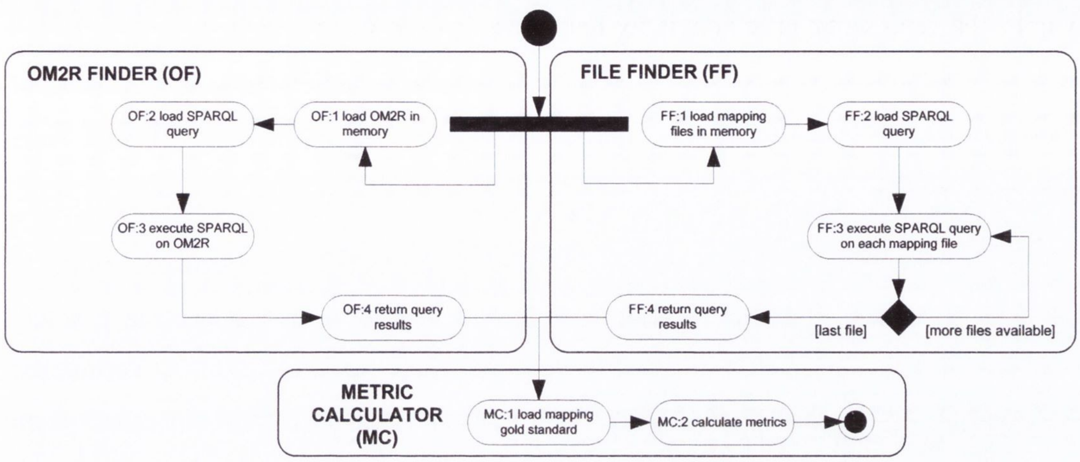

Figure 18 UML activity diagram for experiment 1 on mapping discovery

The File Finder (FF) system loads all available ontology mappings from a central collection into the system (FF 1). In the next step the FF loads a predefined SPARQL query (FF 2) and executes the query on all available mappings (FF 3). As a result a list of relevant mappings is then transferred to a metric calculating component (FF 4).

The OM2R Finder (OF) loads the OM2R ontology (OF 1) and executes the pre-defined SPARQL queries (OF 2 + 3). A list of identified relevant mappings is then transferred to the metric calculating component (OF 4). In a last step the metrics are calculated (MC 1+2). More details on the OM2R Finder can be found in section 5.4.2.

The File Finder, the OM2R Finder and the metric calculating component were developed in JAVA. The open source framework Apache Jena [Ap12] was used to enable SPARQL queries. For this experiment all components were combined into a single JAVA application. The experiment was conducted on 18th September 2009 by the author of the report and the experiment application was executed on an IBM T60 notebook with an IBM CPU T2300 @ 1.66 GHZ and 1 GB RAM in a time frame of 2 hours in total.

5.5 Results of Experiment

This section presents the results of this automated retrieval experiment. The raw data for this experiment can be found in the attached DVD in the folder *A Experiment 1 Automated retrieval experiment*.

5.5.1 Results for Recall and Precision

The first aspect analysed in this experiment was the retrieval effectiveness in terms of how well each individual system can return relevant mappings and withhold none relevant mappings.

Please note the recall and precision results will be presented as normalized values in percentages. The reason for this is that according to the gold standard the absolute numbers of relevant mappings are different for each mapping task. A normalized presentation helps to identify trends across all tasks as the focus of this experiment is on relative performance differences only. Table 11 shows the total number of relevant mappings per discovery task:

Table 11 SPARQL Queries used for Mapping Discovery Task 1

Task	1a	1b	2a	2b	3	4	5	6	7	8	9	10	11	12	13	14
Number of relevant mappings	8	12	9	12	0	8	8	8	8	4	4	16	4	8	16	16

Figure 19 shows the results for recall for both tools in comparison.

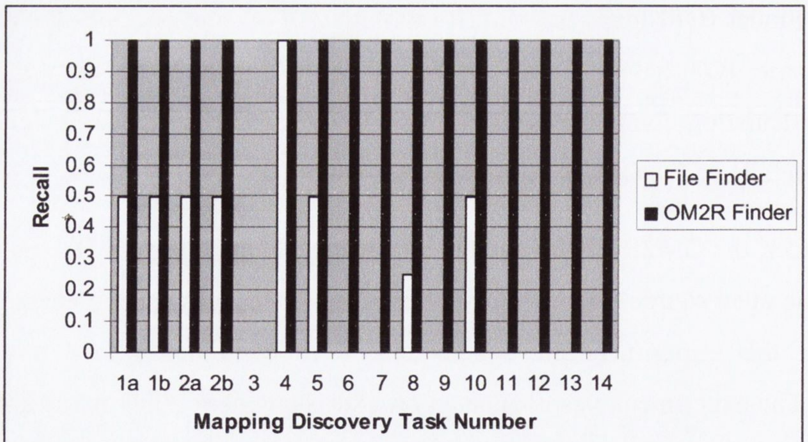

Figure 19 Recall for experiment 1

The graph shows that individual recall results for the OM2R Finder (black bar) and the File Finder (white bar). For example the File Finder shows a recall of 50 % for task 10. This means the File Finder returned 8 out of 16 relevant mappings according to the gold standard. It can be noted that in comparison, the File Finder returns lower recall values for all tasks except task 4. For 7 out of 16 tasks the File Finder was not able to identify any relevant mappings at all. Please note for task 3 the number of relevant mappings is zero and both tools return zero results correctly.

Figure 20 shows the results for precision for both tools in comparison.

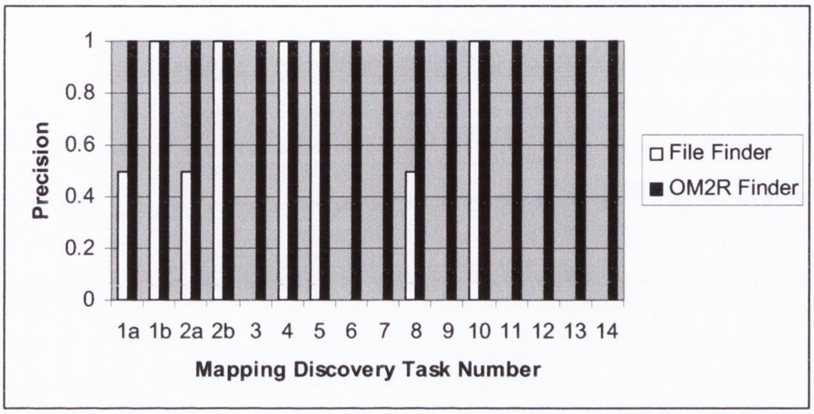

Figure 20 Precision for experiment 1

The graph shows the individual precision results for the OM2R Finder (black bar) and the File Finder (white bar). For example the File Finder achieved a precision of 50 % for task 8. This shows that 2 out of the 4 returned mappings where relevant according to the gold standard. The graph also shows that the OM2R Finder achieved 100 % precision for all tasks. In comparison the File Finder returned lower recall values for task 1a, 2, and 8.

5.5.2 Results for Query Time and Query Complexity

The next analysed aspect focusses on retrieval efficiency in terms of how economical the system is when achieving the objective. Figure 21 shows the query time for each SPARQL query for each tool used to answer the individual discovery tasks:

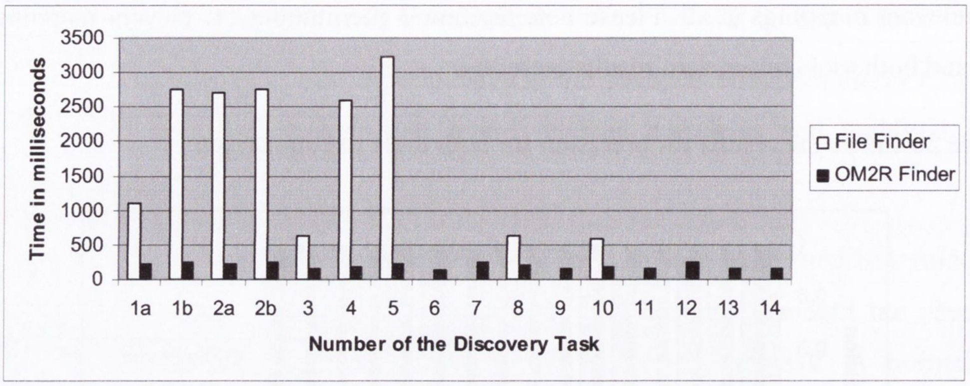

Figure 21 Query time

For example the File Finder required approx. 3250 milliseconds to return the search results for task 5. The graph shows that the query time for the OM2R Finder was lower for all tasks. The only exceptions are tasks 6, 7, 9, 11, 12, 13 and 14 where the File Finder could not return any results and therefore no suitable query could be created in the first place.

In addition, to the query time, the complexity of each used SPARQL query is also relevant for retrieval efficiency. The more triple patterns involved, the more effort is needed to find a matching pattern [Pe09]. Complexity was measured in this experiment by the number of triple patterns used in each query. Figure 22 shows the complexity of each the SPARQL queries applied for each discovery task per tools:

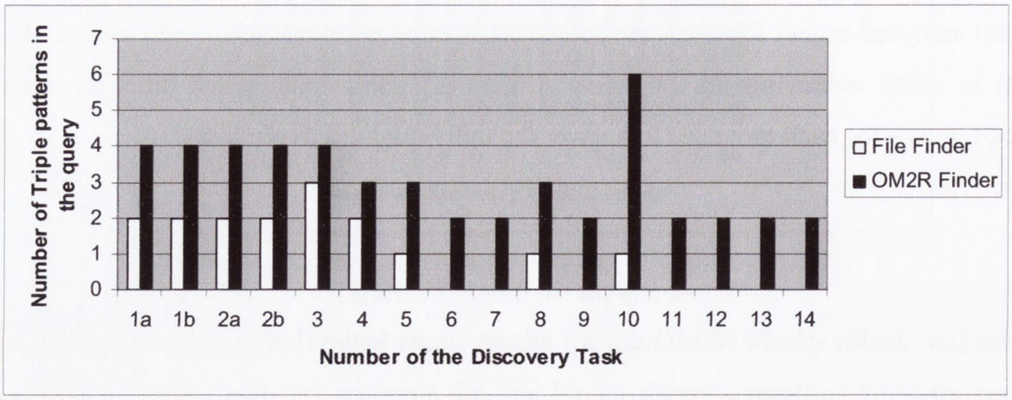

Figure 22 Complexity

For example the query used by the OM2R Finder tool in task 10 contains 6 patterns compared to 1 pattern used in the File Finder. It can be observed that the SPARQL queries used in the OM2R Finder are more complex in the OM2R ontology across all tasks, except those where no query could be generated.

The following section presents the analysis of these results.

5.6 Analysis

As a first step the focus is placed on retrieval effectiveness. Table 12 summarizes the individual recall and precision results. It shows the performance of the File Finder compared to the performance of the OM2R Finder. The symbol "<" represents a lesser performance, "=" represents an equal performance and 0 stands for failed, i.e. the File Finder was not able to retrieve any relevant mapping. Such a view on relative performance differences between the File Finder and the OM2R Finder is helpful as the absolute performance differences are impacted by the specific design of the retrieval tools and mappings.

Table 12 Comparison of File Finder performance for recall and precision with the performance of the OM2R Finder

Task	1a	1b	2a	2b	3	4	5	6	7	8	9	10	11	12	13	14
Recall	<	<	<	<	0	=	<	0	0	<	0	<	0	0	0	0
Precision	<	0	<	=	0	=	=	0	0	<	0	=	0	0	0	0

Table 12 indicates that the OM2R Finder provides better retrieval effectiveness. More specifically, the tool using the OM2R model returned more relevant mappings (higher recall ratings in 15 out of 16 discovery tasks) and the results were more accurate (higher precision ratings in 10 out of 16 discovery tasks).

A possible reason for the lower results of the File Finder was the common lack of meta-data in the mapping representations. The OM2R is designed to document aspects of the ontology mapping lifecycle with a focus on the identification of the addressed ontologies as well as matching generation and management details. The lack of such information in current mapping representations makes it impossible for the File Finder to answer certain discovery tasks. For example no information on mapping context or ontology versioning is stored in the Alignment Format or in SKOS. In addition, available meta-data was represented quite differently in the addressed mapping representations. This made it difficult to extract the required information. For example details about the author is modelled in SKOS [SK12] using the Dublin Core [Du12] concept *creator* and in the Ontology Alignment Format within the meta-data field *provenance*. In contrast, in the OM2R model all meta-data fields were mapped to a common meta-data model which improves recall and precision.

Furthermore, the results gathered in this experiment indicate that the OM2R Finder provides a higher level of retrieval efficiency. The query time of the OM2R Finder was considerably lower for all tasks (9 out of 16 tasks and for the remaining task no query could be generated at all). However, this can be explained by the fact that the File Finder has to query every mapping file individually. In contrast the OM2R Finder only queries the OM2R ontology which acts as central repository for the meta-data.

It can be observed that the SPARQL queries used in the OM2R Finder are more complex than the once used in the File Finder (9 out of 16 tasks and for the remaining task no query could be generated at all). The reason for this is that the OM2R ontology provides more meta-data in a more complex knowledge structure. As a result more triple patterns are needed to identify relevant mappings in the graph based structure of the OM2R. The negative impact on the retrieval time depends more on the specific implementation but less on the actual meta-data model. In addition, the specific way the SPARQL query is designed has also an impact on the complexity of the query which is not limited to the number of patterns [Pe09].

5.7 Conclusion

In this automated mapping retrieval experiment evidence was gathered which indicates that an ontology based OM2R model can result in a higher retrieval effectiveness and efficiency. This is compared to an approach limited to meta-data being represented in the Ontology Alignment Format as the most common current mapping representation. In addition, the experiment provides evidence that an OM2R model based application can support mapping retrieval, which is a vital precursor for mapping reuse as relevant mapping instances need to be identified first before any reuse can be considered.

The insights gained in this experiment cannot be generalized due to the controlled scenario and purpose built tools. However, the results show a trend that a higher amount of meta-data, with a more expressive structure and an explicit focus on ontology mapping lifecycle details can improve the discovery of ontology mappings. This experiment was based on real life mapping discovery tasks. This shows that the OM2R model was able to document the needed meta-data, which demonstrates the relevance of the OM2R for applications such as mapping discovery. Overall it can be concluded that the superior performance of the OM2R model approach suggests that an ontology based OM2R model is feasible from a domain modelling and technical point of view.

The different mapping discovery tasks tested in this experiment demonstrate the wide range of potential relevant meta-data. Meta-data creation is complex and time-consuming [Ca11]. To minimize the creation effort, it is essential to understand the specific extent of relevance of each meta-data fields towards the documentation support for mapping reuse. The next experiment addresses this question by creating a relevance ranking of the proposed meta-data fields in the OM2R model based on an end-user experiment.

6 Experiment 2 – Participant-based Relevance Ranking of OM2R

This section presents an experiment to investigate the relevance of the proposed OM2R model toward the ontology mapping reuse application. The result of this experiment will provide a relevance ranking of each meta-data field based on ratings by the participants. The following sections provide the experimental hypothesis, methodology, setup, data collected, analysis and conclusions.

6.1 Motivation

The decision to reuse existing ontology mappings depends on various factors, and as such the meta-data needed to support this decision is driven by the individual mapping re-use scenario.[88] As such feedback from a diverse group of participants is required to mitigate the impact of the heterogeneity of mapping reuse-cases [Eu06] that exist. This motivates the need to rank the identified list of OM2R meta-data fields for their relevance towards the mapping reuse use-case. A ranked list will help to focus the model and as such reduce the effort needed to populate the model.

6.2 Hypothesis

Based on this motivation the following hypothesis for the experiment was defined:

> *The proposed meta-data fields in the OM2R model are considered relevant for an ontology mapping reuse decision.*

6.3 Methodology

In this experiment participants are presented with two mapping scenarios and asked to rate the meta-data fields of the OM2R model for their relevance regarding mapping reuse. The experiment is conducted in a lab environment with two specific mapping reuse scenarios. See section 6.4.1 for details. The experiment was designed as an anonymous

[88] Please note ontology mapping reuse is chosen as application as it is a high level task and incorporates more specific applications such as mapping retrieval.

questionnaire [Sa83]. The participants were asked to complete the experiment in one session of approximately 20 minutes length.

The method used in this experiment is based on the GQM approach [Ba00a] which provides a structure for evaluation starting with a goal, questions that describe the goal further and metrics to answer the question.

The primary goal of this experiment is to create a relevance ranking of all OM2R fields. The question is therefore for each meta-data field, how the participant rates the field relevant for a given mapping reuse decision in a given scenario. The metric is a simple Boolean value for each participant and each meta-data field in terms of relevant: yes or no.

However, the focus of this thesis is on the model itself. Thus the meta-data fields should be rated for their relevance but not for the specific field content selections - since they depend on the specific scenario. For example the field ontology name could be presented with the field selection "X123" or with a human readable value "private IT ontology". The danger is that a participant could rate the field more relevant simply because of the more human readable label which is easier to interpret field value. Also the field could potentially be rated as less relevant because of the cryptic field value "X123".

To mitigate the impact of the specific meta-data selections on the relevance rating the experiment is conducted in two steps. In a first step the participant is presented with a reuse scenario and the meta-data fields but without any specific meta-data field content. The participant is asked to state if the meta-data fields are relevant and the participant sees only the name and the definition of the field.

In a second step, the same reuse scenario is presented but for every meta-data field a specific field selection shown. In other words the meta-data field is filled with content to document the given scenario, e.g. mapping author :hasName: Hendrik". This results in two relevance ratings given per participant per scenario and in the following section this is referred to as *task oriented relevance metric*.

The task oriented relevance is based on the specific scenario and in this way biased by the specific task and formulation used. It is valuable to gather evidence of the subjective relevance rating of the participant too. In other words, what is the subjective opinion about the relevance of the meta-data fields for ontology mapping reuse independent from the given experiment scenario and setup.

To gather this subjective relevance rating at the end of the experiment each participant was asked to select from all the OM2R meta-data fields a single data fields according to the following three criteria: What is the most useful meta-data field, what is the least useful meta-data field and what is the most difficult to understand meta-data field? The answers to these questions will be referred to as *subjective relevance* ratings.

A combination of the subjective and task oriented relevance ratings per participant will allow the creation of an overall relevance ranking of the OM2R model fields by using the average relevance ranking per field.

The experience of the participants with regards to ontology mappings and mapping documentation can potentially have an impact on their relevance ratings. For example if a participant is not familiar with the individual ontology mapping topic, it is likely more difficult for the participant to rate the relevance of the meta-data fields compared to expert participant who has knowledge of the purpose of the topic documented in the field.

To investigate the impact of the participant experience on the relevance results, the experiment contains questions to capture the experience of the participant in the ontology mapping domain. More specifically, the initial questionnaire contains the following three questions:

- **Ontology creation experience**: It is asked: "How many years of experience do you have in ontology creation / application?" The participant could choose: no experience, less than one year, between one and two year, three or more years of experience or not select any experience level,
- **Ontology mapping experience:** In addition, it is asked: "How many years of experience do you have in creation of ontology mappings/ matchings?" The participant could choose: no experience, less than one year, between one and two year, three or more years of experience or not to select any experience level.

- **Ontology mapping reuse experience:** Furthermore, the participant is asked: “Have you ever reused an existing ontology mapping?” The answer includes yes, no or no answer at all.

These individual experience metrics can be used to group the participants into experts with experience in mappings and novice participants with no experience. More specifically, a participant is classified as an expert if: The participant has answered yes for ontology mapping reuse experience and has one or more years of ontology mapping experience and has one or more years of ontology creation experience.

In the next section the experiment setup is presented.

6.4 Experimental Setup

This section will provide an overview of the scenario, experiment tool which allows the participant to express relevance ratings and details about the targeted participant groups.

6.4.1 Scenario

The chosen use-case for this experiment is mapping reuse. Figure 23 shows the scenario used in this experiment which was derived from the Home Area Network domain [Br12]. The scenario is to enable the integration of multiple information resources to create comprehensive overview of Home Area Network product. The scenario was selected as its content and the source project had an explicit focus on information integration and provided ample opportunities for mapping reuse. The scenario provides the participant with a research relevant context and a clear motivation why the actual mapping is conducted.

A retailer has decided to add Home Area Network (HAN) products to their range of goods. You will play the role of a new product manager. Your task is to create a new product survey as the market for HAN products is very heterogeneous. In particular, you need to learn what products are available and which of them are equal in terms of functionality. Fortunately, all the main vendors and wholesale dealers have published ontologies which describe their individual products and offered services. Analysing all those ontologies yourself would be too time consuming. However, on the Internet many ontology mappings are available which describe the relationships between those ontologies, e.g. product A in ontology X is equal to product A1 in ontology Y. Your Task: We will present you two different sets of ontology mappings (marked red and blue). Please evaluate these ontology mappings and decide if you what to reuse the ontology mapping in your study or not?

Figure 23 Mapping Reuse Scenario for Experiment 2

Based on this reuse scenario two specific mappings examples are provided. A simple text-based representation was used for the target and source elements as well as the explicit mapping representation. The motivation for using a text representation was to keep the experiment simple and allow even novice participants to understand the mappings without the need to interpret a semantic representation such as RDF/XML. Both mapping examples were created by the author of this thesis. The first mapping contained familiar names of products as shown in table 13.

Table 13 Mappings with human readable names for Experiment 2

Source Ontology	Mapping Relation	Target Ontology
Class: DVD Player	is equal to	Class: Digital Video Disc Player
Class: Video Recorder	is equal to	Class: VCR Player
Class: I-Pod	is equal to	Class: I-Pod Player

The second mapping set contains only anonymous product identifiers, such as those often used by inventory systems. These mappings put more emphasis on the meta-data as participants are not able to draw conclusions for the reuse by simply judging if the mapping is valid from the element names. In other words it is easier to see that I-Pod and I-Pod Player refer to the same thing but not so for XZ12B to Digi 115760. Table 14 presents the second mapping.

Table 14 Mapping with ID labels for experiment 2

Source Ontology	Mapping Relation	Target Ontology
Class: XZ12b	is equal to	Class: Digi 15760
Class: IOSONO 3245	is equal to	Class: V46565
Class: K1000	is equal to	Class: P100d

The participant is then asked to rate the relevance of the following OM2R meta-data fields shown in table 15. Please note the names and definition are in some cases slightly different to the final version of the OM2R model. The reason for this is that the result of this experiment was used to improve the meta-data model (see section 3.8.1 for details about the evolution of the OM2R model).

Table 15 Overview of OM2R meta-data fields which the participants are asked to rate for relevance for experiment 2

Title	**Definition**
Ontology name	A descriptive and human-readable label for the ontology, e.g. 'Wild Animals'.
Ontology source	The location of the ontology file in terms of a digital and downloadable representation, usually a URL.
Ontology creation date	An explicit time and date of the creation of the ontology.
Ontology creator	Some details identifying the creator (usually a human) of the addressed ontology.
Ontology language	An ontology language is a formal language used to encode the ontology.
Ontology size	Some indications of how large the ontology e.g. 5 classes, 10 individuals. Please note there are obviously many different ways to specify the size but any indications provided might be helpful for the evaluation process, as bigger ontologies only tend to be more difficult to map.
Ontology design	Some indications on the basic design of the ontology, e.g. a sophisticated and deep hierarchy , a flat class hierarchy with no or few parent-client classes
Matching method	Which method was used to find suitable candidates for a mapping in the addressed ontologies, e.g. manual selection, automated algorithm or mixed
Matching algorithm	If an automated selection was applied, this section provides a descriptive and human-readable label to identify the used matching algorithm.
Matching selectors	Some details identifying the people who manually selected the mapping candidates.
Matching algorithm Implementation	A descriptive and human-readable label to identify the specific implementation of the algorithm. Could be a URL or a specific JAVA class name such as org.jena.stringComparsion. Also helpful is to provide a URL to download the source code.
Matching selectors Backgrounds	Some details about the knowledge and background of the selectors in regard to this matching task.
Matching selection process	Some details of the applied selection process to find matches between the ontology elements.
Mapping objective	Some details about the purpose and aim of the performed mapping.

Mapping requirements	Details about the specific rules which were followed in the mapping process (if any).
Mapping context	Details about the wider context in which the mapping was applied.
Mapping process	Details about the applied mapping process. In other words how are confirmed mappings selected from the generated collection of matching candidates?
Mapping creator	Some details identifying the creator (usually a human) of this mapping.
Date of mapping creation	An explicit time and date of the creation of the mapping.
Mapping language	A mapping language is a formal language used to encode the mapping correlation.
Application using mapping	Details about known application which actively use this mapping.
Source of the mapping file	The location of the mapping file in terms of a digital and downloadable representation, usually a URL
Change notification	Details about the process (if any) used to notify participants on changes in the mapping file. For example, a RSS feed could be used.

6.4.2 Experiment Tool

The experiment tool was implemented as an interactive web page based on PHP and JavaScipt. Each participant was identified by the IP address of the accessing computer and if provided by their e-mail address. The web page was hosted on servers located within the School of Computer Science and Statistics server room. The experiment was conducted in July 2012. The experiment tool was structured into four pages that focus on the activities that are explained below:

1. **Introduction**: On the first page the participant was introduced to the purpose of the experiment. Furthermore, the participant is asked about their experience in ontology creation, ontology mapping and mapping reuse. The users need to approve the experiment disclaimer to continue. More details can be found in section 7.3 Methodology.
2. **Tutorial:** On the next page, the participant was shown a short video including audio commentary that gave an overview of the meta-data model. This four minute long video is available on: *http://www.youtube.com/watch?v=JSLUZG-gvBw*.
3. **Task:** On the following page, the participant was introduced to the scenario of the experiment and the tasks involved as described in section 7.4.1
4. **Relevance rating**: On this page the participant was asked to review the two mapping sets shown in section 6.4.1. For each set, the OM2R meta-data fields were displayed as shown in Figure 24. As explained in the methodology section above the model

review cycle was structured into two stages. First only the meta-data fields were presented, then the fields with the content were.

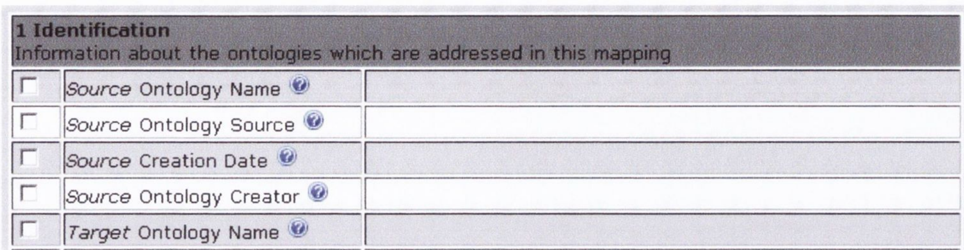

Figure 24 Experiment tool meta-data relevance ranking – only meta-data fields

In a second step, the same reuse scenario is presented but for every meta-data fields a specific field selection is shown. The participant is asked the same question but this time must choose a field with field content displayed, e.g. mapping author :hasName: Hendrik. Figure 25 shows an example of this section in the experiment.

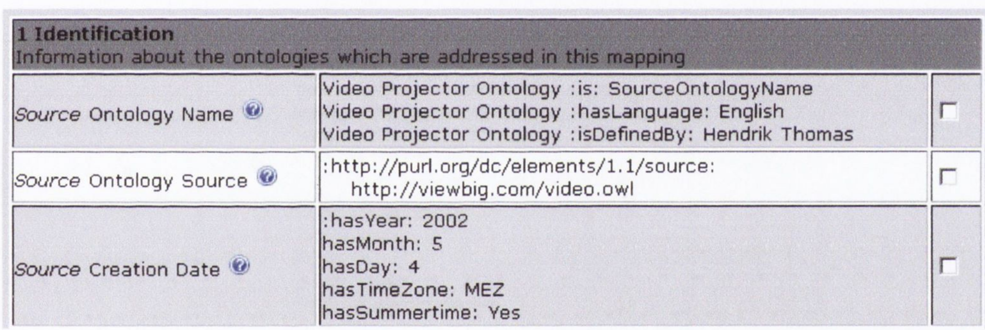

Figure 25 Experiment tool meta-data relevance ranking – only meta-data fields for rating relevance for meta-data fields with field content selection

Figure 26 shows a screen shot of the experiment page. In this section the participants are asked to rate the relevance of the meta-data fields.

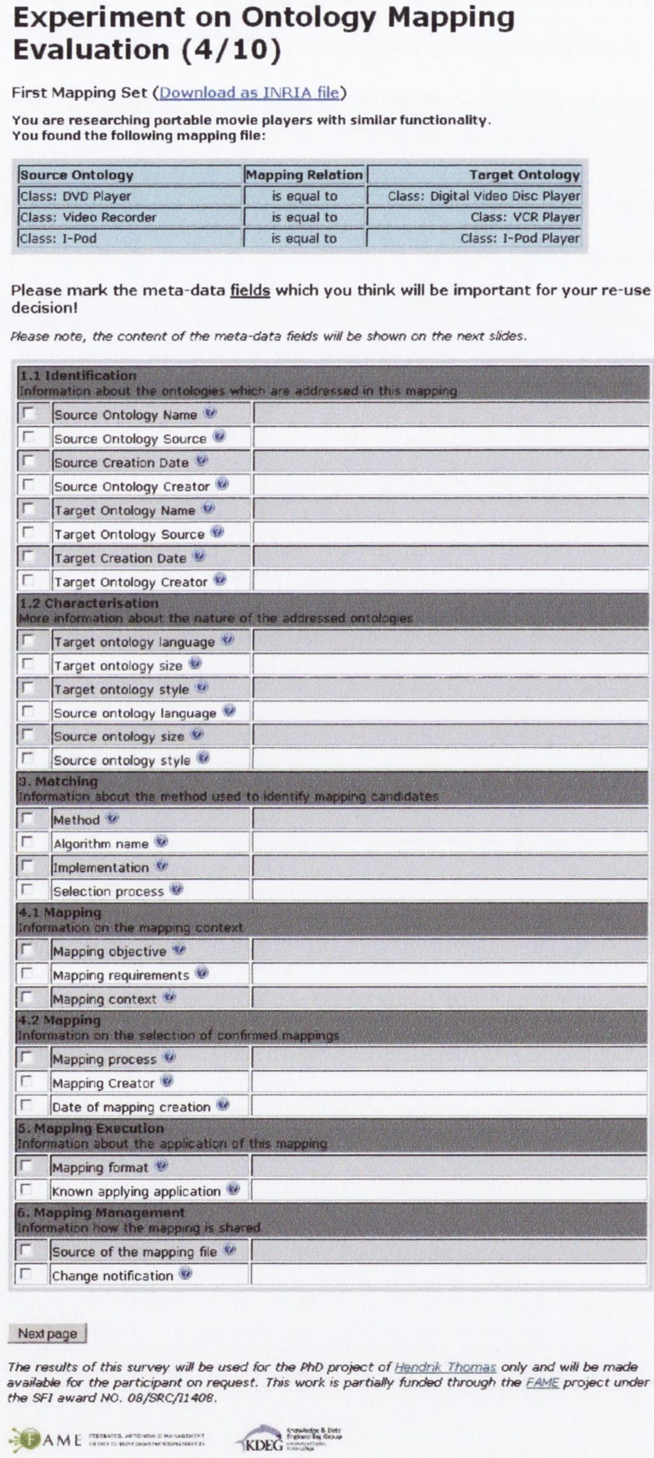

Figure 26 Experiment tool meta-data relevance ranking – only meta-data fields

5. **Subjective Rating:** On the last page, the participant was asked to share their personal opinion about the model in general. Specifically the participant was asked to rate the most useful meta-data field, the least useful and the most difficult to understand meta-data field. One of the key features of the model that was evaluated was the labels used for the meta-data fields. Easy to understand labels are essential. Therefore the participant were asked to judge for every meta-data field if the participant such "like", "dislike" or has "no opinion" on the field's label.

The UML activity diagram in Figure 27 summarizes the activities described above in each page:

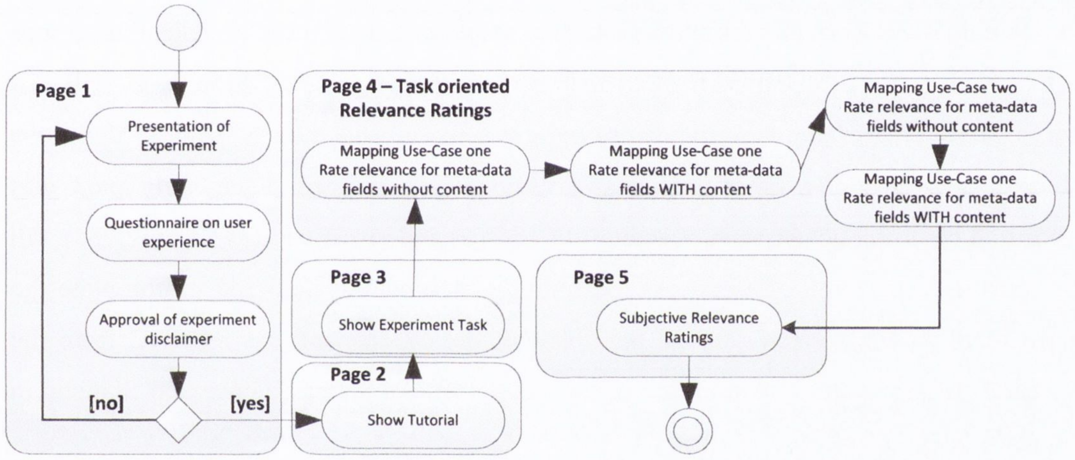

Figure 27 UML activity diagram for experiment 2

Please note a full copy of the experiment interface can be found in the *folder B Experiment 2 Relevance ranking of the OM2R* of the attached DVD.

6.4.3 Participant Target Group

Ontology mapping is a specialised topic (see section 2.2). In this experiment participants would be asked to document two given ontology mapping in the given scenario. At least a basic understanding of ontologies mappings would be helpful but not required. The reason for this is that it cannot be assumed that all OM2R participants have such experience in ontology mapping.

Participants from the university institutions were targeted for this experiment. The institutes were chosen which have known publication record on ontology matching, mapping experience and Semantic Web.

More specifically students, PhD students and post-graduates from the following institutes and communities were invited to volunteer to participate: Knowledge and Data Engineering group Trinity College (Ireland)[89], Information and Knowledge group, Technical University of Ilmenau (Germany)[90], Computer science group, Massey University of Auckland (New Zealand)[91] and W3C Linked Open Data Project Mailing List[92].

Please note there were no conflicting interests within or between the participating groups with regards to this experiment. Participants were recruited through a call for participation email. The only mandatory requirement was that a participant needed to be 18 years or older. No other exclusion criterion was enforced. Each participant was identified by a unique access time stamp, IP address and email address (if provided). All participants were required to access the online experiment system and to complete the experiment in one session. Ethical approval for the experiment was received from the Research Ethics standing committee of the School of Computer Science and Statistic in Trinity College Dublin.

6.5 Results of Experiment

In this section the results of the experiment are presented. More specifically, the details of the participants as well as the results of the task oriented and subjective relevance ratings are discussed.

---

[89] See http://kdeg.cs.tcd.ie/ for contact details.

[90] See http://www.tu-ilmenau.de/ for contact details

[91] See http://www.massey.ac.nz/massey/home.cfm for contact details.

[92] The public-lod@w3.org mailing list provides a discussion forum for members of the Linking Open Data project and the broader Linked Data community. The Linking Open Data project is a grassroots community effort founded in February 2007 as a W3C Semantic Web Education and Outreach Interest Group Community Project.

6.5.1 Participant Experience Profile

In total 49 participants participated in this experiment and provided relevance ratings.[93] The questionnaire regarding ontology creation experience shows that 62 % of the participants had less than one year experience, 16 % had between one and two year, and 22 % had three or more years of experience.[94] Overall the majority of participants (62 % of all participants) had less than one year experience with ontologies.

The next relevant aspect was the experience of the participants with ontology mapping. The questionnaire showed that 2 % of the participants provided no answer to this question, 80 % had less than 1 year experience in mapping/matching creation, 8 % had between one and two year experience, and 10 % had more than three years of experience. The responses show that the majority of participants (80 % of all participants) had less than one year experience with ontologies mapping and matching. Only a small number (18 %) of the participants had more than one year experience.

The third aspect which was analysed was the experience of the participants with ontology mapping reuse. This diagram shows the majority of participants (90 % of all participants) have never reused a mapping at all. Only 10 % of the participants had reused a mapping at least once.

According to the participant classification outlined in the methodology section, the participants can be split into 69 % novices and 31 % experts.

6.5.2 Task-oriented relevance rating

Table 16 provides the results of the task oriented relevance rating where participants had to judge if the individual OM2R meta-data fields are relevant for a reuse decision or not. The data shown is for both scenarios. For each scenario the participant had to rate the relevance twice, once when presented only with the meta-data field name and again time when the field and a content selection was presented. For each meta-data field the total number of participants are listed, who rated the fields as relevant. The percentage is

[93] Please note 50 participants in total accessed the experiment web page. One user filled in all information about his experience but did not provide any relevance rankings.

[94] Please note all participants provided an answer to this question.

calculated by the sum of all relevance ratings per field / total sum of relevance ratings * 100 to provide a normalized relevance rating.

The last column "Overall Mean Relevance %" shows the mean of all normalized % relevance ratings and the list of meta-data fields is sorted according to this ratio. This allows an aggregation of the 4 relevance ratings each user could provide per field.

Table 16 Task oriented relevance ranking of the OM2R meta-data fields

#	OM2R field name	scenario 1 (only fields displayed)	%	scenario 1 (fields & content displayed)	%	scenario 2 (only fields displayed)	%	scenario 2 (fields & content displayed)	%	Overall Mean Relevance %
1	Source Ontology Name	44	90 %	34	75.6 %	33	80.5 %	29	70.7 %	79.14 %
2	Mapping Process	35	71 %	35	77.8 %	35	85.4 %	28	68.3 %	75.72 %
3	Target Ontology Name	39	80 %	32	71.1 %	29	70.7 %	30	73.2 %	73.65 %
4	Source of the Mapping File	28	57 %	33	73.3 %	30	73.2 %	27	65.9 %	67.38 %
5	Matching Selection Process	31	63 %	28	62.2 %	28	68.3 %	30	73.2 %	66.74 %
6	Mapping Requirements	32	65 %	30	66.7 %	30	73.2 %	24	58.5 %	65.92 %
7	Matching Method	31	63 %	31	68.9 %	29	70.7 %	24	58.5 %	65.36 %
8	Target Ontology Language	35	71 %	26	57.8 %	28	68.3 %	24	58.5 %	64.01 %
9	Mapping Objective	30	61 %	26	57.8 %	31	75.6 %	25	61.0 %	63.90 %
10	Source Ontology Source	32	65 %	26	57.8 %	30	73.2 %	23	56.1 %	63.09 %
11	Mapping Language	28	57 %	28	62.2 %	28	68.3 %	25	61.0 %	62.16 %
12	Source Ontology Language	34	69 %	26	57.8 %	26	63.4 %	22	53.7 %	61.06 %
13	Target Ontology Source	28	57 %	26	57.8 %	30	73.2 %	23	56.1 %	61.05 %
14	Mapping Context	31	63 %	20	44.4 %	27	65.9 %	28	68.3 %	60.46 %
15	Change Notification	26	53 %	30	66.7 %	25	61.0 %	25	61.0 %	60.42 %
16	Target Ontology Design	24	49 %	24	53.3 %	26	63.4 %	25	61.0 %	56.68 %
17	Matching Algorithm	27	55 %	26	57.8 %	27	65.9 %	16	39.0 %	54.44 %
18	Source Ontology Design	23	47 %	22	48.9 %	26	63.4 %	21	51.2 %	52.62 %
19	Source Ontology size	19	39 %	24	53.3 %	20	48.8 %	22	53.7 %	48.64 %
20	Target Ontology size	19	39 %	23	51.1 %	19	46.3 %	22	53.7 %	47.47 %
21	Known Applying Application	23	47 %	21	46.7 %	21	51.2 %	18	43.9 %	47.18 %
22	Source Creation Date	22	45 %	16	35.6 %	18	43.9 %	17	41.5 %	41.45 %
23	Matching Algorithm Implementation	14	29 %	20	44.4 %	20	48.8 %	16	39.0 %	40.21 %
24	Target Creation Date	19	39 %	14	31.1 %	17	41.5 %	17	41.5 %	38.20 %
25	Date of Mapping Creation	15	31 %	11	24.4 %	15	36.6 %	14	34.1 %	31.45 %
26	Source Ontology Creator	23	47 %	10	22.2 %	12	29.3 %	10	24.4 %	30.70 %
27	Mapping Creator	16	33 %	9	20.0 %	13	31.7 %	8	19.5 %	25.97 %
28	Target Ontology Creator	14	29 %	8	17.8 %	11	26.8 %	8	19.5 %	23.17 %

Based on the Overall Mean Relevance it can be identified that on average that the ontology names (source ontology name 79.14 % and target ontology name 73.65 %), mapping process (73.65 %), Source of the mapping file (67.38 %) and Matching Selection process (66.74 %) related fields have been selected by most participants as relevant. In contrast the field related to details about the creator of the ontologies (source and target ontology) and mappings have been rated by the least amount of participants. Furthermore, details about the mapping (25.97 %) and source creator (23.17 %) are rated on average by the lowest number of participants in this experiment.

6.5.3 Subjective Relevance Rating

On the final page of the experiment, each participant was asked to share subjective feedback about the relevance of the meta-data fields. For this purpose the participants were asked about each OM2R field as to which is most useful, least useful and most difficult. Table 17 shows the responses with respect to the ratings of the most useful field. For each field it is specified how often it was selected by experts, novices and in total. In addition, a ratio is provided. The ratio is calculated by the sum of all amount of selections per field / total sum of selections * 100.

Table 17 Most useful rated field of the OM2R

Subjective Most useful	Experts	Novices	Total	% of total
Ontology Name	6	4	10	18 %
Mapping Process	3	1	4	7 %
Matching Method	0	3	3	5 %
Ontology Source	0	2	2	4 %
Ontology Design	1	1	2	4 %
Mapping Objective	1	1	2	4 %
Matching Algorithm	0	2	2	4 %
Matching Selection Process	0	1	1	2 %
Ontology size	1	0	1	2 %
Sum of Ratings given			27	

The ontology name and the mapping process fields were rated as most useful. Table 18 presents the number of participants that rated the meta-data field ratio as least useful for the mapping reuse application.

Table 18 Least relevant rated field in the OM2R model

Subjective least useful	Experts	Novices	Total	% of total
Creation Date	2	5	7	24 %
Ontology Creator	4	1	5	17 %
Mapping Creator	2	3	5	17 %
Matching Algorithm Implementation	0	2	2	7 %
Ontology Size	1	1	2	7 %
Ontology Source	1	1	2	7 %
Ontology Name	0	1	1	3 %
Source of the Mapping File	0	1	1	3 %
Known Applying Application	0	1	1	3 %
Date of Mapping Creation	1	0	1	3 %
Mapping Language	1	0	1	3 %
Change Notification	1	0	1	3 %
Sum of Ratings Given			29	

Details about the creator and regarding the creation time have been rated by most participants as least useful. Table 19 shows the subjective rating for the most difficult meta-data field of the OM2R

Table 19 Subjective rated as the most difficult meta-data fields

Subjective most difficult	Experts	Novices	Total	% of total
Known Applying Application	2	2	4	19 %
Mapping Process	1	2	3	14 %
Matching Selection process	3	0	3	14 %
Ontology Design	1	1	2	10 %
Mapping Requirements	1	1	2	10 %
Date of Mapping Creation	0	1	1	5 %
Ontology Name	0	1	1	5 %
Mapping Context	0	1	1	5 %
Matching Implementation	0	1	1	5 %
Matching Selectors	1	0	1	5 %
Mapping Objective	1	0	1	5 %
Matching Algorithm	1	0	1	5 %
Sum of Ratings given			21	

The fields providing details about known application and process details (matching and mapping) are rated by most participants as most difficult to handle.

6.6 Analysis

In total 49 participants participated in this experiment and 69 % are considered novices and 31 % as experts. The high level of experience with ontologies is not unexpected as target groups have been selected with a known background in Semantic Web research. This is also related to the progress shown within the Linked Data community that the application of ontologies is no longer limited to a small group of experts as more and more people are interested in ways to improve knowledge interoperability. An aspect worth mentioning is that 10 % of all participants have reused an ontology mapping at least once. This not insignificant number supports the claim that ontology mapping reuse is of growing importance and needs to be supported by a suitable infrastructure.

The hypothesis for the experiment stated that the OM2R model meta-data fields are relevant for the mapping reuse. To confirm this hypothesis a relevance ranking list of all proposed OM2R fields was created based on the gathered data. For this purpose the subjective and task relevance ratings are combined to a single relevance rating. More specifically, this metric is calculated by adding up the mean of number of task-relevance selections for each scenario and step. Furthermore, the numbers of selections for most useful fields are added and the numbers of selections for least useful fields are subtracted. Thus the combined rankings show the calculated number of relevance selections. Table 20 presents the result of this approach:

Table 20 Combined relevance ranking based on subjective and task relevance ratings

#	OM2R field name	scenario 1 (fields)	scenario 1 (content)	scenario 2 (fields)	scenario 2 (content)	Mean Count	Most useful Count +	Least useful Count -	Combined Ranking
1	Source Ontology Name	44	34	33	29	35	10	1	44.0
2	Target Ontology Name	39	32	29	30	33	10	1	41.5
3	Mapping Process	35	35	35	28	33	4	0	37.3
4	Matching Method	31	31	29	24	29	3	0	31.8
5	Mapping Objective	30	26	31	25	28	2	0	30.0
6	Matching Selection process	31	28	28	30	29	1	1	29.3
7	Mapping Requirements	32	30	30	24	29	0	0	29.0
8	Source Ontology Source	32	26	30	23	28	2	1	28.8
9	Target Ontology Source	28	26	30	23	27	2	0	28.8
10	Source of the Mapping File	28	33	30	27	30	0	1	28.5
11	Target Ontology Language	35	26	28	24	28	0	0	28.3
12	Source ontology Language	34	26	26	22	27	0	0	27.0
13	Target Ontology Design	24	24	26	25	25	2	0	26.8
14	Mapping Context	31	20	27	28	27	0	0	26.5
15	Mapping Language	28	28	28	25	27	0	1	26.3
16	Matching Algorithm	27	26	27	16	24	2	0	26.0
17	Change Notification	26	30	25	25	27	0	1	25.5
18	Source Ontology Design	23	22	26	21	23	2	0	25.0
19	Source Ontology Size	19	24	20	22	21	1	2	20.3
20	Target Ontology Size	19	23	19	22	21	1	2	19.8
21	Known Applying Application	23	21	21	18	21	0	1	19.8
22	Matching Algorithm Implementation	14	20	20	16	18	0	4	13.5
23	Date of Mapping Creation	15	11	15	14	14	0	1	12.8
24	Source Creation Date	22	16	18	17	18	0	6	12.3
25	Target Creation Date	19	14	17	17	17	0	6	10.8
26	Source Ontology Creator	23	10	12	10	14	0	5	8.8
27	Mapping creator	16	9	13	8	12	0	3	8.5
28	Target Ontology Creator	14	8	11	8	10	0	5	5.3

Tab. 1 Combined relevance ranking based on subjective and task relevance ratings

The top relevant rated fields ranked task-relevant (79.14 %) and subjectively "most useful" (18 % of all participants) are the *ontology names*. This basic descriptive information is available in almost all current ontology mapping representations. This top ranking is not unexpected as it is essential for every reuse decision to know which ontologies are addressed in a mapping and where to find the sources.

The second most task-relevant rated (75.72 %) and third most useful rated field (7 % of all participants) is the *mapping process*, which contains details on how the final mapping correspondences have been generated. This shows the need of participants to understand the mapping lifecycle process to enable them to judge if a mapping can be reused in another context.

It is noteworthy that other fields that are related to the mapping process, describing the *matching selection process* and *matching method* respectively, can also be found in the top 10 of ratings. This indicates that participants consider process-oriented information in general as highly relevant. Such process information is currently not available in existing mapping representation formats (see state of the art section).

In addition, the fields *mapping requirements* and *mapping objective* are rated as highly relevant. This demonstrates that participants consider contextual information important. These fields address the WHY related to a mapping, which complements the process information that addresses the HOW a mapping was generated. Process and contextual information is usually quite complex and cannot be represented in the simple key-value fields which are common in current mapping representation formats. This may explain why mapping process information was rated as the second most difficult field (14 % of all participants). The OM2R model addresses this point, as it allows a participant to express the required information in RDF triples of arbitrary complexity and so can provide the required level of detail and flexibility.

It can be observed a clear trend that all creation date and author related details are considered less task-relevant and often rated as least useful. This could indicate that participants are result-oriented and do not really care who generated or when a mapping was generated. This may be valid for this particular experimental scenario where it may have been an inherent participant assumption that the mapping was applicable for the hypothetical example. In real life cases and in particular in a business integration cases the level of uncertainty is much higher, thus information about how up-to-date a mapping is and details about the author may be considered more useful than in this experiment.

For example a user could avoid obsolete mappings and known untrustworthy sources. The matching algorithm implementation field was rated as relevant by only a quarter of the participants (combined relevance rating of 13.5 out of 49). This could indicate that the level of detail provided was too high for the majority of the participants. This corresponds well with a common trend in IT, where functions or services are often reused without a detailed understanding of the underlying implementations. Again in a commercial setting it may be the case that specific implementations are preferred.

It is also interesting that the field *applying application* received a low rating (combined relevance rating of 19.5 out of 49). At the moment the level of interconnectivity between mapping representations is very low. However, this might change soon as embedding and linking to existing vocabularies is also becoming more and more popular in ontologies.

Another point of interest was the question of whether the relevance ranking given was impacted by the experience level of the participants in ontologies and ontology mapping creation. To identify any statistical significance difference between the relevance ratings given by experts and novices participants, an unpaired two-tailed t-test [Ru06] was calculated.

The t-test was selected as it is one of the most commonly used methods for testing a hypothesis on the basis of a difference between sample means [We10a]. The t-test can be used in this setup as it allows a comparison of two small sets of quantitative data when samples are collected independently of one another. Between the expert and novice users no direct relationship can be identified. As such the samples are collected from two different populations and therefore a test for independent samples (unpaired) is suitable.

Table 21 displays the results of the unpaired t-test.

Table 21 Unpaired two-tailed t-tests for relevance rating given by novice and experts

Test Subject	**Sample Mean for Experts**	**Sample Mean for Novices**	**P-value** ($\alpha = 0.05$)	**Statistical Significance**
Task relevance ranking	0.5627	0.4981	$t = 1.6327, p = 0.1084 > \alpha$	NOT ONE

The data indicates that there is no statistically significant difference between the relevance rating given by experts and novices. This provides evidence that novice and experts had a similar judgement of the relevant information and consequently there is promise that the proposed OM2R model can meet the requirements of both groups. This again is important for the creation of consistent documentation which requires support of expert and novice participants.

6.7 Conclusion

The results are biased, to a degree, as the experiment was conducted in a lab environment and the provided scenario and mapping example were designed on purpose to be simple and easy to understand. In a real life re-use situation different priorities may arise from external factors and can create a different relevant ranking for the field. To mitigate this impact, the experiment was designed with a total of four relevance ratings per field, including an additional subjective ranking.

Obviously the findings must also be tempered with the knowledge that they apply to the specific reuse scenarios investigated in the experiment and the limitations of a finite pool of study participants of 49.

Nevertheless the experiment provides evidence to support the hypothesis that the meta-data fields proposed in the OM2R model are relevant for the ontology mapping reuse application. The calculated overall relevance ranking (as a combination of task and subjective relevance rating) shows that *ontology name* and *process information* are considered most relevant by the participants. Date and creator related details emerged at the bottom of the rating but are still considered relevant by a normalized number of 5.3 out of 49 participants. These results indicate that all meta-data fields of the OM2R model are rated relevant by the participants but the extent of the relevance varies from strong agreement (normalized 44 out of 49 participants) to weak agreement (5.3 out of 49). Overall, participants are highly interested in matching and mapping process information. As a consequence a model design is needed for the OM2R model which enables it to express details related to the process and different process steps. This needs to be combined with details regarding mapping requirements and objectives. The lack of that information that was identified as most relevant in existing mapping representations, demonstrates the potential value of the OM2R proposed approach (see section 3.5). The insights gained in this experiment were used to redesign an early version of the OM2R (section 3.8).

In the next chapter the focus is shifted towards the support for consistency provided by the OM2R model.

7 Experiment 3 – User-based evaluation of the OM2R model to support application consistency and logical consistency

This chapter describes the evaluation undertaken to examine the ability of the OM2R model to support the consistent documentation of mappings. This is achieved by comparing an ontology-based representation of the model with those of an index-based baseline representation. The chapter presents the experiment using the following sub-sections motivation, hypothesis, methodology, setup, results and analysis.

7.1 Motivation

A meta-data model not only needs to support generation of relevant meta-data (see chapter 6), but also consistent meta-data to support mapping retrieval and reuse over time. For example, it should allow users to expect relevant meta-data fields while searching and should help to avoid ambiguity [Ch99]. This leads to the question as to what extent the proposed OM2R meta-data model can support ontology engineers in the creation of consistent documentation of ontology mappings. In this regard it is worth considering that the potential users of the OM2R model will most likely have diverse backgrounds. As such the OM2R model should offer a different level of consistency support for users with less experience in the domain.

Another consideration is that the OM2R model offers a meta-data collection but as any model it can be expressed in a range of different structures, e.g. ranging from ontology-based representations to data centric structures such as thesauri or back of book indices [Be01, No08].[95] The specific structure used to represent the model potentially has an impact on the type and amount of information which can be expressed in the meta-data model. For example it is generally accepted that only an ontology can contain an

[95] An common and well-known example for meta-data models is the Dublin Core [DC] vocabulary to describe topics and author details (e.g. <dc:creator>Rose Bush<:/dc:creator>). It is often used in an index presentation but a more expressive OWL model is available, too [DC]. Another example typically expressed in a thesaurus structure is the Library of Congress Classification (LCC) [Ta09].

explicitly typed relationship.[96] The structure of the model impacts the editing process as users commonly experience the meta-data model via editing tools.

Consequently the chosen representation structure has an impact on the outcome of the documentation process and can impact the level of consistency support the tool can provide. An index-based structure is the simplest and most popularly structured used in ontology mapping meta-data representation (see section 3.2 for details). For example, the Ontology Alignment Format is expressed as a RDF(S) file but provides only key-value meta-data fields. In contrast, the OM2R model is expressed by an ontology rather than an index-based structure. A comparison of both structural approaches will provide insight in the impact of the structure and supply a base line to evaluate the contribution of the ontology-based OM2R model.

7.2 Hypothesis

These considerations can be summarized into the following two hypotheses for experiment 3:

First Hypothesis (H1): The OM2R model can support users in creation of more application and logically consistent documentation for ontology mapping if the model is structured as an ontology rather than an index-based structure.

Second Hypothesis (H2): The OM2R model can offer similar support in the creation of application and logically consistent documentation for ontology mappings for users with high ontology mapping experience, as for those with limited ontology mapping experience.

7.3 Methodology

In this experiment participants were asked to document two given ontology mappings examples (see section 7.4.1 experiment scenario) in an online lab environment by using

[96] It is worth noting that some of the library oriented meta-data models such as Dublin Core and LOM are expressed as ontologies (RDF(S)). However, when analysed they still retain their simple index structure. More specifically each field is linked to a single data type value which contains the field content commonly as a string, e.g. <dc:creator>Rose Bush<:/dc:creator>.

the OM2R model. The generated OM2R instances are compared to a gold standard[97] to calculate the achieved level of application and logical consistency. Please see section 7.4.1 for more information as to how the gold standard is defined.

To investigate if the structure chosen for the OM2R model has an impact on the support for consistency (H1), the experiment was conducted in the following two stages. Each stage of the experiment was performed independently and for each stage a new participation campaign was implemented conducted. A participant could participate in both or only one of the experiment stages. This has little influence on the results as the exact same use-case scenarios and task instructions are used in both experiment steps.

- **Stage 1 – ontology-based model representation:** In the first stage the OM2R model is expressed as an ontology.
- **Stage 2 – index-based model representation:** In the second stage the OM2R model is expressed in a key-value-based index structure.

Compared to an ontology-based approach an index-based structure is less expressive and therefore some information defined in the OM2R model cannot be expressed or can be expressed only with limitations. In this experiment the participants will experience both representations of the OM2R model through two custom-built OM2R editors. The specific feature set of each editor depends on the representation of the OM2R model. Section 7.4.4 Ontology-based and Index-based OM2R Editor presents the details.

This two-stage approach allows a comparison of the achieved level of consistency support if the same OM2R model is expressed using two different structures. The index structure was chosen for this comparison as it is a common way to document meta-data, for example in the librarian and research domains [Ba99]. In addition, existing ontology matching notations (e.g. Ontology Alignment Format) use a similar structure (see section 2.5.4). Furthermore, index structure provides a baseline against which the support which the ontology-based structure can be compared, as explicit meta-data models for ontology mapping or ontology mapping documentation tools are not available (see section 2.7).

[97] Please note the OM2R instance of the model that is used as a gold standard, is fully application and logically consistent. The instance was created by the author of this thesis and the consistency was confirmed by a senior postdoctoral researcher in the Knowledge and Data Engineering Group (Trinity College Dublin).

A full overview of the specific questions that the participant is asked to answer and the given scenario is provided in section 7.4.1. To allow a comparison of the results the same documentation task and the same metrics are used in both experimental stages.[98]

Table 22 provides an overview of the metrics used in the experiment. A detailed definition of the metrics can be found in section 1.4. Note that the term meta-data field selections refers to an individual meta-data field and the specific value assigned to it by the participant. For example, for the meta-data field *source ontology label*, the meta-data selection "ontology about cars" is suitable.

Table 22 Metrics for experiment 3

Metric	Definition
recall (application)	= number of relevant meta-data field selections in the documentation instance / total number of relevant meta-data field selections in gold standard
recall (logical)	= number of relevant meta-data field combinations in the documentation instance / total number of relevant meta-data field combinations in the gold standard
precision (application)	= number of relevant meta-data field selections in the documentation instance / total number of meta-data field selections in the documentation instance
precision (logical)	= number of relevant meta-data field combinations in the documentation instance / total number of meta-data field combinations in the documentation instance

To encourage the participant to utilise the provided meta-data (e.g. alternative names) some of the facts were not presented plainly. In other words the scenario instructions do not provide direct information on all the meta-data fields which need to be used. For example, an instruction states that the mapping of interest is expressed in *INRIA*. However, the interfaces show only an alternative name for this mapping language which is *Ontology Alignment API*. In such a way, the participant is motivated to use the help documentation where all names are listed. Such items will be referred to as "vaguely formulated facts" in the instructions for the scenario.

The second hypothesis (H2) aims to investigate if a difference in consistency results can be identified for participants with experience in ontology mapping and those without. To investigate if the OM2R model can provide a similar level of support for generating consistent documentation, the participants in experiment 3 were split into two groups

[98] . To be more specific for this experiment the same definition of consistency as defined in [Ha07] is used as outlined in the section 1.4.

based on their experience level. Please note that participants were recruited from research and university domains that have an explicit focus on Semantic Web and ontology mapping / matching related research. However, this included participants with and without experience in ontology mapping. Full details on the participants profile can be found in section 7.4.2. The grouping of the participants and subsequent analysis is only completed for stage one of the experiment where the OM2R is expressed as an ontology. The second stage had considerable less participants[99] and serves only as a base line for comparison. More specifically, the two participant groups were defined as follows:

- *Experts*: A participant was allocated to this group, if yes was answered in the ontology mapping documentation experience question (question 2) or if some exposure to ontology mapping and familiar with the topic was indicated (that is if "less than one year", "between one and 3" or "more than 3 year experience" in ontology mapping was chosen in reply to question 1). In summary, members of this group are considered to have some exposure to ontology mapping and have familiarity with the topic.
- *Novice*: A participant was categorised into this group if no experience in ontology mapping documentation (question 2) was indicated or if no statement was made in regards to this question. In addition, a participant was assigned to this group if "no experience" was chosen in question one with respect to ontology mapping or no selection at all was made. In summary, members of this group are considered to have limited exposure to ontology mapping and considered not to be familiar with ontology mapping process.
- Based on this categorisation, 40 % of the participants were assigned to the expert group and 60 % to the novice group.

Experiment 3 is based on specific documentation tasks and the actions a participant can perform are very limited and guided by the provided instructions. The intention was to provide an objective view on the level of consistency support offered by the OM2R model.

[99] 48 participants undertook experiment stage 1 and 24 participants undertook experiment stage 2.

However, all participants experience the model through the provided editors. Both editors used in this experiment were deliberately designed to be basic and task specific. This should help to focus the participants on meta-model rather than additional productivity features that one could imagine a fully-fledged editor could feasibly support. The intention was to allow a simple direct correlation between the participants view on the model as experienced through the system.

It is therefore also of interest to gather the subjective feedback of the participant in this experiment regarding the support for consistency. The subjective feedback of the participants is likely more generic and reflects more the high level view of this overall approach and is less impacted by the specific formulation of the individual task. For this purpose the experiment contained a questionnaire to be completed at the end of the task with the following questions. Possible answers for the questions are on a scale of "Do not agree" (1) to "Fully agree" (5).

- *The system will help me to document information correctly.* This question focusses on the subjective opinion of the participant as to whether the model can help to support application consistency.
- *The system will help me to document compatible characteristics, e.g. the ontology is modelled in OWL and encoded in RDF/XML.*[100] This question focusses on the subjective opinion of the participant as to whether the model can help to support logical consistency.

However, it was considered that insights might still be gained by asking the participants to evaluate the usability of the editors. To measure usability the well-known and standardized "System Usability Scale (SUS)" is used [Br96]. The SUS metric scale was chosen as it contains only ten statements and is therefore relatively quick and easy to apply. In addition, SUS is technology agnostic, which means it can be used to evaluate any type of user interface. More specifically, SUS is based on the following 10 item questionnaire shown in figure 28.

[100] It is worth to note that the example can be considered difficult to understand for novice user with no experience in ontology mapping domain. However, the tool and the provided introduction make it very clear what compatible characteristics mean in this experiment.

1. I think that I would such as to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

Figure 28 SUS questions

The SUS uses the following response format: 1 strongly disagree, 2, 3, 4, 5 strongly agree. A high SUS score indicates a high perceived usability performance.[101]

7.4 Setup

This section describes the experimental setup with subsections: on the scenario, the targeted participants and a description of the OM2R index-based and ontology-based editing tools.

7.4.1 Experiment Scenario

In both stages (stage 1 ontology-based and stage 2 index-based) of the experiment the participants are asked to document mappings within the following two use-case scenarios.

- Use-Case 1: ontology language scenario
- Use-Case 2: matching process scenario

The first use-case scenario (Use-Case 1: ontology language scenario) shown in figure 29 focusses on creating meta-data for the semantic language aspects of an ontology mapping, aspects that are common in the Semantic Web domain [St99]. It is more likely that these concepts are familiar to novice participants even if they are not experienced in ontology matching specifics. The participant was given a textual description of the languages chosen for the target and source ontologies as well as the mapping file. The participant was asked to document the notation and formal language for each of these

[101] The individual SUS score for users is calculated as a combined answer score according to the following logic. For odd items: subtract one from the user response. For even-numbered items: subtract the user responses from 5. This scales all values from 0 to 4 (with four being the most positive response). Add up the converted responses for each user and multiply that total by 2.5. This converts the range of possible values from 0 to 100 instead of from 0 to 40.

elements based on a selection of suitable content options defined in the OM2R model, e.g. om2r:source_ontology om2r:hasNotation om2rL:RDF/XML.

The following Figure 29 shows the specific use-case scenario presented to the participant:

The LOM ontology (source) is expressed in RDF/XML and in the language Ontology Web Language DL. The private ontology (target) of Peter is represented in TURTLE and modelled as an RDF document. The mapping is expressed in INRIA as an XML file."

Figure 29 Use-Case 1: ontology language scenario - Documentation Instructions

Table 23 shows the different OM2R fields and field content which the participants needs to use in order to document the relevant information. The fields that need to be documented according to the scenario are highlighted in bold in table 23 and represent the gold standard

Table 23 Use-Case 1: ontology language scenario - Object of Interest, OM2R Field and Field Content

Object of interest:	**OM2R Field**	**OM2R Field Content**
Source Ontology	Notation	**RDF/XML Syntax**, Turtle, N3, XML Topic Map Notation, LTM
Source Ontology	Formal Language	RDF, RDF(S), OWL, OWL Lite, **OWL DL**, OWL Full, SKOS, Topic Maps
Target Ontology	Notation	RDF/XML Syntax, **Turtle**, N3, XML Topic Map Notation, LTM
Target Ontology	Formal Language	**RDF**, RDF(S), OWL, OWL Lite, OWL DL, OWL Full, SKOS, Topic Maps
Mapping	Notation	**RDF/XML Syntax**, Turtle, N3, XML Topic Map Notation, LTM
Mapping	Formal Mapping Language	OWL, Topic Maps, EDOA, **Ontology Alignment Format**

To encourage the participant to utilise the provided meta-data (e.g. alternative names) some of the facts were intentionally formulated vaguely in the instructions for scenario 1. More specifically, the instruction states that the mapping is expressed in INRIA but the default name for the field option is Ontology Alignment API. Furthermore, the file is expressed in XML but the correct field option for the notation has the default label XML / RDF. To document the information according to the gold standard the participants have to use the extended views (see section 4.4.2 OM2R Editor for details) to show all acronyms. In other words the scenario instruction does not provide direct information on what needs to be used for these meta-data fields. This means 2 out of a total of 6 facts are formulated vaguely.

The second use-case scenario (use-case 2: matching process scenario) focusses on the ontology matching process, which is a more challenging task for the naïve participants, as this topic is more specialised and the concepts involved are less commonly known. The intention again with the formulation of the scenario text was to encourage the participant to use the help documentation and field content options offered by the OM2R model. More specifically, the participant was asked to document a given matching process that consists of two steps. The process was presented in an application neutral pseudo code, e.g. MatchingProcessClass A1 = new ClassStructAlignment(). The participant was asked to document the matching method, tool, scope and algorithm. The participant was provided with the instructions shown in figure 30 which outlines what information was needed to be documented.

Peter required a matching tool which can support multiple algorithms, is fast and flexible. He decides to use the Ontology Alignment API 4.2 to create the mapping with the following code:

```
URI uri1 = http://slor.sourceforge.net/ontology/lom.owl;
URI uri2 = http://modelmapping.org/myonto;
MatchingProcessClass A1 = new ClassStructAlignment();
A1.runThreshold(0.25);
MatchingProcessClass A2 = new StringDistAlignment();
A2.maxThreshold(0.75);
```

Figure 30 Use-Case 2: matching process scenario - Documentation Instructions

Table 24 shows the different OM2R fields and field content which the participants needed to use to document the language information of the object of interest. In this scenario the matching process consists of two separate process steps. Thus the same set of fields is shown twice in the interface - one for the first and one for the second process step. Please note the fields that need to be documented according to the scenario are highlighted in bold in table 24 and represent the gold standard:

Table 24 Use-Case 2: matching process scenario - Object of Interest, OM2R Field and Field Content

Object of interest:	OM2R Field	OM2R Field Content
Step 1 & Step 2 of Matching Process	Matching Method	**Automated Matching**, Manual Matching, Mixed Matching
Step 1 & Step 2 of Matching Process	Matching Tools	MAFRA, TM4J, **Alignment API and Server 3.1**, Alignment API and Server 4.2
Step 1 & Step 2 of Matching Process	Matching Algorithm	**ClassStructAlignment()**, MergeTopicsBasedOnNames() MergeTopicsBasedOnPSI(), **StringDistAlignment()**, StrucSubsDistAlignment(), Name Matching()
Step 1 & Step 2 of Matching Process	Element the algorithm is based on	**RDF(S) Class**; Any RDFS Label; **RDFS Labels for Classes**; RDFS Labels for Properties; Topic Association; Topic Names
Step 1 & Step 2 of Matching Process	Matching Scope	**Matching of all elements**, Partial Matching
Step 1 & Step 2 of Matching Process	Applied Threshold	100 % Similarity Measure, **75 % Similarity Measure**, 50 % Similarity Measure, **25 % Similarity Measure**

The second scenario contained three facts which are formulated vaguely intentionally. More specifically, the "matching scope" was not specified in the instruction but the model contains a compatible relation between the algorithm and the correct matching scope. Similarly the "element is based on" was not specified but again a clear compatible relation exists with the algorithm. Furthermore, the applied threshold is displayed as a numeric value (0.75) but the default name in the field content is shown in percentage "75 % Similarity Measure". This means 3 out of a total 6 facts were deliberately formulated vaguely.

The UML activity diagram shown in figure 31 shows the specific activities the participant need to complete. Stage 1 and step 2 are independent and completed by different participants. In each step the participant needed to document two mapping scenarios in the OM2R model. The same scenarios are used in both stages of the experiment but the difference is the editing tool. In stage 1 the tool is based on the ontology-based representation of the OM2R model and step 2 on an index-based representation of the OM2R model.

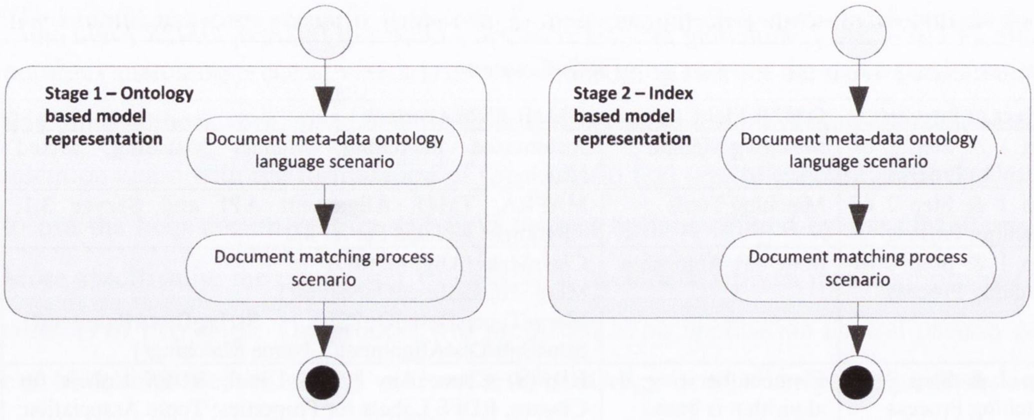

Figure 31 UML activity diagram of user activities

7.4.2 Participant Target Group

Ontology mappings and more specifically, the documentation of ontology mappings are specialised topics (see section 2.2 Ontology Mapping). In this experiment participants were asked to document two scenarios in two stages. At least a basic understanding of ontologies mappings and general Semantic Web principles are helpful but not required. The reason for this is that it cannot be assumed in reality that all users would have such experience in ontology mapping.

Students, PhD students and post-graduates from the following institutes and communities were invited: Knowledge and Data Engineering group Trinity College (Ireland)[102], Information and Knowledge group, Technical University of Ilmenau (Germany)[103], Computer science group, Massey University of Auckland (New Zealand)[104] and W3C Linked Open Data Project Mailing List[105].

Please note there were no conflicting interests within or between the participating groups in regards to this experiment. Participants were recruited through a call for participation email. The only mandatory requirement was that all participants needed to be 18 years or

---

[102] See http://kdeg.cs.tcd.ie/ for contact details.
[103] See http://www.tu-ilmenau.de/ for contact details
[104] See http://www.massey.ac.nz/massey/home.cfm for contact details.
[105] The public-lod@w3.org mailing list provides a discussion forum for members of the Linking Open Data project and the broader Linked Data community. The Linking Open Data project is a grassroots community effort founded in February 2007 as a W3C Semantic Web Education and Outreach Interest Group Community Project.

older. No other exclusion criterion was enforced. Each participant was identified by a unique access time stamp and their IP address. All participants were required to access the online experiment system and to complete the experiment in one session. Ethical approval for the experiment was received from the Research Ethics standing committee of the School of Computer Science and Statistic in Trinity College Dublin.

7.4.3 Experiment System

The experiment was implemented as an interactive web page based on PHP and JavaScipt. The open source framework Apache Jena [Ap12] was used to query the OM2R model via SPARQL queries to generate the interfaces and help documentation pages of the two editing system used. The metrics defined in section 1.4 were calculated by the tool and stored in a collection of CSV files on the server. The ontology language scenario was conducted in March 2012. The matching process scenario was undertaken in September 2012. A copy of the specific experiment web page used can be found in folder *C Experiment 3 Evaluation of the OM2R for consistency support* in this thesis DVD.

Overall the experiment system consists of five individual web pages. The UML activity diagram in figure 32 provides an overview of all pages and the involved activities. The same process is used for both experiment stages. In short the tool is used twice once for stage 1 (ontology-based model representation) and another time for stage 2 (index-based model representation). The only difference is that in the first stage an ontology-based editor is used and in the second stage an index-based editor is used. Section 7.4.4 has a detail description of the editors. The online version of the tool can be accessed on:

http://ontologymappingdocumentation.com.escalade.mochahost.com/2012_Experiment_3_v10/1_intro.jsp

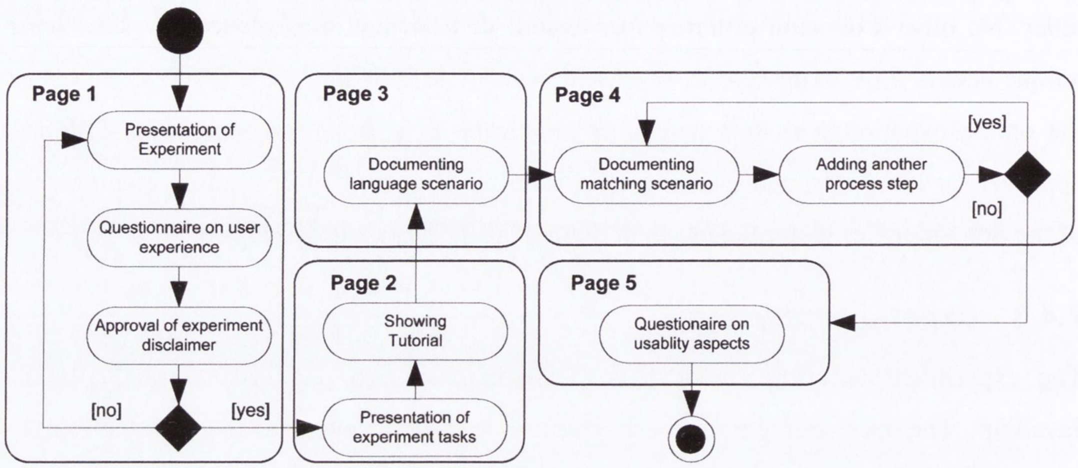

Figure 32 Activity diagram for experiment 3 system

On the first web page the experiment is presented to the participants and it is necessary the participants accept the experiment's disclaimer to continue. Furthermore, the following questions are asked to gather information on the experience level of the participant: How many years of experience do you have in creation of ontology mappings? The participant can choose not to give an answer, no experience, less than one year, between one and two year, three or more years. Have you ever documented the life cycle of an ontology mapping? The participant can choose not to give an answer, yes and no. The participant can also enter an email address which is optional.

The second page shows a short slide tutorial show which introduces the participant to the objective of the research and provides an overview of the functions of the OM2R editor. In addition, the experiment task is presented with the following statement: "We will present you with two specific mapping documentation tasks. Please document all the information presented by using the interface provided." The participant is also shown a PowerPoint presentation Web pages 3 and 4 focus on the actual documentation task. The participant is presented with a short textual description of the documentation scenarios (see section 7.4.1 Experiment Scenario for details). Web Page three presents the scenario focussed on language aspects and web page four the scenario focussed on the matching process. In the last page, the participant is asked a series of standardized questions based on SUS in regards to the usability aspects of the tool. In addition, the subjective feedback question regarding application and logical consistency are placed in the same questionnaire.

7.4.4 Ontology-based and Index-based OM2R Editor

In this experiment, two purpose built OM2R editors were used. Table 25 provides an overview of the functional differences of both editors

Table 25 Overview of the function provided by the OM2R editors

Function	**Experiment Stage 1 Ontology-based Editor**	**Experiment Stage 2 Index-based Editor**
Scenario display	Yes	Yes
Automated interface generation	Yes	Yes
Editing Workflow[106]	Yes	No
Different view types for meta-data	Yes	No
Single external help document	No	Yes
Highlighting of compatible relations	Yes	No

In the following sections the two editors are discussed in more detail.

7.4.4.1 Ontology-based OM2R editor

Figure 33 shows a screen shot of the ontology-based OM2R editor used in this experiment. The individual interface items are labelled with numbers and are used in the subsequent explanation.

[106] No additional descriptive attributes can be attached directly to the meta-data field in an index based representation. As such no workflow or compatible relations can be documented and used in the editor which is purely based on the model.

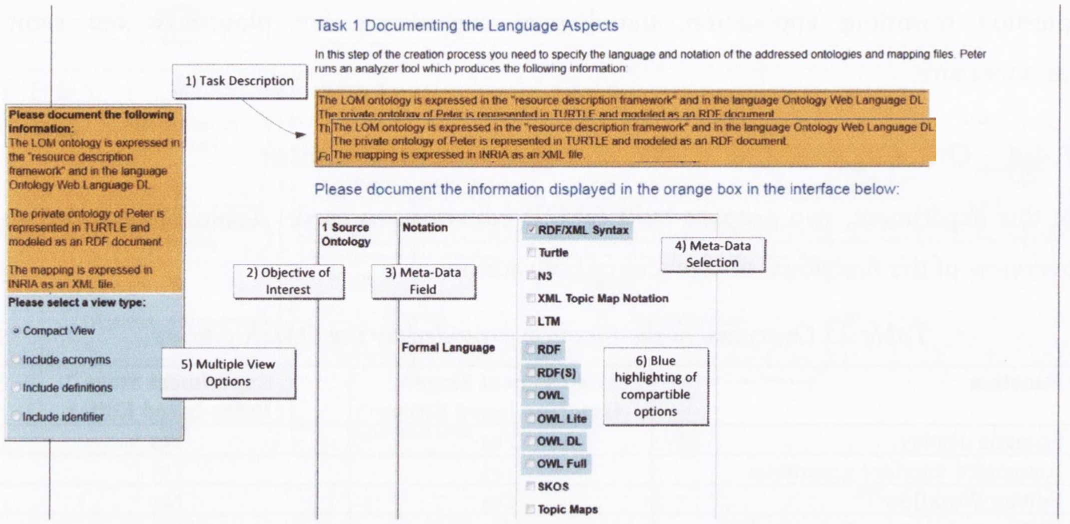

Figure 33 OM2R Editing Tool Screen shot

The first feature of the ontology-based OM2R editor is the automatic generation of the editor interface based on the OM2R model. The editor displays the documentation instructions (1) to the participants that define the scenario information which needs to be documented. The relevant meta-data fields are extracted from the OM2R model automatically and displayed in a text-based interface. The model defines the default name of each meta-data field element, and the sort order of the meta-data fields in the interface are based on a weighting number. The weighting numbers are descriptive fields linked to each meta-data field and used to define an explicit editing work-flow for the editor. In summary, the editor generates the interface as a view of the OM2R model. The interface provides the participants with the relevant objects of interests (2) that the participant needs to use in order to document an ontology mapping. In addition, the relevant meta-data fields (3) and suitable field options (4) are displayed, e.g. XML/RDF, Turtle and so on. The participant needs to tick the appropriate option to express that the meta-data field content is relevant for the given scenario. For example the source ontology is expressed in the notation XML/RDF.

The second feature of the editor is the extensive help information made available to the participant. The help information is based directly on the help documentation in the OM2R model. For each meta-data field and field selection option, the OM2R model provides a default name, a list of acronyms, a short textual definition as well as a URI as a unique identifier (5) for each model element. The editor extracts this meta-data information from the OM2R model and presents the data to the participants as different view types. Each view type is focussed on particular meta-data information and the participant can freely choose between those view types. The following view types are available: show only default names, show all acronyms, show all definitions, and show unique identifier of the relevant subjects. Figure 34 and 35 provides examples of two view types and their presentation in the tool.

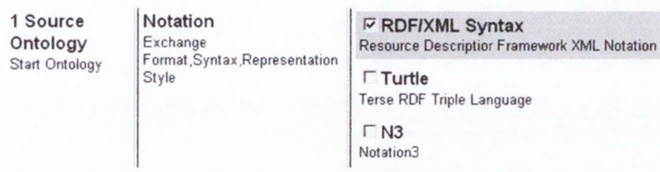

Figure 34 Example of a default name view

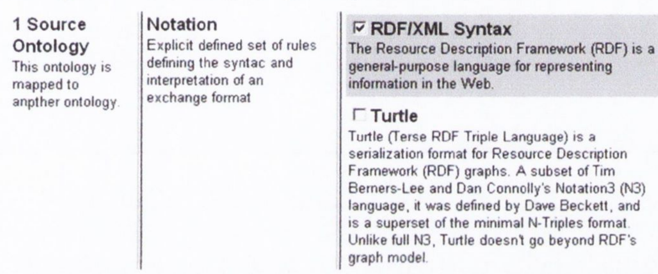

Figure 35 Example of the meta-data view including definitions of the elements

The third feature is that the tool recommends compatible relations (6). The OM2R contains relations which define what field options are compatible to each other. These compatible relations are indicated by displaying using a dynamic blue highlighting. For example if the participant selects the option for notation XML/RDF then the system detects automatically from the meta ontology that the formal language RDF is compatible and thus, both elements are highlighted blue.

7.4.4.2 Index-based OM2R editor

The index-based editor presents the scenario information to be used in the documentation task to the participant. The layout is identical to the ontology-based editor in order to allow a comparison of the results of both experiment stages. The meta-data fields are extracted from the OM2R model automatically and displayed in a text based interface as shown in figure 36. In a flat index structure no weighting information can be represented for each field and therefore this interface shows the meta-data fields as an alphabetically ordered list.

Please note the tool can be access online on:

http://ontologymappingdocumentation.com.escalade.mochahost.com/2012_experimentcf/1_intro.jsp

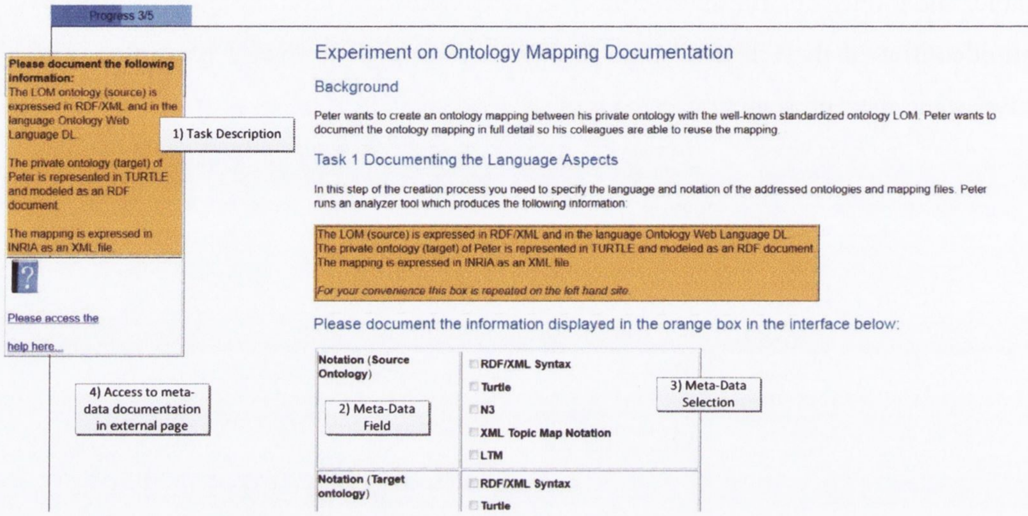

Figure 36 Index-based OM2R Editing Tool

As there are no underlying relations representable between the meta-data fields in the index-based structure, help information is not dynamically generated from the model. Instead the participant is provided a link to a special page where all meta-data fields and the corresponding meta-data are listed (4). This mimics the common way in the Semantic Web domain where help documentation for languages are provided on central resources, e.g. RDF primer [RDF02]. The meta-data fields shown in this editor are listed in alphabetical order as well (2+3). Figure 37 provides a screenshot of a sample of help documentation used in the experiment.

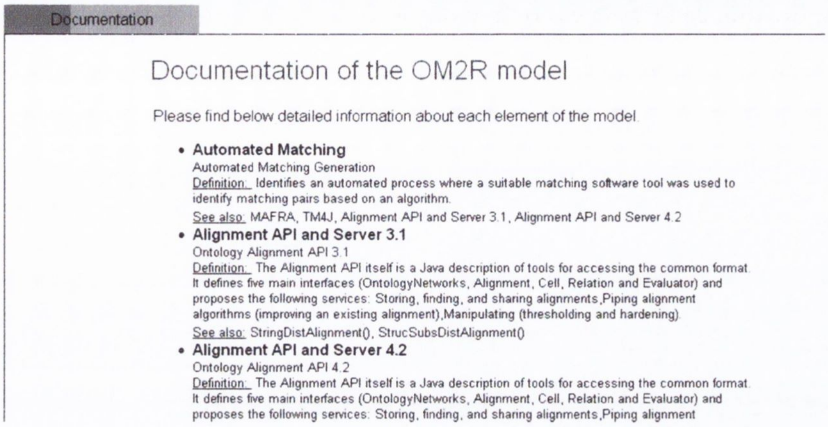

Figure 37 Index-based OM2R Editor - Help Documentation page

The OM2R model representation with an index style used in the experiment cannot contain relations such as compatibility. However, in libraries it is not uncommon to provide simple relations such as "see also" [Ba99]. This is a rather generic relation but often used in back of book indices. In this experiment this "see also" relation is used to express the compatibility relation between meta-data fields. This is undertaken with the intention to provide the participants with as much as possible of the OM2R content within the limitation of the representation structure. More specifically, in the help documentation, under each meta-data element all compatible relations are displayed as "see also" relations.

7.5 Results of Experiment

This section presents the experience profile of the participants and presents the application and logical consistency measurements gathered when the participants created meta-data during the documentation of ontology mapping tasks. The raw data for this experiment can be found in the attached DVD in the folder C *Experiment 3 Evaluation of the OM2R for consistency support*.

7.5.1 Participant Experience Profile

Experiment 3 was conducted in 2 stages (ontology-based model and index-based model) representation. In total 48 participants completed the first stage of the experiment based on the ontology-based OM2R editor.[107] 24 participants completed the second stage of the experiment with index-based OM2R editor.[108] Note that there were 6 months between the two stages and that some of the same participants in stage 1 participated also in stage 2. However, an exact ratio for the overlap cannot be provided as the participation was anonymous.

In the first step of each experiment stage the participants was asked about their experience in ontology mapping. The histogram in Figure 38 shows the distribution of the individual experience levels in relation to the corresponding percentage of participants for the ontology-based stage.

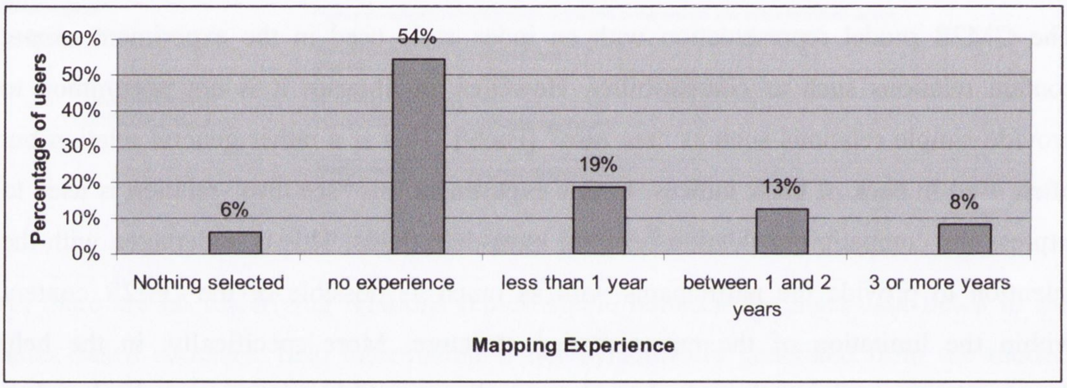

Figure 38 Participant experience distributions for stage 1 - ontology-based

The diagram shows that the majority of participants (54 %) had no experience in ontology mapping at all. In addition, the participants were asked if they had documented the lifecycle of an ontology mapping before. 10 % of the participants stated that they have documented at least one ontology mapping and the majority (83 %) has never performed such a task. 6 % of the participants did not answer this question.

The histogram in figure 39 shows the distribution of the individual experience levels for the second experiment stage which is based on the index structure.

---

[107] According to log files the experiment system was accessed 61 times. Five of these accesses are linked to a test user used to check if the experiment was accessible. These entries have been excluded from the result set.

[108] Please note 26 users accessed the experiment web page and approved the experiment participation form. However, they did not proceed beyond the tutorial page and provided no relevant experiment data. This means 92.3% of the users who accessed the experiment page actually documented some or all of the experiment tasks. According to log files one additional access was registered which is linked to a test account used to check if the experiment was accessible. This entry is excluded from the result set.

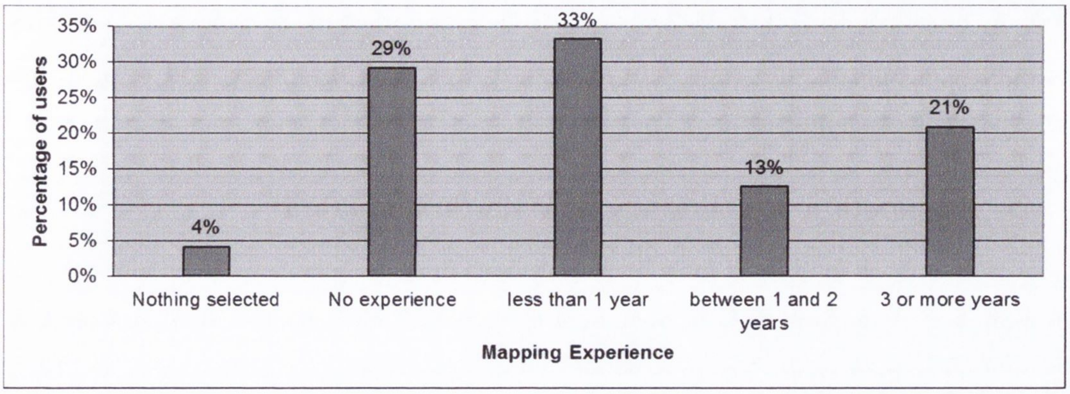

Figure 39 Participant experience distributions for stage 2 – index-based

Figure 39 shows that the majority of participants (67 % of 24 participants) stated that they have at least some level of experience in ontology mapping. In addition, the participants were asked if they have documented the lifecycle of an ontology mapping. 4 % of the participants stated that they have documented at least one ontology mapping lifecycle and the majority (92 %) has never performed such a task. 4 % of the participants did not answer this question.

This data shows that participants in each experiment stage had different levels of experience with ontology mappings and their documentation. This diversity is expected as ontology mapping is a much specialised domain and still not a mainstream topic in the world of the Semantic Web [Sh12]. Based on the difference in the experience level it can be assumed that the documentation in terms of completeness, correctness and inconsistency is likely to be different between these groups. Thus support for creation of consistent documentation is needed to support mapping reuse and discovery over time as both groups will likely be participating in generating and using the meta-data in reality.

6 % for stage 1 and 4 % of the participant did not select anything for their mapping experience. It is unclear why but it can be assumed that the participant simply did not know what ontology mapping was or simply had no experience at all.

Based on the categorisation outlined earlier, for analysis purposes, 40 % of the participants were assigned to the expert group and 60 % to novices for stage 1 of the experiment. The results from the mapping documentation undertaken by the members of each group was used to investigate H2 in terms of whether a higher level of experience with this domain has an impact on the achievement of consistent meta-data generation for ontology mappings using the OM2R model. The second stage of the experiment provides a baseline to compare the consistency support offered by the model and as such the participants are not grouped.

7.5.2 Application consistency of Meta-data created

The first focus of this experiment is on application consistency, which addresses the question: are participants able to document the correct field options when compared to the gold standard? Table 26 provides an overview of the mean of the recall for both experiment stages. The experiment tool documented the meta-data field sections made by the participants in CSV files. The data was then processed in excel to calculate the means and precision based on the gold standard figures automatically. The excel table can be found in *folder C Experiment* 3 of the DVD. The mean was chosen as a well-known method to measure of the central tendency in the distribution of application consistency. This provides a summary number to compare the performance of both tools.

Table 26 Comparison of mean precision and recall for application consistency in experiment stage one and two

Metric	**Experiment Stage 1 (ontology based representation)**	**Experiment Stage 2 (index based representation)**
Mean recall	0.78	0.354
Mean precision	0.818	0.521

Figure 40 and Figure 41 present the distribution of the recall results across the participants for the different tasks in both experiment stages related to measurement of application consistency.

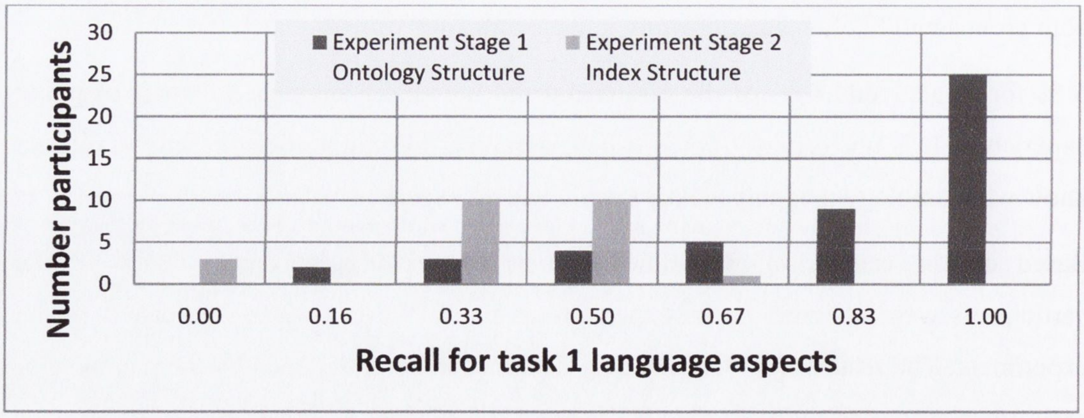

Figure 40 Recall for application consistency for ontology language scenario

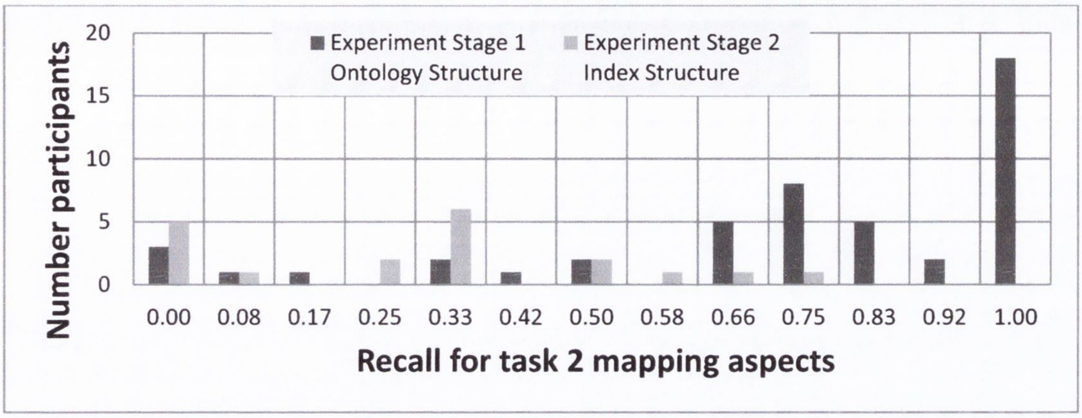

Figure 41 Recall for application consistency for matching process scenario

Figure 40 and Figure 41 shows a high spread across the recall value range. For the ontology-based structure more participants can be found on the right side with high recall values and for the index structure to the left with low recall values.

Figure 42 and Figure 43 both show a high spread across the precision value range but for the ontology-based structure more participants can be found on the right side and for the index structure on the left. Both documentation tasks contain a set of facts which were intentionally formulated vaguely. Please see chapter 7.4.1 Experiment Scenario for details.

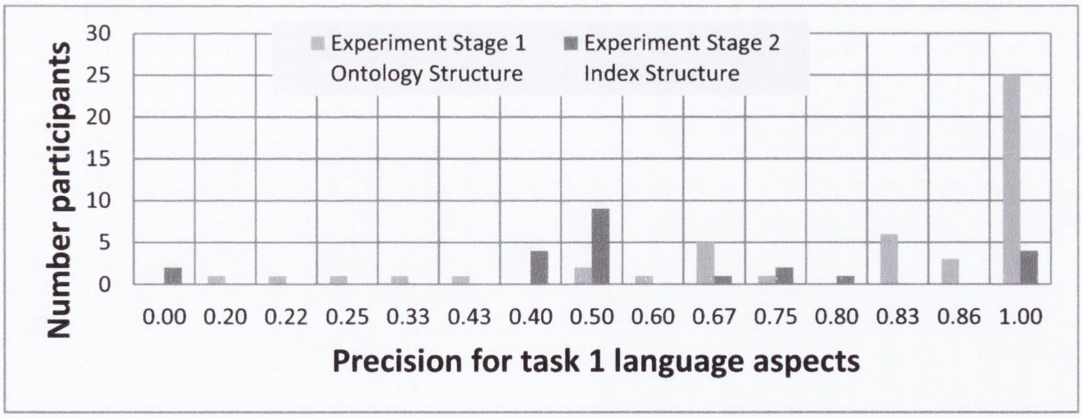

Figure 42 Precision for application consistency for ontology language scenario

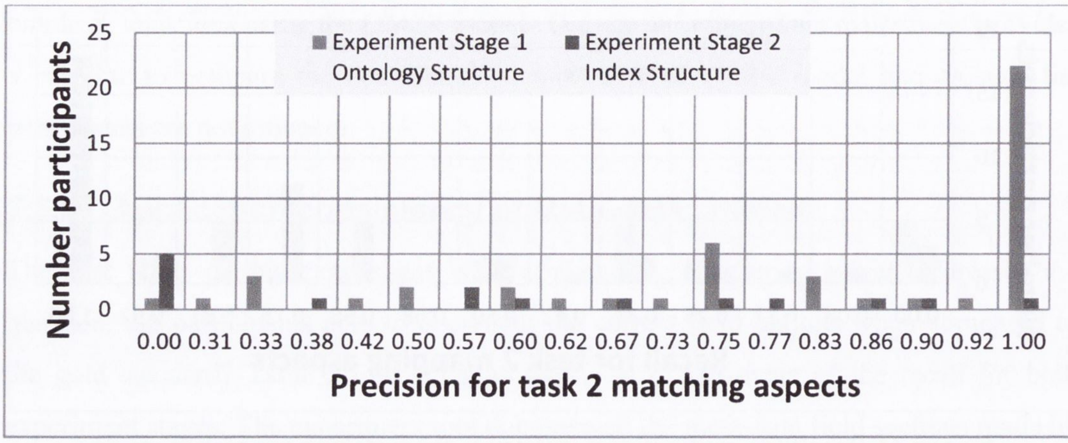

Figure 43 Precision for application consistency for matching process scenario

Tables 27 and 28 provide details of the distribution of the recall results across the participants for both documentation tasks, but focusing on the results for the vaguely formulated facts.

Table 27 Recall results for vague formulated facts for ontology language scenario

	Percentage of Participants	
Number of vague document facts	Experiment Stage 1 Ontology based representation	Experiment Stage 2 Index based representation
0 out of 2	8 %	71 %
1 out of 2	29 %	29 %
2 out of 2	63 %	0 %

Table 28 Recall results for vague formulated facts for matching scenario

	Percentage of Participants	
Number of vague document facts	Experiment Stage 1 Ontology based representation	Experiment Stage 2 Index based representation
0 out of 3+F110	15 %	74 %
1 out of 3	15 %	21 %
2 out of 3	19 %	5 %
3 out of 3	52 %	0 %

7.5.3 Logical consistency of meta-data created

The second aspect investigated in the experiment relates to logical consistency, which focusses on the question: are participants able to document logically consistent relations between the field options? Table 29 provides an overview of the mean of the recall for both experiment stages:

Table 29 Mean Precision and recall for logical consistency

Metric	Experiment Stage 1 (ontology based structure)	Experiment Stage 2 (index based structure)
Mean recall	0.86	0.18
Mean precision	0.85	0.45

Figure 44 and figure 45 show a high spread across the recall value range of the different tasks for logical consistency.

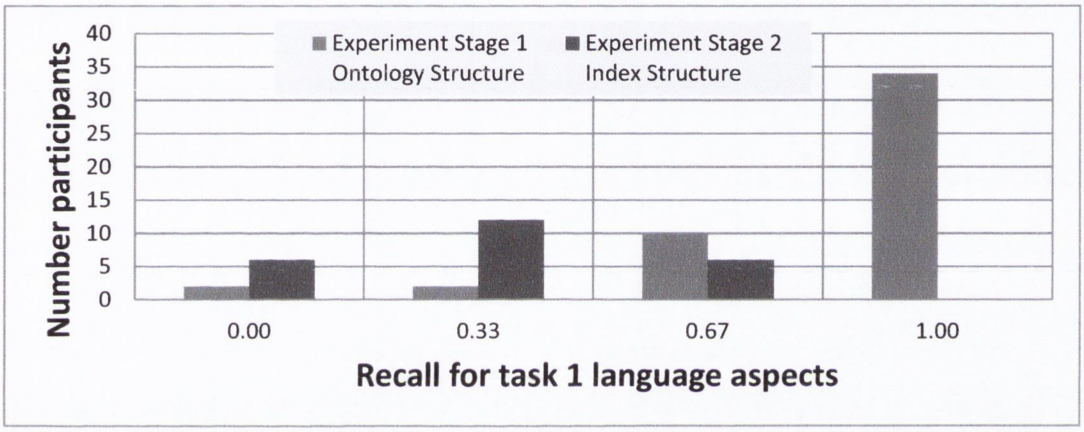

Figure 44 Recall for logical consistency ontology language scenario

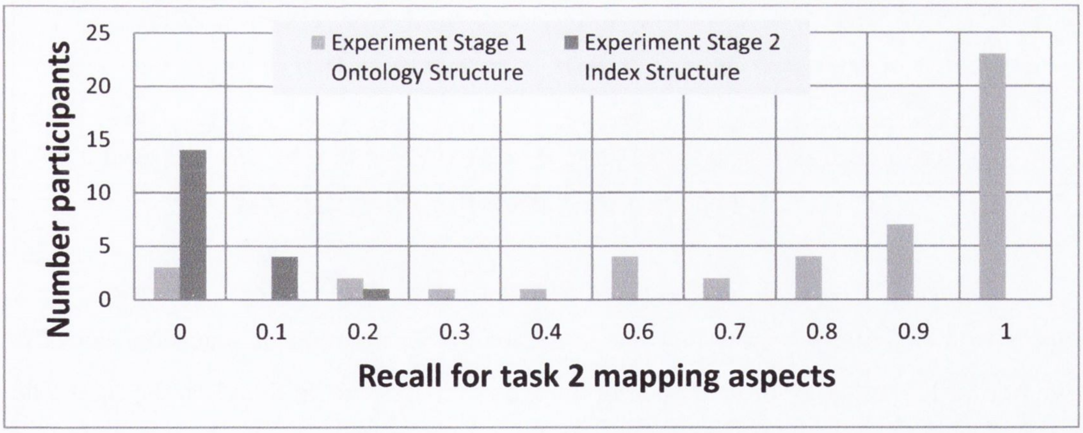

Figure 45 Recall for logical consistency for matching process scenario

Both Figure 44 and figure 45 show a high spread across the recall value range but for the ontology-based structure more participants can be found on the right side and for the index structure to the left particular for task 2 where more relations had to be documented (10 in stage 2 compared to 3 in stage 1).

Figure 46 and 47 show the distribution of the participants across the precision value range of the different tasks for logical consistency.

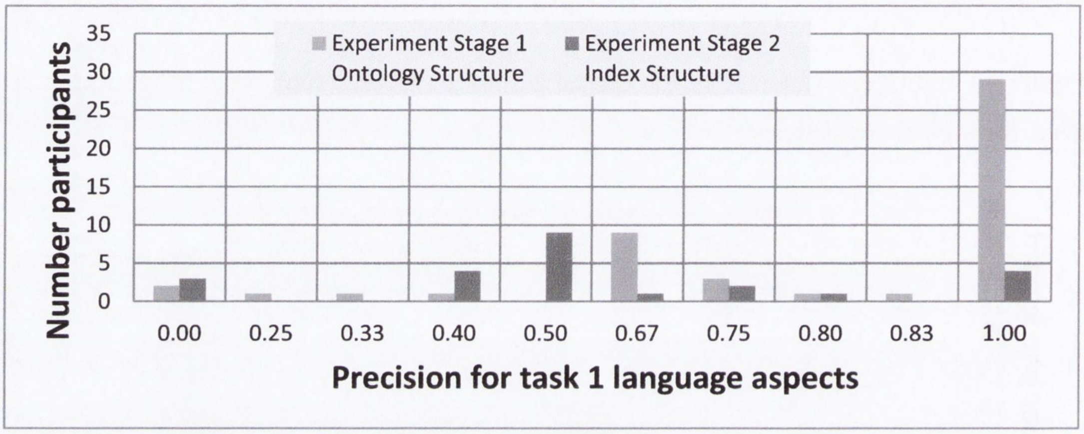

Figure 46 Precision for logical consistency for ontology language scenario

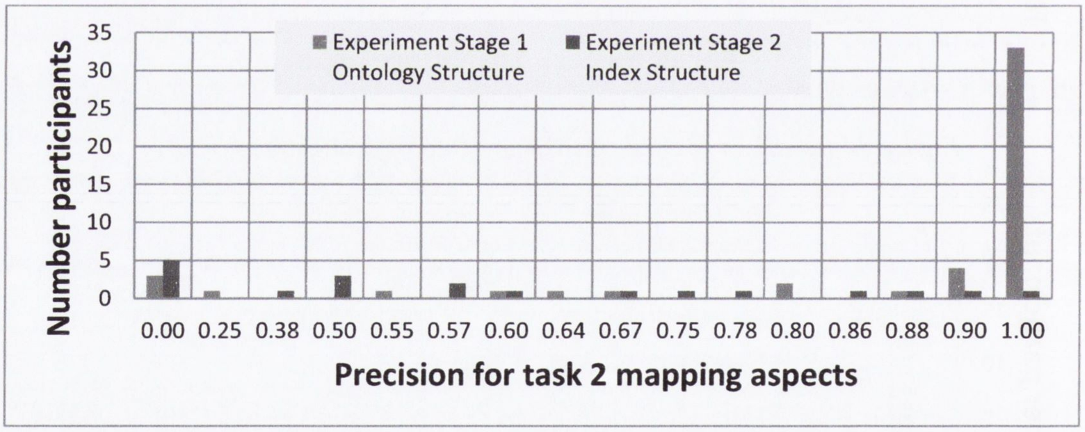

Figure 47 Precision for logical consistency for matching process scenario

Figure 46 and figure 47 show that the spread of the distribution of precision values is high but for the ontology-based structure more participants can be found on the right side and for the index structure on the left. A particularly strong left distribution can be found in task 2 for the ontology structure based tool.

7.6 Usability results

For the first stage of the experiment an ontology-based OM2R editor was used. An average SUS score of 65.1 out of a maximum of 100 was calculated for this stage. If the individual groups are considered an average SUS score of 64.65 was identified for the novice group and 66 for the expert group.

For the second experiment an index-based OM2R editor was used. A mean SUS score of 64.2 out of 100 per participant was calculated. Please note for stage 1 only 18 out of 24 participants completed the usability questionnaire. This is in contrast to stage 2 where 48 users completed the SUS questionnaire. Please note only in stage 1 of the experiment the participants were separated into individual groups.

7.7 *Analysis*

This section presents an analysis of the experiment results with respect to the hypotheses. First the metrics are compared between the ontology-based and index-based representation with respect to H1, namely that "*The OM2R model can support participants in creation of more application and logically consistent documentation for ontology mapping if the model is structured as an ontology rather than an index-based structure*". Second, the participants are grouped based on their ontology mapping related experience in order to test H2, namely "*The OM2R model can offer similar support in the creation of application and logically consistent documentation for ontology mappings for participants with high ontology mapping experience, as for those with limited ontology mapping experience*".

7.7.1 H1: Application consistency – Impact of representation structure

This section focusses on analysing the impact of representation structure of the OM2R model upon the application consistency of the meta-data created by participants.

Table 26 in section 7.5.2 showed that the mean recall for ontology structure was 0.78 and for index structure was 0.35. The table also showed that the mean precision was 0.818 for ontology-based structure and 0.521 for the index structure.

The results show that the achieved mean recall for the ontology-based model was considerably higher than the index-based model. If mean precision is also considered, a similarly high performance of the ontology-based model can be observed. The results arising from the stage with the index-based structure indicate a high error rate. In contrast to the ontology-based model, a higher number of participants achieved higher recall. This suggests that ontology-based model representation can support the creation of application consistent meta-data in a much better way. Thus these results would suggest that an ontology-based representation for the OM2R model would be more suitable for mapping retrieval applications, as the meta-data documentation created have been found to be more accurate and complete.

In addition, it is helpful to consider the distribution of the recall values for application consistency between experiment stage one and two. Figures 40 and 41 in section 7.5.2 show that the results of the ontology-based model were centred more in the right area of the distribution and the index-based model more in the left region. Thus, the distribution shows that with the ontology-based editor more participants achieved a higher recall value for the created meta-data. The opposite is the case for the editor using the index-based representation of the model. For example the figure 42 shows that 45 % of the participants using the ontology-based model achieved a recall of 1.0, which means that they documented all facts correctly. In contrast the highest level of recall (0.7-0.8) with the index-based model was achieved by a mere 3 % of the participants. A similar distribution can be observed in the experiment results if the precision for application consistency (shown in figure 42 and figure 42) is considered.

The distributions indicate that a difference exists between the levels of application consistency achieved by the different representations used for the OM2R model. To verify the difference, it is necessary to check if the difference between both experiment stages is statistically significant. For this purpose an unpaired two-tailed t-test was calculated for each of the precision and recall results for each documentation task.[109] Table 30 presents the results.

[109] The t-test was selected as it is one of the most commonly methods for testing a hypothesis on the basis of a difference between sample means [We10a]. The t-test can be used in this setup as it allows a comparison of two small sets of quantitative data when samples are collected independently of one another. Between the ontology based and index based experiment no direct relationship can be identified as the experiment were conducted at different points in time. Therefore a test for independent samples (unpaired) is suitable.

Table 30 t-tests for recall and precision difference between ontology-based and index – based stage or application consistency

Task		**Mean for Ontology-based Model**	**Mean for index-based Model**	**P-value ($\alpha = 0.05$)**	**Statistical Significance**
Ontology language scenario	Recall	0.815969	0.374988	$T = 7.7969$, $p < 0.0001$[110], $p < \alpha$	Yes
	Precision	0.826833	0.544446	$T = 4.3918$, $p<0.0001$, $p < \alpha$	Yes
Matching process scenario	Recall	0.743063	0.333325	$T = 5.7756$, $p<0.0001$, $p < \alpha$	Yes
	Precision	0.809370	0.402783	$T = 6.4455$, $p<0.0001$, $p < \alpha$	Yes

The data in the table indicate that there is a strong statistical significance difference between the recall and the precision metrics between both models representations for each of the tasks. This provides further confidence that the ontology-based representation of the OM2R model may better support creation of application consistent meta-data and that the variations in the observed measurements is not pure chance (error factor 0.05 %).

In addition, some facts were intentionally vaguely formulated with the goal to drive the participant to use the provided help documentation. Table 31 presents a comparison of the mean recall metric of the total experiment with only data that included the meta-data fields with vague instructions. [111]

Table 31 Recall results for vaguely formulated instructions

Metric	**Experiment Stage 1 (ontology structure)**	**Experiment Stage 2 (index structure)**
Ontology language scenario– Mean recall for vague instructions	0.775	0.146
Matching process scenario- Mean recall for vague instructions	0.685	0.11

For both tasks the recall and precision metrics are considerably lower when an index-based model structure was used. A possible explanation is the limited functionality offered by the index-based OM2R editor, as simply not all information available in the OM2R model could be represented.

[110] To be more precise p = 1.59280339124E-12.

[111] The first documentation task contained two meta-data fields with vague instructions and the second task four meta-data fields.

The index-based OM2R model shows only the default name. All other information can only be found in a single external help documentation page. In contrast the ontology-based tool is able to provide different views of help data and as such makes accessing the documentation easier and faster.

However, it should be noted that it is difficult to draw a general conclusion from such comparisons, as the participant was given a set of plain text instructions in a lab environment in both experiments.

7.7.2 H1: Logical Consistency – Impact of representation structure

In this section the analysis of the achieved level of logical consistency is presented. Logical consistency addresses the ability to support the documentation of an ontology mapping whilst avoiding inconsistent statements. Table 32 provides a side by side comparison of the mean overall metrics for both experiment stages.

Table 32 Mean precision and recall for logical consistency for experiment stage one and two

Metric	Experiment Stage 1 (ontology based representation)	Experiment Stage 2 (index based representation)
Mean recall	0.86	0.18
Mean precision	0.85	0.45

The table shows that the recall values achieved are considerably lower when an index-based ontology structure is used in the experiment. A similar picture can be observed for precision but the low achievement is not as great as with the recall. These results show that participants in stage 2 with the index-based editor had considerable difficulties to document logically compatible meta-data combinations, in fact on average only 1 out of 5 of the expected compatible relations were documented. This represents a very high error rate, with the documentation of an ontology mapping containing incompatible or contradicting relations between meta-data field selections.

One of the reasons for the positive result for the ontology-based representation is that the corresponding OM2R editor is able to provide active support for participants to avoid inconsistent statements in the form of visually highlighting of compatible field data combination. This dynamic assistance is a help for participants as they can identify compatible relations easier. In contrast, with the index-based editor those relationships are only available in help text document as plain text. In addition, due to the limitations of the index structure this relation is expressed as a generic "see also" relation. Thus, it is not clear to the participant that the field content is compatible to another. It indicates some kind of relationship but of an unknown nature. However, even a flat index file can be used to highlight relevant information if a specific software tool or implementation is used. The advantage of the ontology-based model is that the information needed is semantically embedded in the model and not hidden in a specific software code. With such an explicit representation inside the model itself the help information can be maintained and shared more easily. Overall this shows that the OM2R can be used to build suitable tools to assist participants in the creation of consistent documentation with self-describing meta-data structures.

A next step is to identify whether there is a statistical significance difference between the mean precision and mean recall values gathered with respect to logical consistency. For this purpose an unpaired two-tailed t-test was calculated for the mean precision and recall results for each documentation task. Table 33 shows the results.

Table 33 Unpaired two-tailed t-tests for recall and precision difference between experiment stage one and two for logical consistency

Test Subject		**Mean for Ontology-based representation**	**Mean for index- based representation**	**P-value** ($\alpha = 0.05$)	**Statistical Significance**
Ontology language scenario	Recall	0.895838	0.3333	$T = 7.9620$, $p < 0.0001$[112], $p < \alpha$	Yes
	Precision	0.830560	0.402783	$T = 6.5760$, $p<0.0001$, $p < \alpha$	Yes
Matching process scenario	Recall	0.825000	0.031579	$T = 10.2991$, $p<0.0001$, $p < \alpha$	Yes
	Precision	0.870283	0.491226	$T = 4.7722$, $p<0.0001$, $p < \alpha$	Yes

The data in the table indicates that there is a strong statistical significant difference between the recall and the precision metrics between the two model representations used in task 1 and task 2. This data supports the claim that an ontology-based OM2R model is

[112] To be more precise p = 8.58411657276E-12

better able to support the creation of logical consistent documentation, than an index-based structure.

It is worth noting that mean recall for task 1 is 0.3333 but for task 2 it is considerably lower at 0.03. In task 1, the participant had to document only 3 relations with a focus on relatively known language related aspects (notation vs. formal language). Task 2 is far more complex and contained 6 relations. The focus was on matching which is also a less known topic. This is a possible explanation as to why a higher recall and more correct documented facts was achieved in task 1, compared to task 2 where therefore more errors are expected.

7.7.3 H2: Application Consistency - Support for Novice and Experts

In this section an analysis of achieved level of application consistency for the different participant groups is presented. To identify any statistical significance difference between the application consistency results of the expert and novice group, an unpaired two-tailed t-test was undertaken. The tests are based on the precision and recall results for each documentation task using the ontology-based representation. Table 34 displays the mean results for both groups.

Table 34 t-tests for recall and precision difference between novice and experts per task for application consistency for experiment stage one

Test Subject		Mean for Experts	Mean for Novices	P-value ($\alpha = 0.05$)	Statistical Significance
Ontology language scenario	Recall	0.82	0.82	$T = 0.0041, p = 0.9967 > \alpha$	NOT ONE
	Precision	0.76	0.87	$T = 1.7015, p = 0.0956 > \alpha$	NOT ONE
Matching process scenario	Recall	0.75	0.74	$T = 0.2074, p = 0.8366 > \alpha$	NOT ONE
	Precision	0.82	0.8	$T = 0.3379, p = 0.7371 > \alpha$	NOT ONE

These results indicate that there is no statistical significant difference between the mean recall and the mean precision values for experts and novices both in task 1 and in task 2. Please note the difference between the precision reached by experts and novices for task 1 is slightly higher but still not statistically significant enough for an alpha of 0.05.[113] The data provides evidence that the ontology-based representation of the OM2R model may

[113] Since the p-value is 0.0956, i.e. greater than 0.05, it can be concluded that there is no difference between the means. To say that there is a difference is taking a 9.56 % risk of being wrong.

equally support experts and novices to create application consistent meta-data. However, it is difficult to draw a general conclusion from these results as the participants were given clear instructions in a lab environment. As discussed earlier in order to mitigate this effect, the provided instructions for some fields were intentionally vague, and so it is worth focusing on the results related to those fields in particular. [114]

Table 35 provides a comparison of the mean recall values of the total experiment focussed only on the meta-data fields with vague instructions.

Table 35 Recall results for vague formulated instructions

Metric	Total Participants	Expert participants	Novice
Average recall	0.78	0.785	0.776
Ontology language scenario - Mean recall for vague instructions	0.775	0.79	0.76
Matching process scenario - Mean recall for vague instructions	0.685	0.65	0.72

For task 2 the recall for both groups is lower (experts 0.65 and novices 0.72) than the overall average for experiment stage one. These lower results are likely as a result of the increased level of difficulty involved in completing the meta-data fields. Why the experts have such low recall for task 2 in comparison to the novice at this stage is not clear.

This aspect of the experiment perhaps reflects a more realistic scenario, as in real life instruction or available information will most likely be vague. Thus this data can also be considered to provide additional evidence in support of H1 hypothesis that the ontology-based OM2R model can better support participants in creating consistent meta-data documentation of mappings.

7.7.4 H2: Logical Consistency - Support for Novice and Experts

In this section the support for logical consistency is considered. To identify any statistical significant difference between the logical consistencies result of the expert and novice group unpaired two-tailed t-tests were calculated. The tests are based on the precision and

[114] The first documentation task contained two meta-data fields with vague instructions and the second task four meta-data field.

recall results for each documentation task (language and mapping) for the ontology-based representation stage of the experiment. Table 36 presents the results.

Table 36 Unpaired two-tailed t-tests for recall and precision difference between novice and experts per task for logic consistency for experiment stage one

Test Subject		Mean for Experts	Mean for Novices	P-value (α = 0.05)	Statistical Significance
Ontology language scenario	Recall	0.98	0.84	T = 1.6428, p = 0.1072 > α	NOT ONE
	Precision	0.84	0.83	T = 0.0951, p = 0.9247 > α	NOT ONE
Matching process scenario	Recall	0.84	0.8	T = 0.1978, p = 0.8441 > α	NOT ONE
	Precision	0.88	0.86	T = 0.2032, p = 0.8399 > α	NOT ONE

The data indicates that there is no statistical significant difference between the recall and the precision metrics for experts and novices in task 1 and task 2. Please note the difference between the precision reached by experts and novices for task 1 is slightly higher but still not statistically significant enough for alpha = 0.05. As a result the data indicates that the ontology-based OM2R model is able to support experts and novices in creating logical consistent documentation on a similar level.

7.7.5 Application and Logical Consistency – subjective participant feedback

Another aspect of the experiment was the subjective questionnaire where participants had to share their opinion on the support for application and logical consistency the model can offer.

In a first step the participants had to agree (5) or disagree 0) with the statement that the model can support the creation of application consistent documentation. The ontology-based model received a mean of 3.6 and the index-based model received a rating of 3.1. Both ratings suggest that the majority of participants tended to agree with the statement. The similar high rating for both model representations is not unexpected as an index-based presentation is a very common way to access meta-data models and most people would be familiar with such structures and know it strengths. The lower level of approval of 3.1, compared to 3.6 in the experiment stage one, provides some evidence to support the claim that an ontology-based approach can provide a better support for application consistency.

Naturally these subjective ratings need to be considered with caution as the participants have no prior knowledge of the OM2R model itself, and were introduced to the model for the first time in the experiment. In addition, the experiment is strictly controlled and the participants are provided with a model editor which allows very little freedom in its use. In short the experiment is limited to meta-data documentation following a fixed, tightly controlled set of instructions. An application of the model in a real life scenario may produce a different result as naturally not all information needed during such a task is always available and typically the documentation scenario would not be as clear as in this experiment.

In the second step participants have been asked to agree (5) or disagree (0) to the statement that the model can support the creation of logically consistent documentation. Across all participants the mean of the answers was 3.25 out of a maximum of 4 for the index-based representation in experiment stage two. In contrast the mean of the answers was 3.8 for the ontology-based representation in experiment stage one. The visual highlighting is a simple to use feature and makes creation of logical consistent documentation very easy. This might explain the slightly higher confidence rating in the experiment stage one where this feature was available. The lower level of approval in the experiment stage two provides some evidence to support the claim that an ontology-based approach can provide a better support for logical consistency. However, again the weight of this evidence is low, given that as discussed earlier the number of participants is low in the second stage of the experiment (24 compared to 48).

Overall the questionnaire in this experiment showed that the participants agree that an ontology-based model can better support the creation of meta-data that is logical consistent, than an index-based model representation.

7.7.6 Usability

The usability results cannot be interpreted in isolation as the tools used are custom0built for this experiment. The results show that the users perceived the usability aspects of the ontology-based tool slightly higher than the index-based tool. Both experiment tools use the same layout and design. However, the ontology-based tool used the information stored to provide more sophisticated and useful editing features (65.1 vs 64.65 out of

100). This indicates that in average the users considered the ontology-based editor slightly more user friendly than the index-based tool. As the difference is very small (0.9 points out of 100) a general conclusion cannot be drawn. This leads to the question as to how this SUS result can be compared to the meta-data system.

According to a study in 2011[115] a SUS of 68 is considered average and as such both tools received a SUS below average. If those SUS scores are compared to the result of two other semantic mapping tools (73 for CoGZ and 46 for PROMPT [Fa09]) it can be observed that they are placed in the middle of the field. Compared to the SUS score for CoGZ the results are lower.

This insight is not a surprise as the tool was from a usability aspect, very basic and provided only few functions. The design, colour schema and implementation were focussed on the functions rather than on everyday usage. Despite the low sample size the SUS result collected in both experiments can be interpreted as a positive signal. It provides some evidences that even a simple and basic tool can enable an intuitive access to the model. However, overall the SUS does not provide sufficient evidence to judge if an ontology-based or index-based model representation is more user friendly. This seems to depend on the actual implementation more than the model used.

7.8 Conclusion

This experiment provides evidence towards the stated hypothesis 1 that support for the creation of consistent meta-data documentation is better if the OM2R model is expressed using an ontology rather than an index-based structure. It is worth highlighting that in this experiment the ontology-based OM2R provided a considerably superior support for logical consistency, with a mean recall being recorded of 0.18 for an index-based representation, compared to a 0.86 for an ontology-based representation.

The subjective questionnaire supports this claim as participants largely agree with the statement that using OM2R will support creation of application and logically consistent

---

[115] See http://www.measuringusability.com/sus.php for full details

meta-data of mappings.[116] However, the actual outcome of the documentation process is barely logical consistent according to the gold standard. This represents a discrepancy which reinforces the general challenge of using complex and interlinked documentation models [Ed07]. Generally documentation can offer users information about the meta-data fields and their application, but as the state of the art study in section 3.2 has shown, users are usually left without support when it comes to logical consistency. This can lead to the observed subjective overrated confidence as a user has typically no way to validate his results with regards to logical consistency. This experiment shows however that the benefits of an ontology-based OM2R model that supports a manual and automated logical validation of meta-data, can improve the overall consistency level of the mapping documentation.

The experiment could not identify any significant differences in the support for the creation of consistent documentation between experienced and novice users. This provides some evidence to confirm H2 in terms of that OM2R model may support real life applications, where users with diverse experience levels are expected.

A limitation of this experiment is that it is based on two custom-built editing tools which could bias the results. For example, the users had no freedom in terms of editing workflow, access to the full OM2R model or a choice of a more appropriate editing tool. For example, a user might have navigated the instance of the OM2R model better in a dedicated ontology editor such as Protégé. However, the general lack of specialised mapping tools with a focus on documentation made these developments unavoidable in order to offer the user at least a basic interface to use the model. The custom-built tools allow experts but also novice users to use the OM2R by focusing on the model content and remove the challenge to create a valid notation manually. To further mitigate this impact, a simple and straight forward design was chosen for both tools to put again the emphasis on the model fields and to allow a comparison of the results.

In the experiment the ontology-based editor could offer a 2.2 time better recall for application consistency and 4.6 time better recall for logical consistency compared to an

[116] For example experiment stage 1 (section 5.5) users show a high level of confidence (3.25 out of a max of 4) that the model can support logical consistency.

index-based representation of the model. This was achieved utilizing the available semantic information but the offered functionalities (e.g. different view types and active highlighting of options) are only an example of how such information can be presented. Other more sophisticated ways are feasible as to how the expressive information can be made available to users to assist them in the documentation process. For example, a graph based visualisation to highlight the relationships between all elements, automated completion of search terms and so on.

The second limitation of this experiment is the controlled setup and scenario. These were intentionally documented in a complete manner in order to allow the application of a gold standard. However, an application of the model in a real life scenario might produce a different result as naturally not all details about the ontology mapping lifecycle and in particular the involved matching algorithms are available. Furthermore, the documentation scenario would most likely not be as clear as in the experiment. The vaguely formulated facts in the task descriptions were used to mitigate this limitation of a controlled experiment and to make it more challenging for the user. 20 % of each task were deliberately formulated vaguely. The recall results in both tasks show that an ontology-based representation provided better results. For example with the ontology-based model a mean of 1.54 correct formulated facts were achieved out of 2 compared to 0.29 for the index-based model. However, more detailed studies are necessary to identify if the user actually used the offered meta-data or rather used, for example, external resources.

In order to investigate the application of the OM2R model in more realistic scenarios, two practical use-case studies were undertaken and are presented in the following chapter 8 and 9.

8 Case Study – OM2R for OAEI

The Ontology Alignment Initiative (OAEI) conducts annual evaluation campaigns of ontology matching methods [Eu11b]. This section presents a case study which investigates how the OM2R can be applied beneficially for the publication of meta-data of matching challenges in the OAEI. The chapter contains an introduction, the methodology used and an analysis of the meta-data used by the OAEI and how the OM2R model could be applied.

8.1 Introduction

In recent years the number of methods that are being proposed for matching (in this domain often referred to as alignment) [Eu07] of ontologies have increased considerably [Ma12, Sh12]. This creates the need to establish a consensus on the evaluation of these methods [Sh12]. The Ontology Alignment Evaluation Initiative (OAEI) [Eu11a] organizes campaigns to evaluate ontology matching methods every year. The OAEI offers different matching challenges which provide a collection of ontologies and reference ontology matchings.[117] These enable a comprehensive evaluation of the tools and algorithms in a controlled environment. These challenges and the reference ontology matchings evolve from year to year to provide a more effective test environment [Eu11a, Eu11b]. The OAEI represents one of the few organisations which have an immediate need to retrieve, analyse and manage ontology matchings in high detail and on a large scale over time.[118] It can be argued that the OAEI applies reuse of ontology matchings as they are used by different participants for different challenges in multiple test rounds over the years.

This motivates the need for suitable meta-data to document the offered reference ontology matchings, so as to assist OAEI organisers in managing changes, the participants in understanding the challenge details and third party researchers in analysing the results. Currently the OAEI does not use any structured meta-data model to

[117] For example in 2012 the OAEI provided seven distinct challenges. Each challenge contains up to 58 individual alignment tasks [Eu11b].
[118] The first OAEI challenge was conducted in 2004.

document the lifecycle details of ontology matchings. Section 7.8 has shown evidence that the OM2R can be used to document the ontology matchings consistently and as such assist retrieval and mapping reuse.[119] This leads to the main question for this case study:

> *Can the OM2R model be applied for the OAEI beneficially to document alignment challenges consistently to support online consumption, retrieval and further analysis?*

This case study will provide some quantification on the likely benefits in terms of helping challenge administrators and participants consistently create annotated challenge documentation and ontology matching results that are explicit and easy to interpret. The case study described here was first published in the 7th International Workshop on Ontology Matching (OM-2012) [Th12]. The following section outlines the methodology used.

8.2 Methodology

[Eu11b] states the goal of the OAEI as "assessing strengths and weaknesses of alignment/matching systems; comparing performance of techniques; increase communication among algorithm developers". To achieve this goal the various stakeholders in this domain have different requirements regarding the documentation of alignment challenges. Table 37 provides an overview of these which can provide insight into the requirements for meta-data [Eu11a].

---

[119] OM2R has a wider scope as it focuses on mappings but the OAEI is limited to alignments [Fe10]. In our terminology alignments are machine-generated correspondence candidates. These represent an essential step in the creation of mappings that are confirmed correspondences created in the mapping phase. Thus the OM2R covers the documentation of alignments and therefore can be applied to the OAEI domain.

Table 37 Overview of the OAEI stakeholder activity

OAEI Stakeholders	Activity
OAEI Event Organizer	This group is responsible for the overall management of the events, the submissions and the result publication every year.
OAEI Challenge Organizer	They maintain the individual challenges by creating and documenting the specific alignment tasks and reference alignment files.
Challenge Participants	This group uses their matching method to complete the individual alignment task. They submit their results as alignment files or since 2011 as application bundles.
third Party Research	They utilise the published alignment results and metrics to learn more about the matching method, their effectiveness and evolving trends.

The overview shows that the focus of the OAEI organizers and participants is on the effective online publication of the relevant challenge details and the change management of the ontology matchings [Eu08]. The interest of third party researchers is on support for retrieval and analysis of the reference and submitted ontology matchings. To allow for adequate measurements and identification of trends [Eu11a] it is vital that they know precisely how each challenge was conducted and what matching algorithms were applied. This results in a need for a detailed and consistent documentation of the ontology matchings across all stages of the matching lifecycle. Based on these stakeholder requirements, two evaluation dimensions for this case study are defined by the author of this thesis"

The first dimension the **characteristic of the meta-data model (D1)** is evaluated. This includes the *number of supported meta-data fields* which provides an indication of the complexity of the model. Furthermore, the *representation concept for the meta-data* is discussed which can range from an unstructured text based approach to an explicit structured meta-data model.

The second dimension focusses on the meta-data **support for the creation of consistent documentation (D2).** This is difficult to measure and compare as it can be conceptual, tool or process based. In order to guide the comparison this case study will be limited to *aspects relevant for application consistency* which relates to the correctness and completeness of the documentation. In addition, *aspects for logical consistency* are discussed which relate to the avoidance of inconsistent statements in documentation.

To allow such an evaluation, the analysis will focus upon specific individual available meta-data fields in detail. The core interest of the OM2R model is the discovery of matchings and the use of matchings over time. This is reflected in the scope for this experiment. More specifically, the following meta-data fields will be evaluated as they have been identified in previous experiments (see chapter 7) as relevant for mapping retrieval and re-use: *ontology names*, *descriptions*, *unique identifier*, *ontology language* and *change management details*.

This study will be based on the analysis of one representative alignment challenge in order to allow a detailed evaluation of the meta-data supported by the OAEI. The selected challenge needs to be documented extensively to allow the identification of offered meta-data fields and the documentation. It should have also been used in previous OAEI initiative to be able to consider changes over time.

In the recent OAEI challenge of 2012 the following alignment challenges (so called data sets) were offered: Benchmark, Anatomy, Conference, Multifarm, Library, Large Biomedical Ontologies, Instance matching [Eu11b]. If the previous four OAEI evaluation rounds (covering the years 2012, 2011.5, 2011, 2010) are considered, only the following challenges have been used in all four rounds: Benchmark, Anatomy and Conference. If the available documentation for these three challenges for the 2012 round are compared, it can be noted that the webpage for the benchmark challenge contains the most detailed documentation.[120] As a result the *Benchmark* alignment challenge offers the highest amount of information with historic application.[121] The author of this thesis selected this challenge as it provided the most insight. A copy of the web page which documents the Benchmark challenge can be found in this thesis DVD in the folder *F OAEI Case Study*.

[120] The amount of information offered for each challenge is determined by the word count of the documenting web page. The benchmark web page contains 3505 words compared to 702 for the anatomy 702 and 544 for the conference challenge. Please note the benchmark challenge of 2013 will not be considered as it was considerable reduce in complexity to only 2 challenges. Thus it provides considerably reduced insights into meta-data use.

[121] Please see for details: http://oaei.ontologymatching.org/2012/benchmarks/index.html.

8.3 Data - Current OAEI Approach

This section discusses the evaluation dimensions for the current meta-data approach used by the OAEI.

8.3.1 D1 - Characteristic of the meta-data model used by the OAEI

Since 2004 the OAEI has defined that each alignment challenge is to document a single web page to provide the scaffolding needed for participants [Eu11b]. The page must include a short textual description of the dataset and evaluation modalities. An evaluation of the select challenge of OAEI 2012 [Eu12b] showed that the majority of information is provided as unstructured text segments and lists on the HTML page. Only few reference ontologies provide embedded meta-data information themselves. The web page per challenge task represents the main source for meta-data for participants and analysts. It can be observed that the OAEI does not utilise any structured meta-data model but rather a text based ad hoc documentation which varies in structure and content between the different challenges (see 8.2 section for details).

In the example of the OAEI "Benchmark" challenge, 12 specific meta-data elements can be identified. Table 38 provides an overview of these fields. The column "Representation" indicates if the field is presented in an explicit *field* (e.g. embedded in the addressed ontology) or mentioned in the *text* segments only. The column also shows if the information is available for all (A) all ontologies, or only for some (S).

Table 38 Overview of the OAEI stakeholder activity

Meta-Data field	Representation
Name of ontologies	Text (A) Field (S)
Description of ontologies	Text (S) Field (S)
Location of ontology	Text (A)
Creation date	Field (S)
Unique identifier for ontologies	Field (A)
Complexity of the ontology	Text (S)
Design of the ontologies	Text (S)
Notation of Ontologies	Text (S)
Formal Language of Ontologies	Text (S)
Creator	Text (S) Field (S)
Matching Location	Text (A)
Formal Language of the Matching	Test (S)

This table shows that the coverage of ontology mapping lifecycle details is low as the current meta-data focusses primarily on ontology identification, with only one field provided for the matching phase.

8.3.2 D2 - Support for creation of consistency documentation in the OAEI

To gain an insight into the current support for creation of consistent meta-data documentation, it is helpful to focus on individual meta-data fields deemed relevant for retrieval and mapping reuse. For example, to participate in a challenge a user needs to identify the addressed ontologies. The Benchmark web page offers a brief textual description for this purpose. In it the source ontology is referred to with different names such as "reference ontology", "bibliographic ontology" but also in the task section as "test". The web page lists 58 specific alignment tasks where the target ontology is specified. Please see in table 39 two examples displayed for such individual task descriptions [Eu11b]:

104) Concept test: Language restriction – This test compares the ontology with its restriction in OWL Lite (where unavailable constraints ... Ontology : [RDF/XML] [HTML] Alignment : [RDF/XML] **201[-2-4-6-8]) Systematic**: No names - Each label or identifier is replaced by a random one. Ontology : [RDF/XML] [HTML] Alignment : [RDF/XML] [HTML]

Table 39 Alignment Task Description of the Benchmark Challenge

It can be noted that the amount of descriptive information for the target ontologies is not consistent for each task, e.g. see example for test 104 vs. 201. It can be observed the tasks listed on the lower sections of the web page contain less information than on the top. Only 25 % of the target ontologies provided in this challenge have meta-data embedded in their source code, e.g. <dc:description>, <rdfs:label> for task 225. It can also be observed that such embedded information cannot be found in all target ontologies, e.g. such information is missing for task 250 and 303.

Human names are a dominant key for identification, but in the Semantic World an unambiguous identifier for ontologies is essential to allow automated processing. In the challenge the *base url* of each ontology is used for this purpose which is unique for each challenge and each data set, e.g. <rdf:RDF xml:base="http://oaei.ontologymatching.org/ 2012/ benchmarks/ 250/onto.rdf#"> for task 250. Until 2010 the web page claims the same ontology was used for this dataset. However, contradicting this assertion is that each of the ontologies has a unique identifier and is therefore from the perspective of automated process potentially different.

Language related information of the addressed ontologies is also crucial for processing and compatibility issues of matching tools [Sh12]. The OAEI provides information about notation and the formal language in text form on the challenge web page. For example the web page states that the reference ontology is available in RDF/XML. The formal language is mentioned in the text but not consistently and in some cases missing, e.g. see description for alignment task 236.

Information about the changes to reference ontologies is particularly relevant for performance evaluation over time as they can bias the results. The OAEI provides only brief and unspecific textual references for such changes. For example, in 2011 the challenge web page states that the reference ontology for the benchmark data set has been altered and “it no longer the very same dataset that has been used from 2004 to 2010” [Eu11b]. No further details are provided.

Not one of the 58 ontology matchings in the benchmark challenge provides any information about the method or tool used to generate them. Even the standard field defined in the applied alignment format which could refer to a class is not used in OAEI 2012 benchmark.

With regards to application consistency, it can be observed that no formal or structured meta-data model is used by the OAEI. As the above examples demonstrate, the resulting documentation is inconsistent (e.g. different names are used) and not detailed (e.g. missing change details) in terms of lifecycle coverage. The missing formal meta-data model makes it very hard to apply tools for automated checks for logical consistency. In addition, the existing contradicting statements in the documentation are a further obstacle to logical consistency.

8.4 Analysis – Application of the OM2R for the OAEI

This section discusses how the OM2R meta-data model can be applied for the OAEI.

8.4.1 D1: Characteristic of the meta-data model of the OM2R applied for the OAEI

The primary focus of the OAEI is on online consumption by humans as the majority of meta-data are presented in text form and tables. In contrast to the OAEI, the OM2R model is designed with retrieval and automated processing in mind. The OM2R model is expressed as an ontology and therefore all meta-data information is stored as explicit and meaningful triples, e.g. om2:source_ontology hasNotaton rdf/xml = object of interest - typed relation - meta-data field option. In addition, the current documentation used by the OAEI is limited to document the details on individual alignment challenges. This is well suited for challenge participants but limits the view for third party researchers and organizers. The benefit of the OM2R model is that multiple alignments can be documented in one OM2R model instance. This is particularly relevant for challenges such as Benchmark which are designed to be stable over time but underwent some change over time which can be clearly documented in the OM2R.

Table 40 provides a comparison of all meta-data supported in the OAEI and the OM2R model. The fields used are limited to the identification, characterisation and matching phases as these lead up to the creation of alignments which are the focus of the OAEI.[122]

[122] The table shows which fields are provided by the OAEI and the corresponding fields in the OM2R. The column "OAEI Fields" indicates if the meta-data related information is presented by the OAEI in an explicit field (e.g. embedded in the ontology) or was mentioned in an unstructured text segment. The column also tells if the information is available for all (A) addressed target and source ontologies or only for some (S).

Table 40 Comparison of OM2R meta-data fields with those provided by the OAEI

Meta-Data field	OAEI Fields	OM2R - Meta-Data Fields
Name of ontologies	Text (A) Field (S)	SourceOntology :Om2r:human_readable_name: "Biology Top Level Ontology"
Description of ontologies	Text (S) Field (S)	Om2r:description
Location of ontology	Text (A)	Om2r:hasLocation (type url)
Creation date	Field (S)	Om2r:hasCreationDate (type date)
Unique identifier for ontologies	Field (A)	Om2r:hasIdentifier
Ontology version	Missing	Om2r:hasVersion (URI)
Complexity of the ontology	Text (S)	Om2r:hasClassCount 73, hasInstanceCount 3, hasPropertyClass 3
Design of the ontologies	Text (S)	Om2r:hasDesign om2r:deep_hierarchy.
Notation of ontologies	Text (S)	Om2r:hasNotation RDF/XML
Formal language of ontologies	Text (S)	Om2r:hasFormalLanauge OWL
Creator	Text (S) Field (S)	OM2r:Creator
Matching location	Text (A)	Matching Om2r:hasLocation: www (URL)
Formal language of the matching	Test (S)	Om2r:hasformalMatchingLanguage: EDOAL
Notation of the matching	Missing	Om2: hasNotation: RDF/XML.
Matching method	Missing	Om2r:hasMethod (manual, automatic, mixed)
Matching tool	Missing	Om2r:isTool AlignmentServer
Matching algorithm	Missing	Algorithm :encodedIn: Java, Algorithm :hasJavaClass: org.stringComp, Algorithm :hasSource: freecode.org/a.zip
Algorithm is based on	Missing	Om2r:isBasedOn rdfs:label, rdfs:class
Applied threshold	Missing	Om2r:has_Applied_Threshold
Matching scope	Missing	Om2r:hasScope (complete or partial=
Matching requirements	Missing	Om2r:hasMatchRequirements (text)

This table shows that the OM2R model provides a more detailed meta-data model with a wide range of explicit meta-data fields to document the matching lifecycle. More specifically, the OM2R supports 25 explicitly expressed meta-data fields in contrast to only 12 mostly unstructured meta-data elements in the OAEI.

8.4.2 D2: Support for creation of consistency documentation with the OM2R in the OAEI

With regards to application consistency the advantage of the OM2R model is the ontology template which provides the explicit meta-data fields but also specific field content options which a user can select during the editing process. This can make the creation of documentation of mappings easier and potentially more correct as users can simply select from the given options. Furthermore, in contrast to the OAEI approach each OM2R meta-data field is linked to a definition which helps users easily understand the intended meaning of the individual fields [Fu93, Ba99]. This helps OAEI participants but also third party researchers to understand the given documentation over time. The OM2R is expressed as an ontology and can therefore be extended in a flexible manner to accommodate new matching methods. This is particularly vital for the OAEI whose very purpose is to showcase new matching methods every year.

The OAEI offers no support for logical consistency in their documentation. The OM2R can help as it contains explicit relations between the given field options in the model, e.g. compatible relation between *language* and *notation*. Such a rich structure can be used by tools to recommend compatible combinations of meta-data fields. The OM2R model is formally detailed in a machine-interpretable notation such as OWL. This allows the potential for reasoning and as such for automated logical consistency check and re-use of alignments for other purposes in the future [Fe10, Br12].

The use of human generated names and inconsistent descriptions for ontologies in the OAEI are suitable for online consumption by users. However, a more consistent approach is needed to support automated mapping retrieval, analysis and processing. The OM2R model addresses this by providing the explicit fields *Ontology Name* and *Ontology Description* field. Thanks to the ontology based representation additional meta-data can be expressed easily and meaningful if needed. For example, *hasAlternativeName* and *hasNaturalLanguage "German"*.

To help users in interpreting and reusing the provided resources, more explicit information can be helpful. For example reasoning can only be applied to OWL DL and not OWL Lite, thus stating the language as OWL would be too broad. The OM2R model addresses this issue with the fields *Ontology Formal Language*. As there are a number of such languages this field specifies the language, e.g. :hasFormalLanguage: http://www.w3.org/2002/07/owl. In case of OWL it is important to specify the sublanguage, too e.g. :subLanguage: OWL-DL using *Ontology Notation*. Beside the ontology language, the specific exchange notation used to represent the addressed ontology can be specified which is essential for tool support and exchange, e.g. *TargetOntology :hasNotation:* RDF/XML.

In addition, it is worth highlighting that the OM2R model provides a set of 8 explicit fields to document the actual matching process which is in the end the prime interest of the OAEI.

Documentation of change to ontology matchings can be effective if the approach is flexible and not too difficult. The core requirement in this context is the ability to identify a specific alignment version. Similar to the OAEI a URI is used but to avoid any misinterpretations, instead of the base URI an explicit field call *Ontology Identifier* is used to store a unique identifier. This targets the needs for automated application and the OM2R model offers the field *Ontology Version* which provides details about the specific version for human editors.

8.5 Conclusion

This case study has shown that the OAEI meta-data approach is currently unstructured, inconsistent and fragmented. It has been shown that the OM2R model provides a wider range of explicit meta-data fields to enable the documentation of the matching lifecycle in more detail. More specifically, the OM2R supports 25 explicit fields, in contrast to the 12 mostly unstructured meta-data elements in the OAEI.

To evaluate the support of application consistency the representation of 5 specific meta-data fields were analysed. Specific examples demonstrated that the OM2R model can make the editing and the interpretation of the intended meaning easier, less ambiguous and provide a better structure for human and automated consumption. It could also be argued that logical consistency can be improved with the OM2R based on its rich relation and reasoning support.

The analysis undertaken in the case study supports the claim that the OM2R model can be applied for the OAEI initiative and help organizers create more consistent and detailed documentation of challenge mappings for the benefits of participants and researchers. It is also clear that the ontology based approach can help avoid misinterpretation and bias which are critical for the scientific focus of the OAEI performance evaluation. The ability of the model to be extended flexibly is of benefit to OAEI as matching methods evolve constantly from year to year.

As the OM2R model is highly focussed on the specific documentation needs of alignments and the current approach has disadvantages, it is likely that OAEI will find the OM2R model attractive. It is helpful that that the OM2R model can be used In addition, to the recommended Ontology Alignment Format. Thus, no tools or processes need to be changed and together with the embedded documentation a practical roll out of the model in the OAEI is expected to be relatively easy. First interest in this usage of the OM2R was expressed by the OAEI organisers during the OM-2012 workshop presentation [Th12].

A limitation of this case study is the fact that only the benchmark challenge of the OAEI initiative of 2012 was evaluated in detail. Other challenges in different years might provide different meta-data. However, spot checks by the author of this thesis of other challenges in 2012 and 2013 have indicated that this is not the case. A challenge for the practical application of the OM2R model to the OAEI dataset would be that considerable effort that would be needed to document the high number of existing ontology matchings for the benchmark challenge (58 ontology matchings) together with the lack of existing meta-data in any form. It may be more practical for OAEI community to apply the OM2R approach to more recent challenges or new challenges.

To provide further evidence that the OM2R model can be applied in real life scenarios a second case study focusing on federation is presented in the next chapter.

10 Case Study – OM2R for Managed Semantic Interoperability for Federations

This chapter assesses whether the OM2R model could be applied beneficially to the management of ontology mappings within federations. The research described here was first published in the Computer Networks Journal [Fe10].

10.1 Motivation to Apply OM2R for Federations

Research to date in semantic interoperability has a strong focus on the generation of ontology mappings. This side-lines the problems of dealing with the dynamism of both the data and the corresponding mappings which are characteristics of real-world integration problems [Ma12]. Once multiple autonomous sources, such as those in a federation, are considered this variability increases even further over time [Wa12].

An aim of federation research [Be11c, Fa12] is to help users consume data from multiple heterogeneous resources with a minimum of re-integration effort over time [Wa12]. As a consequence, management of relationships needs to be made explicit as opposed to being deeply integrated in the application or service management component. The relationships need to be separated from any information describing the mappings required to use shared capabilities and managed on their own [Fe10]. In the light of on-going changes this separation can bring benefits to the process of maintaining local data integration with remote information capabilities. As most change takes place within known relationships and existing federal contexts, there is a large potential for managed reuse of mappings.

The Federal Relationship Manager (FRM) is an implementation of this relationship focussed federation approach [Wa12]. The FRM offers mapping retrieval and management based on federation context [Be13]. More specifically, the FRM can suggest relevant ontology mappings automatically when a new capability is imported, by searching the set of available mappings and knowing the federal relationship context of the request. Thus the FRM has a strong focus on reuse of these mappings in an ever changing environment of online data consumption.

The FRM is based on meta-data describing the federation context and in particular the available ontology mappings. Figure 48 provides a UML diagram of the FRM activities which are defined in this mapping management process [Br13]. Each activity (highlighted in italic) is discussed in more detail below the graphic

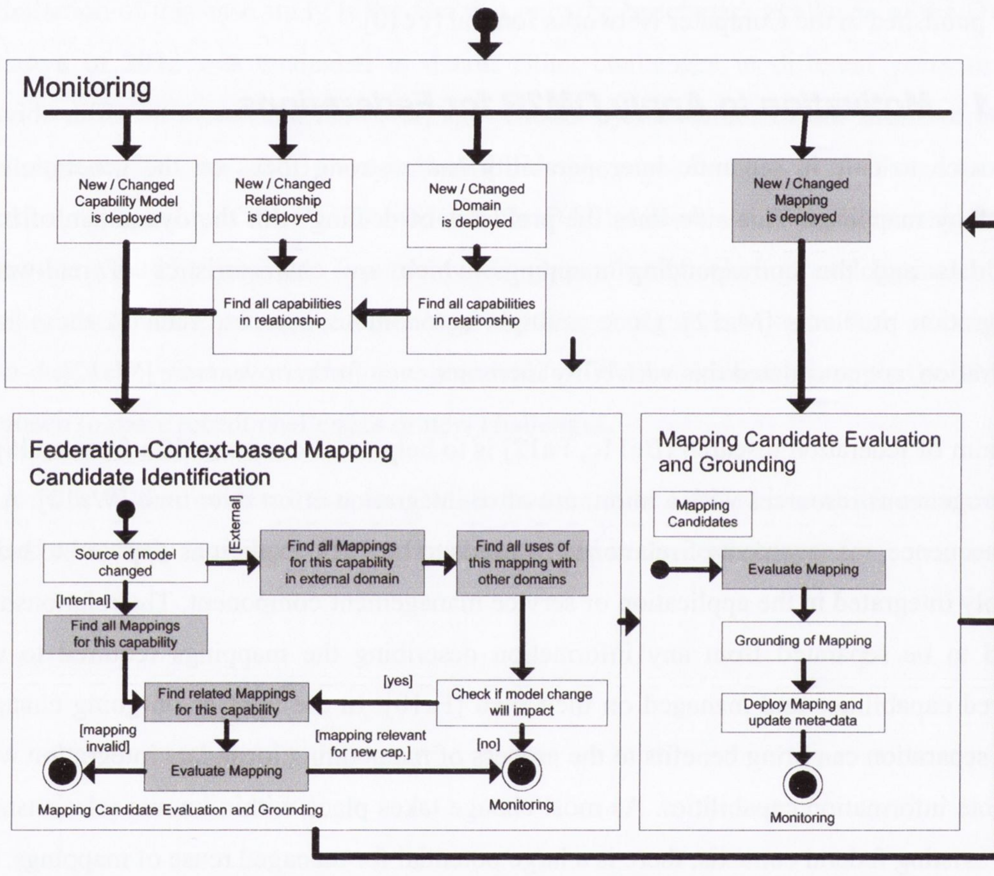

Figure 48 Managed Semantic Interoperability Process (based on Figure from [Be13])

The first FRM activity is *monitoring* where the FRM checks regularly for changes in its managed federation models. For example new capabilities being shared, shared capabilities being modified or relationships modified. The result is a capability set that requires an associated set of mappings to be identified (via federation context-based mapping candidate identification) or a mapping candidate set (that must be evaluated and grounded).

The second activity is the FRM *federation context-based mapping candidate identification* which aims to determine the set of potentially relevant mappings for a given federation context due to a federation-level change. A federation context model must be queried for each of the capabilities to determine the set of relevant mappings.[123] The core of this activity is the discovery of existing relevant ontology mappings. This is the first use-case where the OM2R can be applied beneficially.

The third FRM activity described here is the *mapping candidate evaluation and grounding* which is used to evaluate and refine mapping candidates into concrete mappings. These may be simple mappings (equivalence, broader, narrower) expressed in OWL or complex mappings that require some form of data transformation to be performed. The work published in [Br13] specifies a process for the automated refinement of appropriate simple mappings into complex mappings and generating groundings for those mappings. This activity of evaluation of mappings is the second use-case for the OM2R model.

Currently in the FRM, ontology mappings and their descriptive meta-data are represented in the Ontology Alignment Format or as extended EDOAL ontologies. The main focus is the representation of specific correspondences between the elements in the linked ontologies. Meta-data documenting the mappings are limited (see section 2.7 state of the art) and very limited support for the creation of consistent documentation is offered.

10.2 Hypothesis

This leads to the hypothesis for this case study:

> *The OM2R model can be used to improve the documentation of ontology mappings for Federal Relationship Manager (FRM) activities such as mapping discovery and evaluation.*

[123] If changed or new mapping candidates are detected by this activity then they are handed over to the mapping candidate evaluation and grounding activity, otherwise the system resumes the monitoring state.

Please note the main focus of this use-case is placed on the meta-data model needed to document ontology mappings meta-data for activities in the federation. The actual implementation and meta-data specific editing tools were out of scope.

10.3 Methodology

To investigate the stated hypothesis it was necessary to identify the requirements for ontology mapping meta-data in federation based on a literature research. Based on these requirements the activities in the FRM management process are discussed to define use-cases where the OM2R can be applied beneficially. Based on these use-cases the OM2R model is analysed along two dimensions.

In a first dimension the focus is placed on the **representation of ontology mapping meta-data in federations (D1)**. In other words how is the meta-data represented in the federation regarding ontology mappings to support mapping retrieval and reuse? This provides the high level view where the OM2R model can be placed in the overall federation process.

In a second dimension the **characteristic of the meta-data model for ontology mappings (D2)** is evaluated. This includes a discussion of the individual meta-data fields of the OM2R model that can specifically support the use-cases defined for the FRM. A main feature of the FRM is the automated process to enable mapping identification for reuse between capabilities in the federation. Thus as part of this analysis dimension, it is considered if the meta-data in the OM2R can be generated automatically to support this FRM feature.

10.4 Meta-data Requirements of Federal Relationship Manager

The FRM is designed to manage dynamic relationships between autonomous domains within and across organisations [Be13,Wa12]. FRM is based on capabilities and domains. Capabilities are abstractions of services or resources to be shared between domains[124]. Domains are considered independent entities which share capabilities among each other

[124] Capabilities are shared in capability authorities and arranged in a graph such that some capabilities are identified as narrower or broader reflections of some underlying resources such as a file system.

whereby federations are a sub-type of a domain [Be13b].[125] The FRM keeps track of both the capability model and the domain model as distributed graphs. Together these graph-based models create a so called *federation context* and a model for this purpose was developed based in RDF(S) [Br13].

This federation context model is independent from the actual mappings. This allows the FRM to relate mappings to FRM federations, domains, relationships and capabilities. The relationship meta-data can be generated automatically and maintained for all mappings. Semi-automation is available for the process of maintaining data integration in the presence of change.

Figure 49 provides a summary of an FRM application scenario as defined in [Br13] where the FRM is used as a service quality management application.

- FRM regularly gathers structured service monitoring information from multiple business partners such as resellers, suppliers, and customers.
- Each information domains exist within a specific federation relationship with actors, terms of engagement and data access.
- Over time the number and nature of these relationships change as the business environment evolved, e.g. new resellers are found and supplies are changed.
- Organisational changes are reflected in the data consumption behaviour
- Evolving internal models are selectively combined with external information to extract business intelligence e.g. a reseller connects his payment system to the accounting department of his main customer.

Figure 49 FRM application scenario

In this scenario relevant data sources are represented in the federation context as capabilities. Interactions are enabled by a set of mappings between the local and external models to allow a transformation of instances of one schema or ontology into another. Over time changes are expected on multiple levels such as providing services (represented as capabilities) and relationships. There are challenges for maintaining a system such as this, especially in an efficient manner that re-uses the existing effort in building mappings, as mapping can be purpose built (mapping of a specific supplier with a reseller) or simply reuse of existing standard mappings (such as a EDIFACT approach in B2B relations).

[125] Domains are organized in form of a graph which shows the types of relationships between them, e.g. peer to peer (enabling direct 1:1 sharing of capabilities between organisations) and federal (enabling sharing of capabilities between N domains).

Two key requirements can be identified for the FRM: The first requirement is that the management of change must be on the level of a relationship - explicit abstraction of information resources services, mappings and their federal relationships. The second is that reuse of mappings must be maximised and that integration of new sources with at least partially familiar contexts must be eased [Be13]. In other words the relationship mappings and applications are constantly changing. In addition, new mappings are created or older ones reused in different federation context situations.

Consequently the meta-data model for documenting the mappings needs to allow for flexible extension as the usage of ontology mappings are subject to change over time depending on the federation context changes that may result in new or changes to meta-data fields.

In such a dynamic environment the mapping related meta-data will be used by multiple applications or human stakeholders (in a semi-automated federation process) with different objectives. To accommodate such usage patterns the meta-data for the ontology mappings needs to be consistent and clearly documented to help the involved stakeholders to understand the intended meaning of the individual fields correctly [Ba99].

The FRM processes the federation context automatically to identify suitable mapping sets. This make it necessary for meta-data used for mapping discovery to be available in a machine-interpretable notation such as OWL and can be automatically generated as much as possible to support the automated FRM process.

Table 41 defines the two use-cases of the OM2R model for the FRM which will be analysed in more detail in section 10.6.

Table 41 Meta-Data Requirements for the FRM

Meta-Data Requirements	Motivation
Flexible extension of meta-data model	Dynamic environment leads to changing meta-data over time
Consistent meta-data	Application and logical consistency needed to multiple stakeholders over time
Explicit and clear documentation of meta-data model	Support correct interpretation of intended meaning

As the next step the specific use-cases for the OM2R model in the federation were derived.

10.5 Use-cases for OM2R based on FRM activity

Table 42 summarizes these two use-cases for the OM2R model being applied to the FRM which will be analysed in more detail in the next section.

Table 42 Use-cases of the OM2R for the FDM

Use-Case Title	Definition
Discovery of Ontology Mappings	This use-case focusses on the process of identifying relevant existing ontology mappings for a capability
Evaluation of Ontology Mappings	This use-case addresses the process of evaluation of identified mappings for the suitability to map relations between capabilities

10.6 Analysis

In this section the application of the OM2R model for federations is discussed. The first step shows how the OM2R model can extend the layer model of the FRM. Furthermore, it is presented how the OM2R allows a flexible extension of the meta-data model as well as supporting the creation of consistent and clearly documented meta-data to accommodate the FDM meta-data requirements. In addition, the OM2R meta-data is then discussed in more detail and shows that 62 % of the OM2R model can be generated automatically or deduced. This assists the population of the meta-data to address the dynamic character of federation and in support of the automated FRM activities.

10.6.1 D1: Representation of Mapping Meta-Data in Federations

Currently all ontology mappings in the FRM are represented in the Ontology Alignment Format or as extended EDOAL ontologies. The main focus is the representation of specific correspondences between the elements in the linked ontologies. Consequently all ontology mappings between two ontologies are represented in individual EDOAL ontology files.

[Br13] presented an approach to extend the existing model hierarchy with an additional layer that focusses explicitly on ontology mapping related meta-data by using the OM2R model. The resulting extended layer model for the federal relationship management is shown in Figure 50.

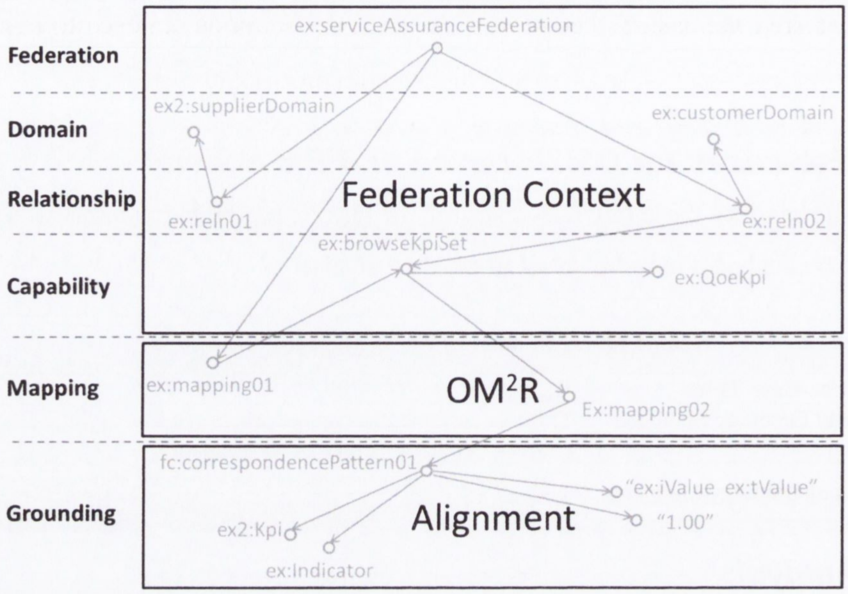

Figure 50 Layer-model for federal mappings based on a Figure found in [Br13]

The core of the FDM approach is the ability to identify how individual mappings are related to FRM federations, domains, relationships and capabilities. This creates a need to bridge between the FRM models of these entities and the actual alignment ontology (or its extension EDOAL). The OM2R model can be used to provide such a bridge and to serve as a model to document the general mapping management related meta-data with an emphasis on mapping lifecycle support – thus linking the specific meta-data with alignment correspondences and external resources such as capabilities in the federation context model.

As the OM2R model is expressed as an ontology, all meta-data information is stored as explicit and meaningful triples. This has the advantage that the OM2R model can be extended flexibly with new meaningful documentation elements. In addition, it can be enriched further with explicit relations between the models to address new reuse-situations. The changes can be made in the model, independent from particular implementation. As demonstrated in experiment 3, an editing tool will be able to process the model to generate a suitable user-interface which can adapt to on-going changes in the FRM easily. In this way the OM2R model can support the fundamental federation approach where relationships need to be managed explicitly as opposed to being deeply integrated in an application or service management component.

Federation has a dynamic character and a focus on reuse. To accommodate these features, meta-data related to ontology mappings need to be consistent to address the needs of multiple stakeholders who will use and reuse the mapping and as such evolve the meta-data [Ba99]. Experiment 3 has already provided evidence to show that the OM2R model can help a user to create documentation for ontology mappings that are application and logically consistent.

The OM2R model provides extensive documentation of the individual element embedded in the model. This supports users in interpreting the meaning and usage of the fields correctly. Manual editing is supported by providing specific field options for each meta-data field that a user may choose from and highlighting of compatible relations, e.g. field formal languages are OWL, RDF(S) and so on.

Current documentation of mappings in EDOAL is limited to one mapping set of correspondences between two data sets at a time. This is suitable for application of mappings but limits management and retrieval options. The benefit of the OM2R approach is that multiple alignments can be documented in one OM2R model. This is particularly relevant for the FRM - it can find application of a mapping in capabilities and the relevant mapping for a capability. Comparison, retrieval and reasoning can be supported better if the different mappings and their versions would be documented in one OM2R model.

10.6.2 D2: Characteristic of the Meta-Data Model

In section 9.5 two use-cases were identified where the OM2R can contribute toward the activities of the FRM. The discovery of ontology mappings in the federation and the evaluation of identified mappings to decide about a possible reuse.

The retrieval and evaluation of mappings in the federation is primarily automatic but can be semi-automatic for suitable cases in the FDM. For both processes it is vital that the model allows the documentation of relevant meta-data for the reuse process. The currently used alignment formats are focussed on the actual alignments but contain only spare life-cycle meta-data, see section 2.7 for full details. Experiment 2 in chapter 7 has shown that the OM2R provides a detailed meta-data model relevant for mapping reuse which supports directly the mapping evaluation as the FMD use-case.

The FRM processes the federation context automatically to identify suitable mapping sets. This makes it necessary that meta-data used for mapping discovery is available in a machine-interpretable notation so that it can be processed and automatically generated as much as possible to support the automated FRM process. In contrast to the current Ontology Alignment Format, which is expressed in RDF(S), the OM2R which is represented in OWL DL, can support automated reasoning and automated processing. Experiment 1 has shown that the OM2R is able to improve precision and recall for automated mapping discovery processes which are key for the first use-case of the FDM.

For the automated processes in the FDM it is vital that the generation of the meta-data can be automated to reduce the manual work-load and speed up the documentation process. OM2R is situated one layer above the actual alignment representation. These meta-data fields in the alignment format can be identified and used to pre-populate the OM2R automatically which can reduce the documentation effort for the documentation creator. Relations in the model can also be used to deduce field options from other previous entered information or extracted information. More specifically, the OM2R offers explicit relations (e.g. compatible) between fields which a suitable interface can interpret and use, e.g. to recommend compatible field combinations[126].

Table 43 provides an overview of the individual OM2R meta-data fields which are relevant for the individual activity state above. The column "Generation" indicates if the meta-data information needs to be entered manually or if it can be extracted from the alignment representation (if suitable meta-data is available), or can be deduced based on

[126] For example if the specific matching algorithm can be identified by parsing the alignment format meta-data, it is possible to deduce the used matching tool (alignment server 3.1) and matching generation type (automated) simply based on the compatible relation expressed in the OM2R. This may not always be possible to the full extent but at least allows a reduction of the number of valid options for a meta-data field.

relations expressed in the OM2R model (e.g. a specific matching algorithm is linked to matching tool linked to matching generation type) as shown above.

Table 43 Overview of OM2R meta-data fields to support FDM use-cases

Meta-Data field	Generation	OM2R - Meta-Data Fields
1) Discovery of Ontology Mappings (each field is available for target and source ontology)		
Name of ontologies	Manual / Extract	SourceOntology :Om2r:human_readable_name: "Biology Top Level Ontology"
Description of ontologies	Manual / Extract	OM2R:description
Location of ontology	Manual	OM2R:hasLocation (type url)
Creation date of ontologies	Manual	OM2R:hasCreationDate (type date)
Creator of ontologies	Manual / Extract	OM2R:hasCreator + details
Unique identifier for ontologies	Manual / Extract	OM2R:hasIdentifier
Notation of ontologies	Manual / Deduced	OM2R:hasNotation RDF/XML
Formal language of ontologies	Manual / Deduced	OM2R:hasFormalLanauge OWL
Formal language of the matching	Manual	OM2R:hasformalMatchingLanguage: EDOAL
Notation of the matching	Manual / Extract	OM2R: hasNotation: RDF/XML.
2) Evaluation of Ontology Mappings		
Complexity of the ontology	Auto generated	OM2R:hasClassCount 73, hasInstanceCount 3 hasPropertyClass 3
Design of the ontologies	Manual	OM2R:hasDesign om2r:deep_hierarchy.
Matching method	Manual / Deduced	OM2R:hasMethod (manual, automatic, mixed)
Matching process	Manual	OM2R:hasProcessStep
Matching tool	Manual / Deduced	OM2R:isTool AlignmentServer
Matching algorithm	Manual / Extract	OM2R:Algorithm :encodedIn: Java, OM2R Algorithm :hasJavaClass: org.stringComp, OM2R Algorithm :hasSource: freecode.org/a.zip
Algorithm is based on	Manual / Deduced	OM2R:isBasedOn rdfs:label, rdfs:class
Applied threshold	Manual / Deduced	OM2R:has_Applied_Threshold
Matching scope	Manual / Deduced	OM2R:hasScope (complete or partial=
Matching requirements	Manual	OM2R:hasMatchRequirements (text)
Location of matching	Manual	OM2R:hasLocation (type url)
Creation date of matching	Manual	OM2R:hasCreationDate (type date)
Creator of matching	Manual / Extract	OM2R:hasCreator + details
Mapping requirements	Manual / Extract	OM2R:hasMapping_Requirements
Mapping process	Manual	OM2R:hasProcessStep
Creation date of mapping	Manual	OM2R:hasCreationDate (type date)
Creator of mapping	Manual / Extract	OM2R:hasCreator + details
Formal language of the mapping	Manual	OM2R:hasformalMatchingLanguage: EDOAL
Notation of the mapping	Manual / Extract	OM2R: hasNotation: RDF/XML.

Table 43 shows that the OM2R model offers detailed documentation with 38 dedicated meta-data fields[127] and field options for documenting the ontology mapping lifecycle compared to only 23 which the ontology alignment notation current offers. 63 % (24 of the total of 38) of the meta-data fields can be populated automatically or deduced based on other selections. This supports the automatic processing in the FRM activities for mapping retrieval and evaluation. Please find in annex F more detailed discussion of the applied fields of the OM2R.

[127] Please note the fields such as ontology name, source are counted twice as they are available for the source and target ontology. This is necessary to make the number comparable to the ontology alignment notation where a similar approach is used.

10.7 Conclusion

In this case-study ontology mapping discovery and retrieval were identified as the two use-cases where the OM2R model can be applied beneficially for the FRM. The OM2R is able to address all identified meta-data requirements of the FRM with a focus on flexible extension and explicit documentation. Experiment 3 has demonstrated the ability of the OM2R model to support consistent documentation which is vital in the dynamic and reuse situations that will be necessary to support within federations over time.

The OM2R model is designed to support an effective online consumption of ontology mapping meta-data. However, the focus of the FRM is on an automated process for managing and using mappings to improve efficiency and speed. Section 9.6.2 has shown that the OM2R model can support this goal as 63 % of the meta-data fields can be pre-populated by harvesting existing fields from the alignment format or by deduction from existing selections (62 %).

One limitation of this case study is the analysis was focussed on the FRM activities which are a particular implementation of the federation approach. Other federation approaches might require different meta-data. However, as the OM2R model is represented as an ontology and can be extended on a model level, it can evolve without changes to tool being necessary. As such, the OM2R model is very capable to cope with new federation developments or changed requirements.

Overall the case study has shown that the OM2R can be a useful addition to the federation approach and help to implement the key requirement to manage relations and their meta-data independent from particular implementation. The OM2R can contribute towards a sharing of a common understanding of the ontology mappings creation and application lifecycle and as such can help to promote and support the reuse of mappings in the context of federations.

11 Conclusion, Contribution and Future Work

This section presents how the individual research objectives stated in section 1.2 were achieved in this thesis. Furthermore, the contribution of the OM2R model towards the ontology mapping and the Linked Data communities are discussed. The chapter concludes with an outlook on further work towards improved mapping meta-data creation and final remarks.

11.1 Achievements / Objectives

In this subsection the achievements towards the research objectives are discussed.

11.1.1 Achievement for RO1: Identify the requirements for meta-data documentation for applications such as mapping reuse

The first research objective of this thesis was the identification of requirements for meta-data documentation for ontology mapping to support mapping reuse and discovery. A first study was conducted in 2009 where 22 mapping representation formats were evaluated based on their use in 13 mapping tools and common matching languages [Th09]. In 2012 another study of meta-data usage for documenting ontology matching in the OAEI was completed (see chapter 8 for details). In 2013 a final evaluation of five tools and four common matching languages were conducted as part of the state of the art review for this thesis (see section 2.3).

These three studies provided an overview of the current state of the art of meta-data use for ontology mappings. They highlight the common features: heterogeneity of meta-data, lack of coverage of the ontology mapping lifecycle, inconsistent use and an absence of documentation of the meta-data model. This helped to derive the key requirements for mapping documentation: expressive, consistent, relevant and easy to interpret documentation with support for ontology mapping retrieval.

It is worth noting that all evaluated tools are designed with ontology matching[128] in mind rather than ontology mapping[129]. This focus on matching related tools was unavoidable as currently no dedicated ontology mapping creation tool exists but plenty of match candidate generators (see section 2.4).

11.1.2 Achievement for RO2: Design of a meta-data model for ontology mapping

The second research objective of this thesis was the design of a meta-data model for ontology mapping documentation that can support the creation of consistency and relevancy. To address these quality aspects four requirements for the OM2R model were derived based on an analysis of common meta-data challenges (see section 3.2) and the state of the art review (see chapter 2). Furthermore, a domain analysis which focussed on the extended ontology mapping lifecycle provided other requirements (see section 3.4). In section 3.5 six design feature were discussed which were designed to support the model meeting these requirements. As a result an ontology-based meta-data model (OM2R) was designed with 32 meta-data fields (see section annex H). The second experiment (see section 6.1) created a relevance ranking of the proposed meta-data fields in the OM2R model. This helped to improve the OM2R model. Section 3.8 shows the details of the evolution of the OM2R model.

11.1.3 Achievement for RO3: Evaluation of the OM2R model

The third research objective was to evaluate the support the meta-data model provides for the creation of consistent and relevant documentation for ontology mappings. To address this objective three experiments were completed.

Experiment one provided evidence that the OM2R model can improve mapping retrieval effectiveness and efficiency compared to retrieval processes based on the existing meta-data model such as Ontology Alignment Format (see section 5.7). More specifically, an OM2R based retrieval tool returned more relevant mappings (higher recall ratings in 15

[128] To provide a consistent terminology within this thesis, ontology matching is defined as the process of identifying equivalence candidate correspondences between *ontology elements* (such as class, properties) based on an automated algorithm or manual evaluation as defined in [Eu04].

[129] Ontology mappings are defined as a process of relating the vocabulary of two ontologies sharing a domain in such a way that the structure of ontological signatures and their intended interpretations are respected [Ka03].

out of 16 discovery tasks) and the results were more accurate (higher precision ratings in 10 out of 16 discovery tasks).

Relevance is a vital quality aspect of meta-data models (see section 3.2.1). The OM2R addresses this aspect in its three evolution stages. The greatest impact on the OM2R model evolution had the result of the relevance rating experiment (see section 6.7). In this second experiment users provided feedback to indicate that all proposed meta-data fields of the OM2R model are relevant. However, the extent of relevancy varied from strong agreement (normalized 44 out of 49 participants) to (5.3 out of 49) for some fields. This resulted in an improvement of the OM2R model, e.g. fields where combined (matching objective and requirements) and others more specialised (the name field for ontologies were split into three more detailed meta-data fields).

Experiment three provided evidence that the model can support application and logical consistency on a high level, i.e. the experiment showed that the participants achieved an 86 % recall for logical consistency and 78 % recall for application consistency (see section 7.5.2 and 7.5.3). Furthermore, the experiment collected evidence to support the claim that an ontology based model representation of the OM2R model can provide a higher support for application and logical consistency compared to an index based representation. More specifically, a recall for logical consistency support of 16 % for an index based representation was measured compared to 86 % for an ontology based structure of the OM2R (see section 7.7.2). These results support the design decisions to represent the OM2R model as an ontology (D1) and to focus on meaningful relation between fields (D2), e.g. the compatible relations allows the recommendation of valid field combinations to the user (see section 4.4.2).

The third experiment could not identify any significant differences in the support for the creation of consistent documentation between experienced and novice users (see section 7.7.3 and 7.7.4. This provides evidence to the claim that the OM2R model can support real life applications where users with a diverse experience levels are expected. This claim is further supported by two case studies which demonstrate how the OM2R model can be applied to the OAEI (see chapter 8) and for federation (see chapter 9).

A limitation of experiment three (see section 7.4.1) is the lack of available real life ontology mappings with suitable meta-data that describe how the mappings were created.[130] It was therefore necessary for the author of this thesis to create the sample ontologies and mappings for this experiment. To mitigate the resulting bias all used ontologies in this experiment where based on scenarios and input provided by the research project FAME [Br12].

11.1.4 Achievement Towards Research Question

The conducted experiments and case-studies provide evidence to show that the OM2R model can support ontology engineers in the creation of consistent and relevant ontology mapping documentation (see section 6.7 and 7.8). Evaluations provided evidence to indicate that the OM2R model can improve mapping discovery (see section 5.7). To sum up the author of this thesis has confidence that the OM2R model can support users in ontology mapping retrieval with more explicit, detailed, predictable and easy to interpret meta-data about the mapping creation lifecycle.

11.2 Contribution

Table 44 provides an overview of all publications created by the author of this thesis in relation to OM2R model, with some brief remarks about what they are about and their relationships to the thesis. The listing is ordered by publication year.

[130] An example mapping from DBpedia to IMDb can be found on http://mappings.dbpedia.org/index.php/Mapping_bg:IMDb_title.

Table 44 Literature Overview of Hendrik Thomas in relation to the OM2R model

Year	Type	Details
2008	Reference	Hendrik Thomas, Bernd Markscheffel, Tobias Redmann, From Subjects to Concept Clouds - Why semantic mapping is necessary, 1st International Conference on Knowledge Federation, Inter University Centre Dubrovnik, Croatia, October 20-22, 2008, 2008
	Research Objective	R1 - Identify the requirements for meta-data documentation for applications such as mapping reuse.
	Summary	The paper highlights the need for Semantic Mappings in the digital library domain to handle federation
2009	Reference	Bernd Markscheffel, Hendrik Thomas, Tobias Redmann, Developing Topic Maps Applications: Lessons Learned from a Digital Library Project, IADIS International Conference e-Society 2009, Barcelona, Spain, February 25-28, 2009, 2009, pp51 - 59
	Research Objective	R2 - Design a meta-data model for ontology mapping documentation that can support the creation of consistent and relevant ontology mapping documentation.
	Summary	The paper presents an approach how Topic Maps can be utilised to management information integration in digital libraries with a focus on mappings and related meta-data
2009	Reference	Thomas, H., O'Sullivan, D., Brennan, R.: Evaluation of Ontology Mapping Representations. In: Workshop on Matching and Meaning, Part of the as part of the AISB 2009 Convention, April 9th 2009. Edinburgh, Scotland 2009
	Research Objective	R1 - Identify the requirements for meta-data documentation for applications such as mapping reuse.
	Summary	The paper presents the results of a state of the art study of 22 common mapping representations with a focus on the supported meta-data, processes and their limitations
2009	Reference	Thomas, H., O'Sullivan, D., Brennan, R.: Ontology Mapping Representations: a Pragmatic Evaluation. In: 21st International Conference on Software Engineering and Knowledge Engineering. SEKE 2009, 1 - 3 July 3, Boston, 2009, pp. 228 - 232.
	Research Objective	R1 - Identify the requirements for meta-data documentation for applications such as mapping reuse.
	Summary	The paper presents results of a state of the art study of 22 common mapping representations used in 13 mapping and matching tools to show current supported meta-data and their characteristics.
2009	Reference	Thomas, Hendrik: Ontology Mapping Management. Poster for European Summer School on Ontological Engineering and the Semantic Web 2009, Cercedilla, Spanien, 5th - 12th June 2009.
	Research Objective	R2 - Design a meta-data model for ontology mapping documentation that can support the creation of consistent and relevant ontology mapping documentation.
	Summary	Presentation of the OM2R model design with a focus on process and layer approach
2010	Reference	Kevin Feeney, Rob Brennan, John Keeney, Hendrik Thomas, Dave Lewis, Aidan Boran, Declan O'Sullivan, Enabling decentralised management through federation, Computer Networks Journal (Elsevier), 54, (16), 2010, p2825-2839
	Research Objective	R3: Evaluate that the meta-data model can support the creation of consistent and relevant documentation for ontology mappings in support of applications such as mapping reuse.
	Summary	The paper presents a concept to enable decentralised management through federation. The OM2R is presented as a model to document meta-data for ontology mappings.
2011	Reference	Thomas, H., Brennan, R. O'Sullivan, D.: MooM - a Prototype Framework for Management of Ontology Mappings. In: Proceedings of the 25th IEEE International Conference on Advanced Information Networking and Applications, Singapore, 22-25 March, 2011, IEEE, pp 548 – 555.
	Research Objective	R2 - Design a meta-data model for ontology mapping documentation that can support the creation of consistent and relevant ontology mapping documentation.
	Summary	The paper presented the requirements for meta-data management for ontology mapping and a prototype tool implementation

Year	Type	Details
2011	Reference	Rob Brennan, Kevin Feeney, Brian Walshe, Hendrik Thomas, Declan O'Sullivan, Explicit Federal Relationship Management to Support Semantic Integration, 1st IFIP/IEEE Workshop on Managing Federations and Cooperative Management, Dublin, 2third May 2011, IEEE, 2011, pp1148-1156
	Research Objective	R3: Evaluate that the meta-data model can support the creation of consistent and relevant documentation for ontology mappings in support of applications such as mapping reuse.
	Summary	The paper highlights how the OM2R can be utilised to management meta-data for ontology mappings in federations
2012	Reference	Thomas, H., Brennan, R., O'Sullivan, D.: Using the OM2R meta-data model for ontology mapping reuse for the ontology alignment challenge - a case study. In proceedings of the 7th International Workshop on Ontology Matching (OM-2012) collocated with the 11th International Semantic Web Conference (ISWC-2012) Vol 946, http://ceur-ws.org/Vol-946/ Boston, MA, USA, November 11, 2012.
	Research Objective	R3: Evaluate that the meta-data model can support the creation of consistent and relevant documentation for ontology mappings in support of applications such as mapping reuse.
	Summary	This paper presents a case-study to show how the OM2R can be applied beneficially to represent and management meta-data for ontology matchings in the OAEI

The key contribution of this thesis is the development of a meta-data model (OM2R) for documentation of ontology mappings for applications such as mapping retrieval towards reuse. The OM2R model is designed as an independent meta-data layer and can therefore be used in parallel with existing tools and mapping representations. OM2R can support the documentation process by providing ontology engineers an expressive representation, embedded self-describing fields, process representation, editing workflows and relevant fields to support mapping retrieval. Thanks to the OM2R ontology based structure the model can be extended flexibly to address new developments and individual mapping reuse scenarios. This is essential as current mapping applications and particular mapping reuse is rare.

The evidence collected in experiment three indicates that the OM2R will help ontology engineers to create more consistent documentation. This relates to application consistency in terms of correct and complete documentation. More specifically, experiment three showed that an ontology-based editor could offer a 2.2 time better recall for application consistency and 4.6 time better recall for logical consistency compared to a tool based on an index-based representation of the model (see segment 7.8). The evidence suggests a particular good support of logical consistency to avoid inconsistent statements (see section 7.7.2).

Overall the OM2R will provide ontology engineers with a meta-data model to create relevant, consistent, explicit and easy to interpret meta-data about the ontology mapping lifecycle.

Thanks to the embedded meta-data documentation in the OM2R model it can support third party researchers and mapping users in understanding the intended meaning of the meta-data fields and their relation between each other. Experiment three has shown that the OM2R model can support domain experts and novice users on a similar level (see section 7.7.3 and 7.7.4), which again assists to more consistent documentation creation in a real life scenario.

The Linked Data community [Um10] will benefit from this work as well. The fast growth of current linked datasets increases the need for better interoperability support to allow easy and flexible consumption of data from heterogeneous sources [He11]. An example of the Linked Data community efforts toward this need is the R2R Framework for Translating RDF data from the Web to a target vocabulary [Bi10].

The focus of this community is on automated processes. The OM2R model can address this particular requirement as it offers an explicit and as such easy process-able representation of the meta-data and rich set of relations between them. The implementation independent nature of the OM2R model makes it use and reused flexibly and easy. It allows the formulation of very precise retrieval queries, e.g. experiment one demonstrated how SPARQL queries can be used for mapping retrieval (see section 5.7). The formal language OWL DL used for the OM2R supports reasoning which can improve further the automated identification of relevant mapping candidates. An example for such an application can be found in the case study on federation presented in chapter 9. Overall the OM2R model will assist mapping retrieval and thus make data consumption and linkage to other repositories easier to maximize network effects.

A minor contribution is a practical guide for the "Ontology Alignment Evaluation Initiative" (OAEI) [Eu11a] that shows how the current limitations of documenting ontology alignments can be improved. The OM2R model can help participants of the OAEI to document their matching submissions by provides a standard template and a detailed model for the documentation. It also assists third party researchers to understand the meaning of the documentation fields for a rigor analysis. Most of all it will benefit the OAEI organisers as the OM2R can improve relevance and consistency of the documentation and thus make the identification of trends over time easier. The case study, first presented at [Th12], show cases that the OM2R model does not have to focus on a complete ontology mapping. It can be used throughout the lifecycle and be used beneficial to document only the identification and matching phase.

Overall, the author of this thesis argues that the OM2R as an improved meta-data model can help to leverage the experience gained in the OAEI to extend its focus from a pure testing platform [Eu11a] to a large scale alignments repository [Eu11b]. Thus the OM2R model can provide active and practical help to the way matching and mappings are documented retrieval and potentially reused on a large scale over time.

11.3 Future Work

The OM2R is a well designed a meta-data model. It provides the foundation for future research efforts which need to focus on the practical application of the model to document existing ontology mappings. So far the OM2R model has been applied by a number of users but only in controlled lab environments (see chapter 5, 6, 7). Only the case study for the OAEI relates to a live usage domain (see chapter 8). A clear objective for a follow up research needs to be the motivation of communities related to mapping and matching to utilise this model.

For this purpose the author of this thesis has already suggested a research-collaboration with the OAEI. The first step for such an effort was made with the case-study published in [Th12]. The aim of the OAEI and of the matching community is shifting more and more towards the management related aspects [Sh12]. As a next step the author hopes to motivate the OAEI to invite their participants to submit their matching results with a corresponding OM2R instance. The explicit research focus of OAEI and the positive feedback gained in a research conference [Th12] makes this possible. The OAEI can be used in parallel to the existing processes and as such make the integration easier. Overall this collaboration could provide valuable insights and help to improve the model further on the basis of practical experiences.

Another promising application area for the OM2R is the Linked Data community. The R2R Framework is an example to represent and generate mappings of Web dataset to a target vocabulary [Bi10].[131] Currently the framework focusses on the actual mapping pairs and provides only very limited meta-data, e.g. r2r:targetDataset specifies the target data source.[132] The OM2R tool could be used to offer more details on how the mapping was created and used. This would assist greatly in the mapping discovery for the Linked Data community as demonstrated in section 5.7.

The research presented in this thesis also highlighted that an adequate tool support is needed to utilise the extensive and expressive information available in the OM2R model to its full potential. The OM2R editor created for experiment three demonstrates how such benefits can be achieved e.g. embedded meta-data and recommendation of compatible field combinations (see section 4.4.2). However, the tools were designed with the purpose in mind to evaluate the individual experiment hypothesis. The editing tool offered limited functions and a basic interface design. Experiment three included a SUS question on usability aspects of the editing tool. The OM2R editor achieved a score of 65.1 out of 100. This is lower than the average score of 68 identified in a study on matching tools in 2011[133] (see section 7.7.6). It is worth noting that despite these low usability results, the achieved support for relevance and consistency is considerable which are based in the model content rather on the specific implementation. However, this still implies further research on tools is needed to investigate the full potential of how the expressive mapping meta-data offered by the OM2R model can be used to support meta-data creation and mapping reuse.

To address the current limitations of the OM2R tools the author suggests the development of an open source web based editor which allows an on the fly generation of OM2R model instances. Such a tool can only be successful if it is combined with an

[131] More information can be found on http://wifo5-03.informatik.uni-mannheim.de/bizer/r2r/
[132] More information can be found on http://wifo5-03.informatik.uni-mannheim.de/bizer/r2r/
[133] See http://www.measuringusability.com/sus.php for full details

online repository where OM2R model instances can be uploaded, stored and retrieved. Thus the next research should combine a user friendly meta-data creation with easy consumption of the meta-data.

Another promising research direction is the aim to make the meta-data on ontology mapping more adaptive and less static. The expression of the OM2R model as an OWL DL provides the basis for reasoning. It needs to be investigated further how reasoning can be used to populate the meta-data to the extent that the meta-data can monitor changes to the application of the mappings automatically.

11.4 Final Remarks

The OM2R model offers ontology engineers an easy and implementation neutral meta-data model to document the creation lifecycle of ontology mappings. This helps to prepare the Semantic Web and Linked Data community for the future where knowledge integration will be an increasingly common and vital task where more descriptive information on ontology mappings will be needed for retrieval and reuse decisions. The OM2R model gives the creators the ability to create relevant and consistent documentation of mappings which is vital in such a heterogeneous and ever changing domain.

It is the hope of the author of this thesis that the OM2R model will make sharing and reusing of ontology mappings easier and more popular thanks to a shared understanding of the ontology mapping creation and better supported mapping retrieval.

Bibliography

[Ab89] Abbott, A.: *The system of professions: An essay on the division of expert labour*. Chicago: University of Chicago Press., 1989.

[An04] Antoniou, G., van Harmelen, F.: *A Semantic Web Primer (Co-operative Information Systems)*. The MIT Press, 2004.

[Ap12] Apache Jena: *Java Framework for building Semantic Web application* http://jena.apache.org , 2012.

[Ar11] Arenas, M., Perez, J.: *Querying semantic web data with SPARQL*. in: Proceedings of the thirtieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems (PODS '11), 2011, pp. 305 – 316.

[As06] Assem, M. v. et. al.: *A Method to Convert Thesauri in SKOS*, in: Proceedings of the third European Semantic Web Conference (ESWC 2006), Budva, Montenegro, Berlin, Springer, 2006, pp 95–109.

[Ba99] Baeza-Yates, R., Ribeiro-Neto, B.: *Modern Information Retrieval*. ACM Press, Addison-Wesley, New York, NY 1999, pp. 75 – 78.

[Ba00] NISO, *the National Information Standards Organization 2000.*

[Ba00a] Basili, V. R., Caldiera, G., Rombach, H. D., *Goal Question Metric Approach*, ftp://ftp.cs.umd.edu/pub/sel/papers/ gqm.pdf, 2000.

[Ba09] Bangor, A., Kortum, P.: *Determining What Individual SUS Scores Mean: Adding and Adjective Rating Scale*. in: JUS Journal of Usability Studies, Vol 4 Issue 3 may 2009, pp 114–123.

[Be01] Berners-Lee, T., Hendler, J., Lassila, O.: *The Semantic Web - A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*, Scientific American Magazine, http://www.sciam.com/article. cfm?id=the-semantic-webUNDprint=true, 2001.

[Be04] Beckett, D. (eds.): *RDF/XML Syntax Specification* (Revised), http://www.w3.org/TR/rdf-syntax-grammar/, 2004.

[Be04a] Beigi, M., Calo, S., Verma, D.: *Policy Transformation Techniques*. In: Policy-based Systems Management, in: proceedings of IEEE Policy 2004, Yorktown, NY, 2004.

[Be05] Bernstein, P. A.: *The many roles of meta data in data integration*. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, Baltimore, Maryland, USA, June 14-16, 2005.

[Be08] Bernstein, P. A., Haas, L. M.: *Information integration in the enterprise*. in: Commun. ACM 51, 9 September, 2008, pp 72-79.

[Be11] Bellahsene, Z., Bonifati, A., Rahm E. [eds.]: *Schema Matching and Mapping*, Springer, 2011

[Be11b] Z. Bellahsene, A. Bonifati, and E. Rahm, Eds., *Schema Matching and Mapping*. Springer, 2011.

[Be11c] Bernstein, P. A., Madhavan, J., Rahm, E.: *Generic Schema Matching, Ten Years Later*. in: Proceedings of the 37th International Conference on Very Large Data Bases (PVLDB) Volume 4, 2011, pp 695-701.

[Be11d] Brennan, R., Feeney,K., Walshe, B. at al: *Explicit Federal Relationship Management to Support Semantic Integration*, in: proceeding of the 1st IFIP/IEEE Workshop on Managing Federations and Cooperative Management, Dublin, 2third May 2011, IEEE, 2011, pp1148-1156.

[Bi00] Bisbort, Al., Osborne, L. B: The Nation's Library: The Library of Congress, Washington, D. C. (Library of Congress, 2000.

[Bi10] Bizer, C., Schultz, A.: The R2R Framework: Publishing and Discovering Mappings on the Web. 1st International Workshop on Consuming Linked Data (COLD 2010), Shanghai, November 2010.

[Bi09] Bizer, C.,Heath, T. Berners-Lee, T.: *Linked Data - The Story So Far*. In: International Journal on Semantic Web and Information Systems (IJSWIS), Volume 5, Issue 3, 2009.

[Bo03] Bouquet, F. et al.: C-owl: *Contextualizing ontologies*. In: Proceedings of the 2nd International Semantic Web Conference, ISWC'03, Lecture Notes in Computer Science, Florida, USA, Springer Verlag, pp. 164–179 2003.

[Br04] Bruijn, Jos d., Polleres: *Towards an Ontology Mapping Specification Language for the Semantic Web* DERI Technical Report 2004-06-30.

[Br05] Bruijn, Jos d. et al.: *OWL DL vs. OWL flight: conceptual modeling and reasoning for the Semantic Web.* in WWW'05: Proceedings of the 14th international conference on World Wide Web, New York, USA, 2005, pp 623–632.

[Br12] Walshe, B. *Identifying Complex Semantic Matches. In:* The Semantic Web: Research and Applications Lecture Notes in Computer Science Volume 7295, 2012, pp 849-853.

[Br13] Brennan, R., Walshe B., O'Sullivan, D.: *Managed Semantic Interoperability for Federations*: In: Journal of Network and Systems Management Submitted to Journal JNSM

[Br13b] Brennan, R., Feeney, K, Song, Y, O'Sullivan, D.: *Secure Federated Monitoring of Heterogeneous Networks*, in: IEEE Communications Magazine, 2013 (to appear)

[Br13c] Brennan, R., Feeney, K., Odhrán, G: *Publishing Social Sciences Datasets as Linked Data: a Political Violence Case Study*, in Proceedings ENRICH 2013 at SIGIR 2013, Dublin, 1 August 2013.

[Br96] Brooke, J.: *SUS: a "quick and dirty" usability scale".* In: P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland. Usability Evaluation in Industry. London: Taylor and Francis, 1996.

[Ca03] Caplan, P. et al.: Metadata fundamentals for all librarians. Chicago, IL: American Library Association, 2003.

[Ca11] Carterette, B., Voorhees, E. M.: *Overview of Information Retrieval Evaluation.* in: Current Challenges in Patent Information Retrieval, The Information Retrieval Series Volume 29, 2011, pp 69-85.

[Ch99] Chowdhury, G. G.: *Introduction to Modern Information Retrieval*, London: Library Association, 1999.

[Ch12] Leduc, C., Euzenat, J.: *Project homepage of the NeOn alignment plugin,* http://neon-toolkit.org/wiki/2.3.1/Alignment, 2012.

[Ch13] Chiarcos, C.at al: *Towards Open Data for Linguistics: Linguistic Linked Data.* in: New Trends of Research in Ontologies and Lexical Resources, 2013, pp 7-25.

[Co97] Cooper, W. S.: A *definition of relevance for information retrieval,* In: Information Storage and Retrieval Volume 7, Issue 1, 1971, Pages 19–37.

[Ch06] Choi, N. Song, I.-Y., Han, H.: *A survey on ontology mapping.* In:SIGMOD Record, vol. 35, no. 3, 2006, pp. 34–41.

[Cr09] Cruz, I. F., Antonelli, F. P, Stroe, C.: *Agreementmaker: Efficient matching for large real-world schemas and ontologies*, in: PVLDB, vol. 2, no. 2, pp. 1586–1589, 2009.

[Da10] David, J., Euzenat, j. Scharffe, F. Trojahn dos Santos, C.: *The Alignment API 4.0*,IOS Press, 2010 http://www.semantic-web-journal.net/sites/default/files/swj60_1.pdf.

[Da11] David, J., Euzenat, J., Scharffe, F., Trojahn dos Santos, C.: *The alignment API 4.0.* in: Semantic Web, vol. 2, no. 1, 2011, pp. 3–10.

[De12] Dempsey, L: Thirteen Ways of Looking at Libraries, Discovery, and the Catalog: Scale, Workflow, Attention. In: EDUCAUSE Review Online, http://www.educause.edu/ero/article/thirteen-ways-looking-libraries-discovery-and-catalog-scale-workflow-attention, 2012.

[Di00] The American Heritage: *Dictionary of the English Language*, Fourth Edition, Houghton Mifflin Company, Published by Houghton Mifflin Company, 2009

[Du02] Duval, E.: *Metadata Principles and Practicalities.* in: D-Lib Magazine, April 2002 Volume 8 Number 4, 2002

[Do05] Doan, A. Halevy, A: *Semantic Integration Research in the Database Community: A Brief Survey* . in: AI Magazine, Special Issue on Semantic Integration, 2005.

[Du02] Duval, E.: *Metadata Principles and Practicalities.* in: D-Lib Magazine, April 2002 Volume 8 Number 4, 2002.

[Du10] Durbin, M.: *Maintaining Standards - managing metadata consistency across collections and tools.* In: Proceedings The 5th International Conference on Open Repositories (OR2010) Madrid, Spain, 6-9 July 2010

[Du12] *Dublin Core Metadata Initiative (DCMI)*, http://dublincore.org, 2012

[Dr12] *Drupal Project page*, http://drupal.org/, 2012

[Ed07] Edvardsen, L. Himyr F. et al: *Metadata Challenges in Introducing the Global Ieee Learning Object Metadata (Lom) Standard in A Local Environment.* In: proceeding of Webist 2007 - third International Conference on Web Information Systems and Technologies, 2007, pp, 427-432

[Eu04] Euzenat, J., *An API for ontology alignment*, in: Proceedings of the International Semantic Web Conference (ISWC 2004), Springer, Berlin, Germany, 2004, pp. 698-712.

[Eu05] Euzenat, J. : *Alignment infrastructure for ontology mediation and other applications*. In: Proceddings International Workshop on Mediation in Semantic Web Services (MEDIATE), 2005, pp. 81– 95.

[Eu06] Euzenat et al. *D2.2.6: Specification of the delivery alignment format*. Specification of the delivery alignment format, 2006.

[Eu07] Euzenat, J. Shvaiko, P. *Ontology matching*. Springer, 2007.

[Eu07a] Euzenat, J., Scharffe, F., Zimmermann, A.: Expressive alignment language and implementation. Deliverable 2.2.10, Knowledge Web NoE, 2007.

[Eu08] Shvaiko, P., Euzenat, J.: *Ten Challenges for Ontology Matching*. in: Proceedings of the 7th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE), 2008.

[E08a] Euzenat, J.: Alignment API and server (version 3.2) Documentation, 2008 http://gforge.inria.fr/docman/view.php/117/5036/alignapi.pdf.

[Eu11a] Euzenat, J., Meilicke, C., Shvaiko, P. et al.: *Ontology Alignment Evaluation Initiative: six years of experience*. in: Journal on Data Semantics, Volume XV Edition 6720, pp 158-192, 2011.

[Eu11b] Euzenat, J. et al: *Results of the Ontology Alignment Evaluation Initiative 2011*. in: Proceeding of the 6th ISWC workshop on ontology matching, Proceedings of the 10th International Semantic Web Conference ISWC-2011, 2011, pp 85-110, 2011.

[Fa09] Falconer S.: *Cognitive support for semi-automatic ontology mapping*, Thesis, 2009.

[Fa12] *Federated autonomic management of end-to-end communication services (FAME)*, SFI-funded research project, home page http://www.fame.ie/, 2012.

[Fu93] Fugmann, R.: *Subject Analysis and Indexing: Theoretical Foundation and Practical Advice*. Frankfurt a. M., Indeks, 1993.

[Fe10] Feeney, K., Brennan, R., Keeney, J., et al. : *Enabling decentralised management through federation*. In: Computer Networks, 54(16):2825–2839, 2010.

[Fr12] *Friend of a Friend* project homepage http://www.foaf-project.org/docs/specs

[Ga06] Garshol, L. M., Moore, G.: *ISO/IEC JTC1/SC34, Information Technology – Document Description and Processing Languages*, http://www.isotopicmaps. org/sam/sam-model/, 2006

[Gh09] Ghazvinian, A., Noy, N. F. , Jonquet, C. et al : *What Four Million Mappings Can Tell You About Two Hundred Ontologies*. in: Proceedings of International Semantic Web Conference (ISWC), Washington DC 2009.

[Gl09] Glaser, H., Jaffri, A., Millard, I. C.: Managing Co-reference on the Semantic Web. In: Linked Data on the Web (LDOW2009), Madrid, Spain, 2009.

[Gi12] Gibbon, D.C. Liu,Z. , Basso, A., Shahraray B.: *Automated Content Metadata Extraction Services Based on MPEG Standards*. In: The Computer Journal December 6, 2012

[Gr94] Gruber, T. R.: *Towards Principles of Ontologies Used in Knowledge Sharing*. In: International Journal of Human-Comper Studies, Nr. 43, S. 907-928, 1994.

[Ha05] Haase, P.: *Ljiljana Stojanovic: Consistent Evolution of OWL Ontologies*. ESWC 2005: 182-197 ESWC 2005.

[Ha08] Haase, P. Lewen, H., Studer, R. et al.: *The NeOn Ontology Engineering Toolkit* . in: WWW 2008 Developers Track, 2008.

[Ha09] Hanif, M. S., Aono, M.: *An efficient and scalable al-gorithm for segmented alignment of ontologies of arbitrary size*. In: Journal of Web Semantics, vol. 7, no. 4, pp. 344–356, 2009

[He11] Heath T., Bizer C.: *Linked Data: Evolving the Web into a Global Data Space (1st edition)*. Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool.2011

[Ho02] Hjørland, B. & Sejer Christensen, F.: *Work tasks and socio-cognitive relevance: a specific example*. In: Journal of the American Society for Information Science and Technology, 53(11), 2002, pp 960-965

[Hu08] Hu, W., Qu, ,Y. Cheng, G.: *Matching large ontologies: A divide-and-conquer approach*. in: Data and Knowledge Engineering, vol. 67, no. 1, 2008, pp 140–160.

[IS04] *ISO/IEC 11179-1 Information technology — Metadata registries (MDR) - Part 1: Framework, Second edition* 2004-09-15 http://standards.iso.org/ittf/PubliclyAvailableStandards/ c035343_ISO_IEC_11179-1_2004(E).zip.

[Je09] Jennings, B., Brennan, R., Donnelly, W. et al.: *Challenges for federated, autonomic network management in the future Internet*. In: 1st IFIP/IEEE International Workshop on Management of the Future Internet, New York, 5 June 2009, IEEE, 2009.

[Je11] David, J., Euzenat, J., Scharffe, F., Trojahn dos Santos, C.: *The Alignment API 4.0, in:* Journal Semantic Web Issue Volume 2, Number 1, 2011. http://iospress.metapress.com/content/4164891n48p5v826/

[Ke52] Kleene, S.,: *Introduction to Metamathematics*, North-Holland Publishing Company, Amsterday, New York, 1952.

[Ka03] Kalfoglou, Y., Schorlemmer, M.: *Ontology mapping: the state of the art*. In: The Knowledge Engineering Review, 18 (1), 2003, pp. 1-31

[Ke11] Feeney, K., Foley, S. N., Brennan, R.: *A Trust Model for Capability Delegation in Federated Policy Systems*. in: Sixth International Conference on Risks and Security of Internet and Systems, Timioara, Romania, Sept 26-28, 2011, 2011, pp47 – 54

[Ku98] Kuny, T., Cleveland, G.: The Digital Library: Myths and Challenges. in Journal of International Federation of Library Association and Institutions (IFLA) Vol. 24, 2, 1998, pp. 107–113

[La07] Lambrix P, Tan H, Jakoniene V, Strömbäck L. : *Biological Ontologies*. In: Baker, Cheung, (eds), Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences, pp 85-99, Springer, 2007.

[La08] Lanzenberger, M., Sampson, J.: Making Ontologies Talk: *Knowledge Interoperability*. in: Semantic Web, in IEEE Intelligent Systems, November/December 2008, pp 72-86.

[La09] Lambrix P, Liu, Q: *Using partial reference alignments to align ontologies*. In: Proceedings of the 6th European Semantic Web Conference, LNCS 5554, 188-202, Heraklion, Greece, 2009.

[Lo09] *LOM (Learning Object Metadata)* http://ltsc.ieee.org/wg12/, 2009.

[Ma06] van Assem, M., Malaisé, V., Miles, a. et al.: *Method to Convert Thesauri to SKOS*. In: The Semantic Web: Research and Applications Lecture Notes in Computer Science Volume 4011, 2006, pp 95-109

[Ma11] Manenti, L. Sartori, F.: *Metadata support to retrieve and revise solutions in case-based reasoning*. in: International Journal of Metadata, Semantics and Ontologies, 2011 Vol.6, No.3/4, pp.185 – 194

[Ma99] Madnick, S.: *Metadata Jones and the tower of Babel: The challenge of large-scale semantic heterogeneit*. In: Proceedings IEEE Meta-Data Conf, 1999.

[Mc04] McGuinness, D. L., Harmelen, F. van (eds.): *OWL Web Ontology Language - Overview*, http://www.w3.org/TR/owl-features/, 2004.

[Me09] Meilicke, C.: *The Relevance of Reasoning and Alignment Incoherence in Ontology Matching*. In: Proceeding ESWC 2009 Heraklion Proceedings of the 6th European Semantic Web Conference on The Semantic Web: Research and Applications, 2009, pp. 934 – 938.

[Me10] van der Meij, L., Isaac, A., Zinn, C.: *A web-based repository service for vocabularies and alignments in the cultural heritage domain*. In: Proceedings 7th Extended Semantic Web Conference (ESWC), 2010, pp. 394–409.

[Ma11] Manenti, L. Sartori, F.: *Metadata support to retrieve and revise solutions in case-based reasoning*. In: International Journal of Metadata, Semantics and Ontologies, 2011 Vol.6, No.3/4, pp.185 – 194.

[Ma12] Marshall S. et al: *Emerging practices for mapping and linking life sciences data using RDF - a case series*. in: Journal of Web Semantics: Science, Services and Agents, Volume 12, 2012 Semantic web journal Volume 2, Nr.1 pp. 3-10, 2011.

[Ma12] *MARC 21 Format for Bibliographic Data that resulted from Update No. 15* (September 2012) http://www.loc.gov/marc/bibliographic/, 2012

[Me09] *METS (Metadata Encoding and Transmission Standard)*, http://www.loc.gov/standards/mets/, 2009.

[Me13] Meskill, B., Balasubramaniam, S., Brennan, R.: *Federation Lifecycle Management Incorporating Coordination of Bio-inspired Self-management Processes*. in: Journal of Network and Systems Management, 2013 (to appear).

[Mi09] Miles, A., Bechhofer, S.: *SKOS Simple Knowledge Organization System Namespace Document - HTML Variant 18 August 2009 Recommendation Edition*, http://www.w3.org/TR/2009/REC-skos-reference-20090818/, 2009.

[Mo09] Moser T. et al.: *Semantic Service Matchmaking in the ATM Domain Considering Infrastructure Capability Constraints*. In: Proceedings 21st International Conference on Software Engineering & Knowledge Engineering. SEKE 2009, 1-3.06, Boston, pp. 222 – 227, 2009.

[Na04] Najjar, J. et al.: *Usability Evaluation of Learning Object Indexation: The ARIADNE Experience*. In: Proceedings of ECEL 2004 European Conference on e-Learning, pp. 281-290, 2004.

[Ni93] Nielsen J.: *Usability Engineering*, Academic Press, San Diego/New York, 1993.

[Ni04] NISO. Understanding Metadata. Bethesda, MD: NISO Press,http://www.niso.org/publications/press/UnderstandingMetadata.pdf, 2004.

[No04] Noy, N.: *Semantic Integration: A Survey of Ontology-based Approaches Sigmod Record*. In: Special Issue on Semantic Integration, 2004.

[No08] Noy, N. , Griffith, N., Musen, M: *Collecting community based mappings in an ontology repository*. in: proceedings of 7th International Semantic Web Conference (ISWC), 2008, pp. 371– 386.

[Ni09] NISO: *A Framework of Guidance for Building Good Digital Collections. In:* The third edition MD: NISO Press. http://www.niso.org/publications/rp/framework3.pdf, 2009.

[OM12] OMV - Ontology Metadata Vocabulary, Sourceforge project page, http://sourceforge.net/projects/omv2/, 2012.

[OW12] OWL 2 Overview http://www.w3.org/TR/owl2-overview/

[Pe06] Peterson, E.: *Beneath the Metadata – Some Philosophical Problems with Folksonomy*, in: D-Lib Magazine, 12(11), http://www.dlib.org/dlib/ november06/peterson/11peterson.html, 2006.

[Pe08] Haase, P., Lewen, H., Studer, R, et al.: The NeOn Ontology Engineering Toolkit in: WWW 2008 Developers Track 2008.

[Pe09] Pérez, J., Arenas, M., Gutierrez, C.: *Semantics and complexity of SPARQL*. in: ACM Trans. Database Syst. 34, 3, Article 16, 2009.

[Pr10] Prateek, J. et. al.: Ontology alignment for linked open data. in: Proceedings of the 9th international semantic web conference on The semantic web (ISWC10), Volume Part I, pp. 402-417.

[Pe97] Pepijn, R. S. V., Dean, M. J., Bench-capon, T. J. M. , Shave, M.: *An analysis of ontological mismatches: Heterogeneity versus interoperability*. in: AAAI, Spring Symposium on Ontological Engineering, Stanford, USA, 1997.

[Pe06] Peterson, E.: *Beneath the Metadata – Some Philosophical Problems with Folksonomy*. in: D-Lib Magazine, 12(11), http://www.dlib.org/dlib/ november06/peterson/11peterson.html, 2006.

[Ra04] Rasmussen, E.: *Information Retrieval Challenges for Digital Libraries*. In: Proceedings of the 7th International Conference on Asian Digital Libraries (ICADL'04), Shanghai-New York: Springer, 2005 pp. 93–103.

[Re47] Reichenbach, H.: *Elements of Symbolic Logic*, Dover Publications, Inc. New York, ISBN 0-486-24004-5, 1947.

[Ro06] Roddick, J. F., Vries, D.: Reduce, Reuse, Recycle : *Practical Approaches to Schema Integration, Evolution and Versioning*. In: Advances in Conceptual Modeling - Theory and Practice, Lecture Notes in Computer Science Volume 4231, 2006, pp 209-216

[Ro12] Brennan, R., Walshe, B., O'Sullivan, D.: *Managed Semantic Interoperability for Federations*. In: Journal of Network and Systems Management, Submitted to Journal JNSM.

[Ru06] Ruxton, GD.,: The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test, Behavioural Ecology, 2006

[Sa83] Salton, G., McGill, M. J.: Introduction to Modern Information Retrieval, New York: McGraw-Hill, 1983.

[Sa09] Sa09 http://www.las.org.sg/sjlim/SJLIM20094Sampath.pdf.

[Sc10] Scott, J., Beth P.: *Trading Consistency for Scalability in Scientific Metadata*. In: Proceeding ESCIENCE '10 Proceedings of the 2010 IEEE Sixth International Conference on e-Science Pages, 2010, pp. 292-299.

[Sh04] Shvaiko, P., Euzenat J.: *A Survey of Schema-based Matching Approaches*. in: Journal of Data Semantics, Long version in DIT Technical Report DIT-04-87, 2004.

[Sh05] Shreeves, S. L. et al.: Is "Quality" Metadata "Shareable" Metadata? The Implications of Local Metadata Practices for Federated Collections. Proceedings of Twelfth National Conference of the Association of College and Research Libraries. Chicago, IL: Association of College and Research Libraries., 2005.

[Sk09] *SKOS Simple Knowledge Organization System*, Reference, W3C Recommendation 18 August http://www.w3.org/TR/skos-reference/, 2009.

[St09] Staab, S., Studer R. (eds): *Handbook on Ontologies*. Springer Verlag, Heidelberg. 2004, 2nd edition, 2009.

[Su07] O'Sullivan, D., Wade, V., Lewis, D.: *Understanding as We Roam*. In: IEEE Internet Computing Journal, Volume 11, Issue 2, 2007, pp. 26 – 33.

[Sh08] Shvaiko,P. Euzenat, J.: *Ten challenges for ontology matching*. in Proc. 7th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE), 2008, pp. 1163–1181.

[Sh12] Shvaiko, P., Euzenat, J.: *Ontology matching: state of the art and future challenges*. in: Journal of IEEE Transactions on Knowledge and Data Engineering, 2012 in press.

[St07] Stvilia, B., Gasser, L., Twidale, M. B., Smith, L. C.:A framework for information quality assessment. Journal of the American Society for Information Science and Technology, 58, , 2007, pp 1720-1733.

[Ta09] Taylor, A. G., Joudrey, D.N: *The organization of information*. third ed. Englewood: Libraries Unlimited, 2009.

[Te11] Tejo-Alonso, C. et al: *Metadata for Web Ontologies and Rules: Current Practices and Perspectives*. in Computer and Information Science Volume 240, 2011, pp 56-67

[Th09] Thomas, H., O'Sullivan, D., Brennan, R.: *Ontology Mapping Representations: a Pragmatic Evaluation*. In: 21st International Conference on Software Engineering and Knowledge Engineering. SEKE 2009, 1 - 3 July 3, Boston, 2009, pp. 228 - 232.

[Th09a] Thomas, H., O'Sullivan, D., Brennan, R.: *Evaluation of Ontology Mapping Representations*. In: Workshop on Matching and Meaning, Part of the as part of the AISB 2009 Convention, April 9th 2009. Edinburgh, Scotland 2009.

[Th09c] Thomas, Hendrik: *Ontology Mapping Management*. Poster for European Summer School on Ontological Engineering and the Semantic Web 2009, Cercedilla, Spanien, 5th - 12th June 2009.

[Th11] Thomas, H., Brennan, R. O'Sullivan, D.: *MooM - a Prototype Framework for Management of Ontology Mappings*. In: Proceedings of the 25th IEEE International Conference on Advanced Information Networking and Applications, Singapore, 22-25 March, 2011, IEEE, pp 548 – 555.

[Th12] Thomas B.: *Libraries, languages of description, and linked data: a Dublin Core perspective*. In: Library Hi Tech, Vol. 30 Iss: 1, 2012, pp.116 – 133

[Th12] Thomas, H., Brennan, R., O'Sullivan, D.: *Using the OM2R meta-data model for ontology mapping reuse for the ontology alignment challenge - a case study*. In proceedings of the 7th International Workshop on Ontology Matching (OM-2012) collocated with the 11th International Semantic Web Conference (ISWC-2012) Vol 946, http://ceur-ws.org/Vol-946/ Boston, MA, USA, November 11, 2012.

[Te11] Tejo-Alonso, C. et al: *Metadata for Web Ontologies and Rules: Current Practices and Perspectives*, in Computer and Information Science Volume 240, 2011, pp 56-67

[Um10] Umbrich, J. et al.: *Towards Dataset Dynamics: Change Frequency of Linked Open Data Sources.*/ In Proceedings of the Linked Data on the Web Workshop (LDOW2010), Raleigh, North Carolina, USA, April 27, 2010.

[Wa12] Walshe, B, Brennan, R., O'Sullivan, D.: *Correspondence Pattern Attribute Selection for Consumption of Federated Data Sources*. In: IFIP/IEEE DANMS 2012, Hawii, 16 April 2012.

[Wa12] Wallace, C. Golbreich, C. (eds): *OWL2 Web Ontology Language New Features and Rationale* (Second Edition) W3C Recommendation 11 December 2012 http://www.w3.org/TR/2012/REC-owl2-new-features-20121211/#Simple_metamodeling_capabilities, 2012.

[We07] Hu, W., Qu., Y. : *Discovering simple mappings between relational database schemas and ontologies*. In: International Semantic Web Conference. 225-238, 2007.

[We10] Weidman, S., Arrison, T.: *Steps Toward Large-Scale Data Integration in the Sciences*. ISBN: 0-309-15443-X, 2010.

[We10a] Weiers, R. M.: Introduction to Business Statistics, ISBN-10: 0538452196, 2010.

[We12] Westbrook, R. N.: Metadata Clean Sweep: A Digital Library Audit Project.in: D-Lib Magazine, May/June 2012 Volume 18, Number 5/6, http://dlib.org/dlib/may12/westbrook/05westbrook.html 2012.

[Xi08] Xiaodong Z., Zhiqiu Z., Guohua S.: Description Logic Based Consistency Checking upon Data Mining Metadata. In: Rough Sets and Knowledge Technology, Lecture Notes in Computer Science Volume 5009, 2008, pp 475-482.

[Ya11] Ya-Ning, C. et al.: Metrics for Metadata Quality Assurance and Their Implications for Digital Libraries. In: Digital Libraries: For Cultural Heritage, Knowledge Dissemination, and Future Creation Lecture Notes in Computer Science Volume 7008, 2011, pp 138-147.

[Za12] Zaveri A, Rula A. Maurino A. et al: *Quality Assessment Methodologies for Linked Open Data.* Under revision at IJWS Internation Semantic Web journal http://www.semantic-web-journal.net/system/files/swj414.pdf 2012.

DVD Structure

The following section outlines the structure and content of the DVD attached to this thesis.

- Folder *A Experiment 1 Automated retrieval experiment* offers
    - the raw data of the experiment results,
    - setup related information like the used sample ontologies and
    - the source code of the tool used for mapping retrieval.
- *Folder B Experiment 2 Relevance ranking of the OM2R* contains
    - a copy of the experiment interface,
    - the raw experiment results and
    - the approved application for the ethical commission
- *Folder C Experiment 3 Evaluation of the OM2R for consistency support* provides
    - a copy of the experiment interface,
    - the raw experiment results,
    - the approved application for the ethical commission and
    - details where the experiment and the OM2R editor can be accessed online.
- *Folder D OM2R Model* offers
    - the OM2R model as an OWL DL ontology expressed in the RDF/XML notation and
    - the OM2R project files to view the model and an automated generated interface in Protégé 3.
- *Folder E Thesis* offers this thesis as a Word and PDF document
- *Folder F OAEI Case-Study* contains a copy of the homepage which documents the benchmark alignment challenge and all offered meta-data.

Appendices

The appendices present support information for this thesis.

- Appendix A offers an overview of the meta-data fields which are defined for the common Ontology Alignment format.
- Appendix B discusses in detail how the OM2R can be used to document OAEI challenges with a focus on the identification and matching phase of the lifecycle.
- Appendix C shows the questionnaire form that was used to identify common mapping discover tasks.
- Appendix D provides the result of the questionnaire on mapping discover tasks which are used in section 3.4.4 to deduce design requirements for the OM2R.
- Appendix E lists the OM2R ontology version 1 that was used for the first experiment on mapping discovery, see section 5.4.4.
- Appendix F includes supporting information how the OM2R can be applied to support federation
- Appendix G discusses the impact of the design decision to represent the OM2R model as an OWL DL ontology and not an OWL 2 ontology
- Appendix H contains a copy of the OM2RL model expressed as a OWL DL ontology

Appendix A. Ontology Alignment Format Meta-Data

The specification of the Ontology Alignment Format lists the following standard meta-data field:

http://knowledgeweb.semanticweb.org/heterogeneity/alignment#xml

BOOLEAN telling if the file can be read as XML

http://knowledgeweb.semanticweb.org/heterogeneity/alignment#level

STRING

http://knowledgeweb.semanticweb.org/heterogeneity/alignment#type

STRING identifying the type of alignment (1:1, 1:*, ?:+...)

http://knowledgeweb.semanticweb.org/heterogeneity/alignment#uri1

URI the uri of the first ontology

http://knowledgeweb.semanticweb.org/heterogeneity/alignment#uri2

URI the uri of the second ontology

http://knowledgeweb.semanticweb.org/heterogeneity/alignment#onto1

URL the location of the first ontology

http://knowledgeweb.semanticweb.org/heterogeneity/alignment#onto2

URL the location of the second ontology

http://knowledgeweb.semanticweb.org/heterogeneity/alignment#semantics

STRING the intended interpretation of a correspondence

The following meta-data extensions for listed in the specification:

http://knowledgeweb.semanticweb.org/heterogeneity/alignment#id

URI identifying the alignment.

http://knowledgeweb.semanticweb.org/heterogeneity/alignment#method

CLASSNAME of the generating matching method (or operation).

http://knowledgeweb.semanticweb.org/heterogeneity/alignment#methodVersion

STRING identifying the method version.

http://knowledgeweb.semanticweb.org/heterogeneity/alignment#relation

URI identifying the classname implementing the relation structure.

http://knowledgeweb.semanticweb.org/heterogeneity/alignment#confidence

URI identifying the classname implementing the confidence structure.

http://knowledgeweb.semanticweb.org/heterogeneity/alignment#derivedFrom

URI the alignment from which this one is issued, if applicable

http://knowledgeweb.semanticweb.org/heterogeneity/alignment#parameters

STRING the parameters used with the generating method

http://knowledgeweb.semanticweb.org/heterogeneity/alignment#certificate

STRING a certificate from an issuing source

http://knowledgeweb.semanticweb.org/heterogeneity/alignment#time

DURATION of the matching process.

http://knowledgeweb.semanticweb.org/heterogeneity/alignment#limitations

STRING the validity range of the correspondence

http://knowledgeweb.semanticweb.org/heterogeneity/alignment#properties

STRING the properties satisfied by an alignment or correspondences

http://knowledgeweb.semanticweb.org/heterogeneity/alignment#pretty

STRING a short descriptive name of the alignment.

http://knowledgeweb.semanticweb.org/heterogeneity/alignment#label1

STRING a short descriptive name for the first aligned entity (in correspondences).

http://knowledgeweb.semanticweb.org/heterogeneity/alignment#label2

STRING a short descriptive name for the second aligned entity (in correspondences).

http://knowledgeweb.semanticweb.org/heterogeneity/alignment#provenance

STRING identify the tools and versions who created the alignments (concatenated by)

Appendix B. Application of OM2R for the OAEI

In this appendix it is discussed how the OM2R meta-data fields can be applied beneficially for the OAEI. For this purpose the individual lifecycle phase are analysed.

1. Meta-data fields for the Characterisation phase

The first two use-cases (1+2) focus on the identification of the ontologies which are addressed in the mapping. Two stakeholders have to be considered, i.e. human users and automated processes such (e.g. online match system like www.sameAs.org[134]). Names are the most common identifier for users [Pe06] but names can be a difficult to handle as entities can have more than one name. However, in some cases entities have no commonly accepted name and can only be identified via a textual definition.[135] The OM2R addresses these requirements with the meta-data field *om2r:has_Human_Readable_Name* and *om2r:has_description*. Please note an entity can have more than one names or description and one name can be defined as a default name. In contrast names are not ambiguous enough for automated processing and therefore the OM2R provides the fields *om2r:has_Unquie_Identifier* which allows the specification of a unique URI. These three fields are obvious quite generic and not limited to description of ontologies. Thus in the OM2R model these fields can be used to describe the resulting mapping but also any other model element. Thus these three fields provide the necessary information to understand the intended meaning of the model element and are considered as the embedded OM2R documentation.

Use-Case 2 defines a more specific scenario where a user needs to find mappings between specific elements contains in an ontology. Therefore a reference to an ontology as a whole is not detailed enough. To resolve this task it is necessary to parse the ontologies. For this purpose OM2R offers the field *om2r:has_location* which provides an URL where the file can be downloaded. The relevance of this information is

[134] The web page sameAs.org offers a service to help a user find co-references between different data set based on the owl statement owl:sameAs.

[135] For example aqua planning describes an effect if a car loses traction if a street is wet. This is a common name but there is no name so far for the same situation if the street is wet due to spilt milk and not water.

demonstrated by the fact that all other mapping representation formats provide similar information, e.g. INRIA:

The state of art study in chapter 2 has shown that only sparse information regarding the author of an ontology is provided. In the light of a possible reuse of a mapping it is important to be able to contact the author to negotiated reuse permissions or to manage change of interlinked mappings [Be12, Br13c]. A simply name to identify the author as proposed by the INRA format is not unambiguous enough [Eh07]. The OAEI addresses this need by providing the following field *om2r:mapping_creator*. Please note, existing ontology templates such as FOAF for contact details can be used to help identify the creator more accurately if such information is available e.g. *om2r:dcFirstName* Hendrik.

Another aspect which can help to identify a specific version of an ontology is the *om2r:creation_date* which defines the explicit time and date of the creation of the ontology in the OM2R. As various date formats exist, an explicit and unambiguous representation is needed. Thus internally the creation date will be represented as follows in the OM2R: *om2r:creation_date :hasYear: 2010, om2r:creation_date:hasMonth: 5, om2r:creation_date:hasDay: 4.*

In the next step of the lifecycle the ontologies are characterised to assess the complexity and challenges for the upcoming matching. Vital components are language related aspects. Use-case 3 and 4 cover the case where only ontologies expressed in a certain formal language or notation are needed. This is not limited to the actual targeted ontologies but also vital for the mapping representation (use-case 5) as the used mapping representation format can determine what equivalence relations are supported, e.g. simple equivalence mappings or more complex mappings (narrow, broader etc.). Some of the mapping representation formats provide details about the language of the addressed ontologies (e.g. INRIA) but usually only as a textual reference. This is often not sufficient as one needs to know the specific formal language used, e.g. reasoning can only be applied to OWL DL not OWL Lite, thus stating the formal language as OWL is too broad.

The OM2R addressees this need with the field *om2r:formal_language.* As there are a number of such languages for ontologies, this field specifies the language om2r:hasFormalLanguage language:http://www.w3.org/2002/07/owl. Beside the ontology language, the specific exchange notation used to represent the address ontology can be specified which is essential for tool usage and exchange, e.g. om2r:TargetOntology om2r:hasNotation RDF/XML.

2. Meta-data fields for the Matching phase

Matching is the core process step in the ontology mapping lifecycle for the OAEI and a focus of most of the current tools [Sh12]. Matching is a complex process [Su12] and details about the applied methods are required to understand how mapping candidates were identified. This is vital to decide if the resulting mappings can be applied in a different context. The OM2R offers the following fields to document details about the matching process (use-case 6 to 12) from a top-down approach. It starts with the field *om2r:matching_method* which defines what general method used to identify the mapping candidates, e.g. either automated process, manual selection or mixed approach. Depending on this choice the OM2R provides more and more details further down the model, e.g. matching tool leads to matching algorithm. This allows a user to focus on a needed level of detail depending on the specific retrieval or mapping reuse-case. Figure 48 shows the provided meta-data fields

Automated selection of mapping candidates

Om2r:Matching_Tool: Specified the tool which was used to generate the alignment, .e.g. hasMatchingTool

Om2r:Matching_Algorithm: If an automated selection was applied, this section provides a descriptive, human-readable label to identify the matching algorithm used. For example: matching :basedOn: Levenshtein distance, Levenshtein distance :isDefinedIn: http://en.wikipedia.org/wiki/Levenshtein_distance

Om2r:Matching_Algorithm Implementation: A descriptive, human-readable label to identify the specific implementation of the algorithm. Could be a URL or a specific JAVA class name such as org.jena.stringComparsion. Also helpful is to provide a URL to download the source code. For example: Algorithm :encodedIn: Java, Algorithm :hasJavaClass: org.jena.stringComparsion, Algorithm :hasSource: http://www.freecode.org/123.zip

Om2r:Applied_Threshold: Defines the specific value of the similarity measure which needs to be passed in order to justify a matching pair based on the assumptions of the individual algorithm,e.g. om2r:has_Applied_Threshold

Om2r:Matching_Scope: Defines the scope or area the matching is applied. In particular if all elements are matched to each other or only a particular subset, e.g. om2r:hasScope – complete or partial

Om2r:Element_Matching_is_based_on: Defines the elements which are analyzed by the algorithm to identify the matching pairs, e.g. RDFSLabelForClass

Figure 51 OM2R Fields to describe the manual and automated matching process

These fields focus on the actual matching selection but more information is needed in regards to the context and objective in which the matching was applied. This concludes in specific requirements for a given matching which can be documented in the field *om2r:has_Matching_Requirements* (Use-case 14). Again this is vital to judge the reusability from an application rather than an implementation perspective.

3. Meta-data fields for the Mapping Phase

This phase focusses on the creation of mappings in terms of confirming the mapping candidates which were identified in the previous matching process. For this purpose the OM2R provides the field om2r:has_Mapping_Requirements which documents details about the rules or conditions which were applied. This is particular relevant for the planning of the mapping approach which can be described in the field om2r:mapping_process. The identified committed and approved mappings can then be rendered into different mapping formats in order to enable processing and sharing. Thus the field om2r:notation and om2r:formal_mapping_language are provided. Similar to the ontology identification phase, the author and the creation time of the mappings can be described, i.e. om2r:creation_date and om2r:editor.

4. Meta-data field for the Management Phase

Once the mapping has been created they are used, shared and thus require management, e.g. see section 9.3 where an example is shown how a reference mappings in the evaluation challenge of the OAEI has changed over the years. This creates the need to find mappings in a particular version (use-case 14). To address this, each mapping is linked to a source element in the model that contains the link to access the file and a last change date.

If a mapping reuse-case is considered it is important to know what other application are using a mapping actively. This is vital to handle the impact of cascading changes to mapping sources. The OM2R offers for this the field *om2r:application_using_mapping*. This is complemented by the field *om2r:change_notification* which provides details about the process (if any) used to notify users on changes in the mapping file. For example, a RSS feed could be used. With such different applications it is likely that different requirements come into play and the OM2R model is likely to evolve over time to address them. Therefore each OM2R field is linked to a version to clearly state what version of the OM2R model it belongs too.

Appendix C. Questionnaire on Mapping Discovery

Please find below a copy of the questionnaire send out as part of experiment 1 (see chapter 5 used to identify common mapping discovery task:

Questionnaire on Ontology Mapping Discovery Tasks

Research question: What are common ontology mapping discovery tasks?

Based on an analysis of the ontology mapping lifecycle the following mapping tasks have been identified. As an expert on ontology matching and mapping we would such as to ask you to:

1. **Rate the list of mapping discovery tasks below for likely <u>frequency</u> and <u>relevancy</u> to mapping applications.**
2. **Please tell us about any other common mapping discovery tasks.**

A.) <u>Tasks related to ontologies and mappings</u>

Find mappings between specific source and target ontologies or ontology elements

Motivation: to support information exchange between two specific knowledge models

Example: [source: `tcd_federaton.owl`, target: `usd_federation.owl`]

[ ]	Highly relevant	[ ]	Medium relevant	[ ]	Irrelevant
[ ]	Frequent[ ]	Occasional	[ ]	Rare	

`Comments:`

Find mappings for a particular ontology or an ontology element

Motivation: identify interoperable ontologies which could be reused in order to reduce the modeling efforts

Example: search for available mappings for `fame_security.owl` or for the class `<http://fame.ie/federation#security_level>`

[ ]	Highly relevant	[ ]	Medium relevant	[ ]	Irrelevant
[ ]	Frequent[ ]	Occasional	[ ]	Rare	

`Comments:`

Find mappings between a specific source ontology to any target ontology with specified characteristics, e.g. ontology language

Motivation: the intended reuse scenario dictates what ontologies can be reused, e.g. if reasoning is required only ontologies expressed in OWL DL or OWL Lite can be reused

Example: search mappings for `fame_security.owl` with any OWL DL based ontologies

[ ]	Highly relevant	[ ]	Medium relevant	[ ]	Irrelevant
[ ]	Frequent[ ]	Occasional	[ ]	Rare	

`Comments:`

Find mappings expressed in <u>particular mapping format</u>

Motivation: depending the application only a specific mapping format maybe supported

Example: find mappings expressed in SWRL or the RDF INRIA format

[ ] Highly relevant [ ] Medium relevant [ ] Irrelevant
[ ] Frequent[ ] Occasional [ ] Rare

Comments:

Find mappings for specific correspondence types

Motivation: depending on reuse scenario simple equivalence mappings or more complex mappings (narrow, broader etc.) may be required

Example: find broader mapping between `tcd_han.owl` and `usd_han.owl`

[ ] Highly relevant [ ] Medium relevant [ ] Irrelevant
[ ] Frequent[ ] Occasional [ ] Rare

Comments:

B.) <u>Tasks related to the mapping generation process</u>

Find mappings created either automated, manually or in a combination

Motivation: fundamental differences automated and manual creation, thus the reuse scenario defines what is appropriate

Example: find only manually created mappings for a given ontology

[ ] Highly relevant [ ] Medium relevant [ ] Irrelevant
[ ] Frequent[ ] Occasional [ ] Rare

Comments:

Find mappings created by a specific matching type

Motivation: many different matching approach are possible, e.g. Shvaiko and Euzenat 2004 provided a general classification and this could be used as a search criteria

Example: search mappings based on structural or terminological matching

[ ] Highly relevant [ ] Medium relevant [ ] Irrelevant
[ ] Frequent[ ] Occasional [ ] Rare

Comments:

Find mappings created by a specific matching algorithm implementation

Motivation: narrow search down to a specific matching implementation

Example: find mapping created by Ontology Alignment API and in particular the classStructureAlignment method

[ ] Highly relevant [ ] Medium relevant [ ] Irrelevant
[ ] Frequent[ ] Occasional [ ] Rare

Comments:

Find mappings created by a given matching algorithm configuration

Motivation: depending on the applied parameters or thresholds an algorithm can produce quite different matching results

Example: find mappings created with a given algorithm with a given configuration

[ ] Highly relevant [ ] Medium relevant [ ] Irrelevant
[ ] Frequent[ ] Occasional [ ] Rare

Comments:

Find automated created mappings based on matches with a high confidence level

Motivation: most automated matching applications provide a confidence level which can be used to filter the results

Example: find mapping based on matched with an confidence level >= 80 %

[ ] Highly relevant [ ] Medium relevant [ ] Irrelevant
[ ] Frequent[ ] Occasional [ ] Rare

Comments:

Find manually created mappings depending on the involved users

Motivation: depending on skills and backgrounds different involved users can create different mapping, thus may not be suitable for reuse scenarios (e.g. product engineers vs. customers)

example: search manual mapping created by TCD students

[ ]	Highly relevant	[ ]	Medium relevant	[ ]	Irrelevant
[ ]	Frequent[ ]	Occasional	[ ]	Rare	

Comments:

C.) <u>Tasks based on management related meta-data</u>

Find mappings created by a <u>particular author</u>

Motivation: trust a specific author or what to manage your own mappings

Example: find mapping create by Declan O'Sullivan

[ ]	Highly relevant	[ ]	Medium relevant	[ ]	Irrelevant
[ ]	Frequent[ ]	Occasional	[ ]	Rare	

Comments:

Find mappings for a particular <u>ontology version</u>

Motivation: ontologies are evolving and therefore mapping must constantly be checked if they are still valid. Thus identifying an ontology-based an URI is not specific enough, e.g. creation or update time are needed

Example: find mapping for the newest version of an ontology

[ ]	Highly relevant	[ ]	Medium relevant	[ ]	Irrelevant
[ ]	Frequent[ ]	Occasional	[ ]	Rare	

Comments:

Find a specific version of a mapping

Motivation: mapping are also evolving and need to be managed

Examples: find newest mapping between two given ontologies

[ ]	Highly relevant	[ ]	Medium relevant	[ ]	Irrelevant
[ ]	Frequent[ ]	Occasional	[ ]	Rare	

Comments:

Find mappings created for a particular context

Motivation: the reason why a particular mapping is required has a major impact on the mapping creation process. Thus a reuse of this mapping may only be appropriate in a similar or compatible context.

Example: find mappings relevant for the FAME project Scenario E

[ ]	Highly relevant	[ ]	Medium relevant	[ ]	Irrelevant
[ ]	Frequent[ ]	Occasional	[ ]	Rare	

Comments:

D.) <u>Please add further relevant / frequent mapping discovery tasks</u>

- I think "context" is a catch all for any remaining tasks but…
- Mapping discovery based on mappings created for the day and month of the year.
- Mapping discovery based on previous or similar events.
-

The results of this survey will be used for the PhD project of Hendrik Thomas only and will be made available for the participant on request. This work is partially funded through the FAME project under the SFI award NO. 08/SRC/I1408.

Appendix D. Results of the questionnaire on mapping discovery tasks

Figure 52 presents the results of the questionnaire used to verify the proposed common mapping task. Please see section 3.4.3 for more details. In this questionnaire the participants were asked to rate the collection of mapping discovery tasks for their likely frequency of occurrence (possible answers: 2 = 100 % frequent, 1 = 50 % - occasional, 0 = 0 % - rare) and relevance to mapping applications (possible answers: 2 = 100 % - highly relevant, 1 = 50 %- medium relevant, 0 = 0 % - irrelevant). The results in the graph are presented as normalized value in percentage to allow an easier comparison between both scales. The blue bar indicates the mean relevance rating and the red bar the mean for likely occurrence frequency.

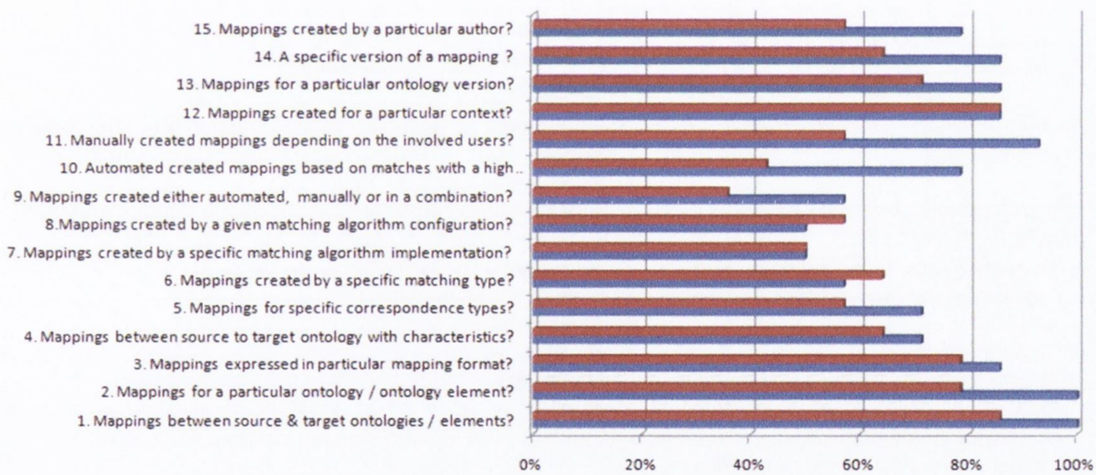

Figure 52 Results of the Questionnaire on Ontology Mapping Discovery

The participants suggested the following other relevant mapping discovery task:

- Mapping discovery based on mappings created for the day and month of the year.
- Mapping discovery created by a specific author
- Mapping discovery based on previous or similar events.
- Find mappings based on popularity – number of times they have been reused.
- Improve mapping findings by grouping similar mappings together.

Appendix E. Example of the OM2R ontology used in the first experiment


```xml
<?xml version='1.0'?>
        <rdf:RDF xmlns='http://fame.org/moom#'
                        xml:base='http://www.animals.fake/animals#'
                        xmlns:rdfs='http://www.w3.org/2000/01/rdf-schema#'
                        xmlns:dc= 'http://purl.org/dc/elements/1.1/'
    xmlns:rdf='http://www.w3.org/1999/02/22-rdf-syntax-ns#'>

  <rdfs:Class rdf:about='http://fame.org/moom#metadata_set' />

  <rdf:Description rdf:about='http://fame.org/moom#mapping_metadata_1'>
   <source_ontology>http://kdeg-vm-7.cs.tcd.ie/experiment1/tcd_fame_v1.owl</source_ontology>
   <source_ontology_version>1</source_ontology_version>
   <target_ontology>http://kdeg-vm-
7.cs.tcd.ie/experiment1/telefonica_fame_owlfull.owl</target_ontology>
   <target_ontology_version>1</target_ontology_version>
   <target_ontology_version>1</target_ontology_version>
   <source_ontology_language rdf:resource='http://fame.org/moom#owl_dl'/>
   <target_ontology_language rdf:resource='http://fame.org/moom#owl_full'/>
   <mapping>http://kdeg-vm-
7.cs.tcd.ie/moomdrupal/experiment1/tcd_v1_telefonfic_editDist.rdf</mapping>

<source_ontology_uri>http://phaedrus.cs.tcd.ie/moomdrupal/fame/psi/telefonica_fame_owlfull.owl</sourc
e_ontology_uri>

<target_ontology_uri>http://phaedrus.cs.tcd.ie/moomdrupal/fame/psi/telefonica_fame_owlfull.owl</target_
ontology_uri>
   <mapping_language rdf:resource='http://fame.org/moom#INRIARDF'/>
   <matching_gen_type rdf:resource='http://fame.org/moom#automated'/>
   <matching_type rdf:resource='http://fame.org/moom#terminological'/>
   <mapping_author_fn>Declan</mapping_author_fn>
   <mapping_author_ln>O'Sullivan</mapping_author_ln>
   <matching_type>terminological</matching_type>
   <matching_impl>fr.inrialpes.exmo.align.impl.method.EditDistNameAlignment</matching_impl>
   <matching_user>Not one</matching_user>
   <matching_conf>strlevel=1</matching_conf>
   <mapping_context>FAME</mapping_context>
   <mapping_creation_time>2009-06-01</mapping_creation_time>
   <mapping_version>1</mapping_version>
   <rdf:type rdf:resource='http://www.w3.org/2000/01/rdf-schema#Class'/>
   <rdfs:subClassOf rdf:resource='http://fame.org/moom#metadata_set'/>
```

Appendix F. Application of OM2R for Federation

In this appendix it is discuss how the OM2R meta-data fields can be applied beneficially for the federation use-case. For this purpose the two key tasks of the federation process presented in section 9.1 analysed.

1. OM2R for Federations - Use-Case Mapping Discovery

A key task for the FRM phase "Federation context-based mapping candidate identification" is to discovery relevant mappings. OM2R supports this activity with the fields: (Target and Source) Ontology Name and (Target and Source) Ontology Description field. These fields are designed primarily for human data consumption and makes management of mapping easier if done manually. However, these descriptors can also be used for an automated discovery by using a string comparison which is common for matching algorithms. The addressed ontologies (target and source) can meta-data embedded in their source code, e.g. <dc:description>, <rdfs:label>. These can be used to populate the above mentioned OM2R fields automatically.

The discovery process in the FRM is automated and an unambiguous identifier for the ontologies is essential to identify what ontology are subject of a mapping. To avoid any miss interpretations the OM2R provides an explicit field Ontology Identifier where such a unique identifier can be stored. The OM2R field Ontology Location and Mapping Source allows users and automated system to retrieve the specific download location of the addressed source codes. The identifier can be harvested from the alignment format as well as the source location.

Important again for the FDM are the language related aspects of the ontology files due to compatibility issues with the processing tools. This is particular vital for the federation contact as various tool with different language processing capabilities share resource and services in the federation. OM2R offers the fields Ontology Formal Language and Notation for this purpose and to make such information explicit. Similar information is also relevant for the actual matching process.

2. OM2R for Federations - Use-Case Mapping Evaluation

As part of the federation context-based mapping candidate identification process step, the FDM needs to decide if an identified mapping is suitable for a given set of capabilities

A key focus is naturally the actual correspondences contained in the matching files. The FDM uses a semi-automated process but depending of the amount of contain correspondences a manual review is not possible in many cases. This brings the actual process which was used to create the mapping into focus. The decision made there can provide an indication if the mapping is suitable form a qualitative and quantitative perspective, e.g. manual matching by domain export vs. automated matching based on strings alignment format provides a corresponding meta-data field such as <method> and <parameters>136 which provide ne detail and are very high level.

In addition, matching creation is usually conducted in more than one step. The process can be documented in the OM2R by explicit modelling the individual steps which can then be linked to further details, e.g. for step1 the matching type automated was used but in step 2 a manual approach. Mapping particular process can be documented in a similar approach such as in matchings. Furthermore, the OM2R provides a field Matching Requirements for details of the specific requirements which needed to be fulfilled to apply the matching. These can be relevant for an evaluation depending on the reuse scenario.

[136] Please see the following URL for more details http://alignapi.gforge.inria.fr/labels.html

Appendix G. Impact of expressing the OM2R model in OWL DL vs. OWL 2

This appendix discusses the implication of the design decision to represent the OM2R model in an OWL DL ontology rather an OWL 2 ontology. In this thesis OM2R is expressed an OWL DL ontology. A key design decision of the model was that every element in the OM2R model should have some descriptive meta-data attached. In OM2R a simplex modelling style for meta-data documentation is defined with the following data properties are: om2r:has_description, om2r:has_unquie_identifier, om2r:has_human_Readable_Name. The default name is indicated in the string with a tailing * in the string value. This is considered a work-a-round but sufficient for the OM2R editors and common ontology editors such as Protégé. However, the implementation choice of using OWL-DL has the limitation that these data properties can only be filled in for instances but not classes. However, the requirement R2 demands that all elements including class have a meta-data attached. In the OM2R a work-a-round in applied where for every class a specific instance is created. The name of this instance matches the one given to the class but with a lower capital and a leading "c_" character. In OWL Version 2 this issue is handled with a concept call punning[137] which refers to the fact that OWL2 DL relaxes the separation between classes and instances, e.g., "Eagle, to be used for both a class, the class of all Eagles, and an individual, the individual representing the species Eagle belonging to the (meta) class of all plant and animal species". This means in OWL Version 2 it would be possible to assign the meta-data data properties to a class as well. In OWL DL version 1 this is not possible.

To sum up the impact of choosing OWL 2 instead of OWL DL is low and limited to the way meta-data is expressed for classes. In OWL 2 the data properties can be linked directly to the class in OWL DL a proxy instance is used. No further implications can be identified as all class and relation used in the OWL DL version of the OM2R model can be implement the same way in OWL 2.

---

[137]http://www.w3.org/TR/owl2-new-features/#Simple_metamodeling_capabilities
http://ontogenesis.knowledgeblog.org/1076

Appendix H. OWL DL version of the OM2R model

In this appendix the final version of the OM2R model is expressed as in the OWL 2 Manchester syntax[138] as it provides the most compact form to show the complete model.

```
Prefix: : <http://www.modellmapping.org/om2rdemo.owl#>
Prefix: p1: <http://www.owl-ontologies.com/assert.owl#>
Prefix: j.0: <http://www.modellmapping.org/om2rdemo.owlm2r:>
Prefix: owl: <http://www.w3.org/2002/07/owl#>
Prefix: rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
Prefix: xml: <http://www.w3.org/XML/1998/namespace>
Prefix: xsd: <http://www.w3.org/2001/XMLSchema#>
Prefix: xsp: <http://www.owl-ontologies.com/2005/08/07/xsp.owl#>
Prefix: rdfs: <http://www.w3.org/2000/01/rdf-schema#>
Prefix: swrl: <http://www.w3.org/2003/11/swrl#>
Prefix: swrlb: <http://www.w3.org/2003/11/swrlb#>
Prefix: protege: <http://protege.stanford.edu/plugins/owl/protege#>
Ontology: <http://www.modellmapping.org/om2r>
Annotations:
    owl:versionInfo "om2r version 3"^^xsd:string,
    rdfs:comment "This onology provides the meta-data model for documenting ontology mappings with a
       focus to support ontology mapping retrieval and reuse. To enable this support the model focues
       on the ontology mapping lifecycel which includes the following phaes
1 Ontology Identification Phase
2 Matching Phase
3 Mapping Phase
4 Mapping Management"@en,
    rdfs:label "OM2R Version 3"@en
AnnotationProperty: owl:versionInfo
AnnotationProperty: rdfs:isDefinedBy
AnnotationProperty: rdfs:label
AnnotationProperty: rdfs:comment
Datatype: rdf:PlainLiteral
Datatype: xsd:anyURI
Datatype: xsd:date
Datatype: xsd:int
Datatype: xsd:string
Datatype: xsd:dateTime
ObjectProperty: was_applied_to_algortyhm
Range:
Matching_Algorithm
ObjectProperty: has_matching_requirments
Domain:
Matching_Process_Steps
Range:
Matching_Requirments
ObjectProperty: hasAppliedThreshold
Domain:
Matching_Process_Steps
Range:
AppliedThreshold
ObjectProperty: has_mapping_requirments
Domain:
Mapping_Process_Steps
Range:
Mapping_Requirements
ObjectProperty: has_matching_scope
Domain:
Matching_Process_Steps
Range:
Matching_Scope
ObjectProperty: has_process_details
Domain:
Mapping_Process_Steps
Range:
Mapping_Process_Details
ObjectProperty: belongs_to_process_step
Domain:
Mapping_Process_Details
```

[138] Please see the following link for more information on this syntax http://www.w3.org/TR/owl2-manchester-syntax/

```
 or Matching_Process_Details
Range:
Mapping_Process_Steps
 or Matching_Process_Steps
ObjectProperty: compatible_to
Domain:
Formal_Language
 or Matching_Algorithm
 or Matching_Method
 or Matching_Tool
 or Notation
Range:
Element_the_matching_is_based_on
 or Formal_Language
 or Matching_Algorithm
 or Matching_Tool
 or Notation
ObjectProperty: has_mapping_process_step
Domain:
Mapping_Process
Range:
Mapping_Process_Steps
ObjectProperty: ElementTheAlgorymIsBasedOn
Domain:
Matching_Process_Steps
Range:
Element_the_matching_is_based_on
ObjectProperty: date_of_mapping_creation
Domain:
Mapping
 or Ontology
Range:
Creation_Date
ObjectProperty: hasApplicationThatUsesTheMapping
Domain:
Mapping
Range:
ApplicationThatUseTheTheMapping
ObjectProperty: is_Addressing
Domain:
Mapping
Range:
Ontology
ObjectProperty: belongs_to_matching_process
Domain:
Matching_Process_Steps
Range:
Matching_Process
ObjectProperty: has_matching_algorithm
Domain:
Matching_Process_Steps
Range:
Matching_Algorithm
ObjectProperty: has_process_step
Domain:
Matching_Process
Range:
Matching_Process_Steps
ObjectProperty: has_creator
Domain:
Mapping_Process_Steps
Range:
Mapping_Creator
ObjectProperty: has_source
Domain:
ApplicationThatUseTheTheMapping
 or ChangeNotification
 or Mapping
 or Matching_Algorithm
 or Matching_Tool
 or Ontology
Range:
Source
ObjectProperty: has_version
Domain:
Class_Representator
 or Matching_Tool
 or Model_Elements
 or OM2R_Model
 or Ontolooy_Mapping_Lifecycle_Entities
Range:
```

```
OM2R_model_version
ObjectProperty: is_Step
Domain:
Lifecycle_Process_Steps
 or Matching_Process_Steps
Range:
Steps
ObjectProperty: has_matching_method
Domain:
Matching_Process_Steps
Range:
Matching_Method
ObjectProperty: has_formal_language_mapping
Domain:
Mapping
Range:
Mapping_Formal_Language
ObjectProperty: has_formal_language
Domain:
Ontology
Range:
Formal_Language
ObjectProperty: belongs_to_step
Domain:
Model_Elements
Range:
Lifecycle_Process_Steps
ObjectProperty: has_notation
Domain:
Mapping
 or Ontology
Range:
Notation
ObjectProperty: hasChangeNotification
Domain:
Mapping
Range:
ChangeNotification
ObjectProperty: has_language
ObjectProperty: is_Target
Annotations:
rdfs:comment "defines if the ontology is source or target for the mapping"^^xsd:string
Domain:
Ontology
Range:
Mapping_Targets
ObjectProperty: has_matching_tool
Domain:
Matching_Process_Steps
Range:
Matching_Tool
ObjectProperty: has_domain
Domain:
Ontology
Range:
Domain
DataProperty: dcAddress
Domain:
Mapping_Creator
DataProperty: textValue
Domain:
Mapping_Requirements
 or Matching_Requirments
 or Process_Description
Range:
xsd:string
DataProperty: hasMonth
Domain:
Creation_Date
DataProperty: dcFirstName
Domain:
Mapping_Creator
DataProperty: hasYear
Domain:
Creation_Date
DataProperty: hasHumanReadableName
Domain:
Abstract
 or Class_Representator
 or Model_Elements
 or OM2R_Model
```

```
 or Ontolooy_Mapping_Lifecycle_Entities
Range:
xsd:string
DataProperty: hasUniqueIdentifier
Domain:
Abstract
 or Model_Elements
 or OM2R_Model
 or Ontolooy_Mapping_Lifecycle_Entities
DataProperty: hasDescription
Domain:
Abstract
 or Model_Elements
 or OM2R_Model
 or Ontolooy_Mapping_Lifecycle_Entities
DataProperty: hasTimeZone
Domain:
Creation_Date
DataProperty: dcLastName
Domain:
Mapping_Creator
DataProperty: has_editing_priority
Domain:
Class_Representator
 or Model_Elements
 or Ontolooy_Mapping_Lifecycle_Entities
Range:
xsd:int
DataProperty: lastCheckedDateAndTime
Domain:
Source
Range:
xsd:dateTime
DataProperty: dataValue
Domain:
Creation_Date
Range:
xsd:date
DataProperty: hasDay
Domain:
Creation_Date
DataProperty: urlValue
Domain:
Source
DataProperty: dcEmail
Domain:
Mapping_Creator
Class: Matching_Requirments
SubClassOf:
Matching_Process_Details
Class: Matching_Details
SubClassOf:
Model_Elements
Class: Matching_Tool
SubClassOf:
Matching_Process_Details
Class: Creation_Date
SubClassOf:
Descriptors
Class: ApplicationThatUseTheTheMapping
SubClassOf:
Management_Details
Class: Language
SubClassOf:
Model_Elements
Class: Notation
SubClassOf:
Language
Class: Mapping_Process_Steps
SubClassOf:
Mapping_Details
Class: Mapping_Creator
SubClassOf:
Mapping_Process_Details
Class: Descriptors
SubClassOf:
Model_Elements
Class: Mapping_Details
SubClassOf:
Model_Elements
Class: Matching_Process_Steps
```

```
SubClassOf:
Matching_Details
Class: Matching_Process
Annotations:
rdfs:comment "Acts as a proxy for the process used to identify matching pair betwen the target and the
     source ontology."^^xsd:string
SubClassOf:
Matching_Details
Class: ChangeNotification
SubClassOf:
Management_Details
Class: Abstract
Class: Matching_Method
SubClassOf:
Matching_Process_Details
Class: Mapping_Targets
SubClassOf:
Abstract
Class: Model_Elements
Class: Mapping_Formal_Language
SubClassOf:
Language
Class: Element_the_matching_is_based_on
SubClassOf:
Matching_Process_Details
Class: Domain
SubClassOf:
Descriptors
Class: Mapping_Process
SubClassOf:
Mapping_Details
Class: Matching_Process_Details
SubClassOf:
Matching_Details
Class: Steps
SubClassOf:
Abstract
Class: Matching_Algorithm
SubClassOf:
Matching_Process_Details
Class: Matching_Scope
SubClassOf:
Matching_Process_Details
Class: Ontolooy_Mapping_Lifecycle_Entities
Class: OM2R_model_version
SubClassOf:
Abstract
Class: Ontology
Annotations:
rdfs:isDefinedBy <http://www.modellmapping.org/om2rdemo.owl#Demo_Book_Ontology>
SubClassOf:
Ontolooy_Mapping_Lifecycle_Entities
Class: Management_Details
SubClassOf:
Model_Elements
Class: Lifecycle_Process_Steps
SubClassOf:
Abstract
Class: Mapping
SubClassOf:
Ontolooy_Mapping_Lifecycle_Entities
Class: Process_Description
SubClassOf:
Mapping_Process_Details
Class: OM2R_Model
Class: Mapping_Process_Details
SubClassOf:
Mapping_Details
Class: Source
SubClassOf:
Descriptors
Class: Mapping_Requirements
SubClassOf:
Mapping_Process_Details
Class: Class_Representator
SubClassOf:
Abstract
Class: AppliedThreshold
SubClassOf:
Matching_Process_Details
Class: Formal_Language
```

```
SubClassOf:
Language
Individual: c_hasChangeNotification
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasHumanReadableName  "relation c_hasChangeNotification "^^xsd:string,
 hasDescription  "links a mapping to particular method to communicate changes"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_hasChangeNotification"^^xsd:string
Individual: c_has_process_details
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_has_process_details"^^xsd:string,
 hasHumanReadableName  "relation c_has_process_details "^^xsd:string,
 hasDescription  "relation of a process to its process details"^^xsd:string
Individual: c_compatible_to
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasDescription      "Relations    between    logical    consistent    compatible    meta-data    field
       combinations"^^xsd:string,
 hasHumanReadableName  ""^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_compatible_to"^^xsd:string,
 has_editing_priority  "0"^^xsd:int
Individual: c_Mapping_Process_Details
Types:
Mapping_Process_Details,
Class_Representator
Facts:
 belongs_to_step  Mapping_Step,
 has_version  OM2R_version_3,
 hasDescription  "Mapping Process details"^^xsd:string,
 has_editing_priority  "2"^^xsd:int,
 hasHumanReadableName  "Mapping Process Details"^^xsd:string,
 hasUniqueIdentifier  ""^^xsd:string
Individual: c_hasDay
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_hasDay"^^xsd:string,
 hasHumanReadableName  "relation c_hasDay"^^xsd:string,
 hasDescription  "links a day to a date"^^xsd:string
Individual: c_dcFirstName
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_dcFirstName"^^xsd:string,
 hasDescription  "links the first name to a person"^^xsd:string,
 hasHumanReadableName  "relation c_dcFirstName"^^xsd:string
Individual: c_Process_Description
Types:
Process_Description,
Class_Representator
Facts:
 belongs_to_process_step  Step_1_mapping_process,
 belongs_to_step  Mapping_Step,
 has_version  OM2R_version_3,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_Process_Description"^^xsd:string,
 has_editing_priority  "1"^^xsd:int,
 hasHumanReadableName  "mapping process description"^^xsd:string,
 hasDescription  "Specifics of the mapping proces"^^xsd:string
Individual: c_Creation_Date
Types:
Creation_Date,
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasHumanReadableName  "Date of a creaton"^^xsd:string,
 hasDescription  "Date of a creation"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_Creation_Date"^^xsd:string
Individual: c_Management_Details
Types:
Management_Details,
Class_Representator
Facts:
```

```
 has_version  OM2R_version_3,
 belongs_to_step  Mapping_Management_Step,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_Management_Details"^^xsd:string,
 has_editing_priority  "500"^^xsd:int,
 hasDescription  "Provides details for mapping management over time"^^xsd:string,
 hasHumanReadableName  "Mapping Management Details"^^xsd:string
Individual: Matching_Requirments_for_step2
Types:
Matching_Requirments
Facts:
 belongs_to_process_step  Step_2_matching_process,
 belongs_to_step  Matching_Step,
 textValue  "Matching must be fast"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_Matching_Requirments"^^xsd:string,
 hasDescription  "Represents matching requirements"^^xsd:string,
 has_editing_priority  "1"^^xsd:int,
 hasHumanReadableName  "Matching Requirements"^^xsd:string
Individual: Matching_Step
Annotations:
rdfs:comment "Information how confirmed mapping are selected from the mapping candidates"^^xsd:string
Types:
Lifecycle_Process_Steps
Facts:
 is_Step  Step_2,
 hasDescription  "Matching Process step of the ontology mapping lifecycle"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#Matching_Process_Step"^^xsd:string,
 hasHumanReadableName  "Matching Process step"^^xsd:string
Individual: c_has_Mapping_requirmentsClass_Representator_14
Types:
Class_Representator
Facts:
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#c_has_Mapping_requirmentsClass_Representator_14"^^xsd
      :string,
 hasDescription  "links mapping to specific mapping requirments"^^xsd:string,
 hasHumanReadableName  "relation c_has_Mapping_requirments"^^xsd:string
Individual: c_has_version
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasHumanReadableName  "relation c_has_version"^^xsd:string,
 hasDescription  "links om2r model element to a particular version of the model"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_has_version"^^xsd:string
Individual: c_has_Mapping_Process_Step
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#c_has_Mapping_Process_Step"^^xsd:string,
 hasHumanReadableName  "relation c_has_Mapping_Process_Step"^^xsd:string,
 hasDescription  "links process details to mapping process step"^^xsd:string
Individual: c_is_Target
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasHumanReadableName  "relation c_is_Target"^^xsd:string,
 hasDescription  "links ontology to a target like source"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_is_Target"^^xsd:string
Individual: c_ChangeNotification
Types:
ChangeNotification
Facts:
 belongs_to_step  Mapping_Management_Step,
 hasDescription  "Systems used to notify user about changes to the mapping"^^xsd:string,
 hasUniqueIdentifier  ""^^xsd:string,
 hasHumanReadableName  "Change Notification Method"^^xsd:string,
 has_editing_priority  "1"^^xsd:int
Individual: c_belongs_to_matching_process
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasDescription  "Defines relations between matching steps and the mathcing process"^^xsd:string,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#c_belongs_to_matching_process"^^xsd:string,
 hasHumanReadableName  "relation_belongs_to_matching_process"^^xsd:string
Individual: Step_4
Types:
```

```
Steps
Facts:
 hasDescription  "Step 4 in a process"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#Step_4"^^xsd:string,
 hasHumanReadableName  "Step 4"^^xsd:string
Individual: Demo_Book_Ontology
Types:
Ontology
Facts:
 is_Target  Source_Ontology,
 date_of_mapping_creation  creation_data_2014-02-03,
 has_version  OM2R_version_3,
 has_notation  OWL_DL,
 has_domain  Book,
 has_source  Demo_Ontology_Book_URL,
 has_formal_language  RDF_XML,
 hasDescription  "Book ontology"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#Demo_Book_Ontology"^^xsd:string,
 has_editing_priority  "1"^^xsd:int,
 hasHumanReadableName  "Book ontology"^^xsd:string
Individual: Step_3
Types:
Steps
Facts:
 hasDescription  "Step 3 of a process"^^xsd:string,
 hasHumanReadableName  "Step 3"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#Step_3"^^xsd:string
Individual: c_mapping_formal_languages
Types:
Mapping_Formal_Language,
Descriptors
Facts:
 belongs_to_step  Mapping_Step,
 hasDescription  "Formal mapping languages"^^xsd:string,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#c_mapping_formal_languages"^^xsd:string,
 hasHumanReadableName  "Formal mapping languages"^^xsd:string,
 has_editing_priority  "200"^^xsd:int
Individual: Library
Types:
Domain
Facts:
 belongs_to_step  Identification_Step,
 textValue  "This ontology represents element relevant for libraries"^^xsd:string
Individual: StrucSubsDistAlignment
Types:
Matching_Algorithm
Facts:
 belongs_to_process_step  Step_1_mapping_process,
 has_source  StrucSubsDistAlignment_Source,
 belongs_to_step  Matching_Step,
 hasHumanReadableName  "StrucSubsDistAlignment"^^xsd:string,
 hasDescription  "Algorithm based on StrucSubsDistAlignment"^^xsd:string,
 has_editing_priority  "10"^^xsd:int,
 hasUniqueIdentifier  ""^^xsd:string
Individual: OM2RM_Model
Types:
OM2R_Model
Facts:
 has_version  OM2R_version_3,
 hasHumanReadableName  "OM2R model for documenting ontology mappings to support mapping retrieval and
      reuse"^^xsd:string,
 hasDescription  "This onology provides the meta-data model for documenting ontology mappings with a
      focus to support ontology mapping retrieval and reuse. To enable this support the model focues
      on the ontology mapping lifecycel which includes the following phaes 1 Ontology Identification
      Phase  2 Matching Phase 3 Mapping Phase  4 Mapping Management"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#OM2RM_Model"^^xsd:string
Individual: c_ontology_formal_languages
Facts:
 hasHumanReadableName  "Ontology "^^xsd:string
Individual: c_OM2R_Model
Types:
OM2R_Model,
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasDescription  "Represent the different versions of the model"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_OM2R_Model"^^xsd:string,
 hasHumanReadableName  "OM2R model"^^xsd:string
Individual: c_Lifecycle_Process_Steps
Types:
```

```
Lifecycle_Process_Steps,
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasDescription  "Represents the process steps of the ontoloyg mappong lifecycle"^^xsd:string,
 hasHumanReadableName  "Process steps of the ontology mapping lifecycle process"^^xsd:string,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#c_Lifecycle_Process_Steps"^^xsd:string
Individual: c_ElementTheAlgorymIsBasedOn
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasDescription  "links the algorithm to a particular element it is based on"^^xsd:string,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#c_ElementTheAlgorymIsBasedOn"^^xsd:string,
 hasHumanReadableName  "relation c_ElementTheAlgorymIsBasedOn"^^xsd:string
Individual: Step_2
Types:
Steps
Facts:
 hasDescription  "Step 2 in a process"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#Step_2"^^xsd:string,
 hasHumanReadableName  "Step 2"^^xsd:string
Individual: Step_1
Types:
Steps
Facts:
 hasDescription  "Step 1 in a process"^^xsd:string,
 hasHumanReadableName  "Step 1"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#Step_1"^^xsd:string
Individual: RDF_XML
Types:
Notation
Facts:
 belongs_to_step  Identification_Step,
 compatible_to  OWL_Lite,
 compatible_to  OWL_FULL,
 compatible_to  OWL_DL,
 compatible_to  OWL_2,
 hasDescription  "RDF/XM"^^xsd:string,
 has_editing_priority  "200"^^xsd:int,
 hasHumanReadableName  "RDF/XM"^^xsd:string,
 hasUniqueIdentifier  "http://www.w3.org/TR/REC-rdf-syntax/"^^xsd:string
Individual: Step_2_matching_process
Types:
Matching_Process_Steps
Facts:
 hasAppliedThreshold  percent75,
 has_matching_requirments  Matching_Requirments_for_step2,
 has_matching_method  Automatic_Matching,
 belongs_to_matching_process  Matching_process_between_books_and_library_ontology,
 has_matching_scope  Partial_Match,
 has_matching_algorithm  StrucSubsDistAlignment,
 has_matching_tool  Ontology_Alignment_API_Version_42,
 is_Step  Step_2,
 belongs_to_step  Matching_Step,
 ElementTheAlgorymIsBasedOn  Element_rdfs_class,
 has_version  OM2R_version_3,
 hasHumanReadableName  "Second step in the matching process"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#Step_2_matching_process"^^xsd:string,
 hasDescription  "Second Step in the matching process"^^xsd:string,
 has_editing_priority  "2"^^xsd:int
Individual: c_Matching_Requirments
Types:
Matching_Requirments,
Class_Representator
Facts:
 has_version  OM2R_version_3,
 belongs_to_step  Matching_Step,
 hasDescription  "Represents matching requirements"^^xsd:string,
 has_editing_priority  "1"^^xsd:int,
 hasHumanReadableName  "Matching Requirements"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_Matching_Requirments"^^xsd:string
Individual: Turtle
Types:
Notation
Facts:
 compatible_to  OWL_2,
 compatible_to  OWL_FULL,
 compatible_to  OWL_Lite,
```

```
 belongs_to_step  Identification_Step,
 compatible_to  OWL_DL,
 hasUniqueIdentifier  "http://www.w3.org/TR/2014/REC-turtle-20140225/"^^xsd:string,
 has_editing_priority  "400"^^xsd:int,
 hasHumanReadableName  "*Turtle"^^xsd:string,
 hasDescription  "Turtle"^^xsd:string,
 hasHumanReadableName  "Turtle (Terse RDF Triple Language) is a serialization format for Resource
      Description Framework (RDF) graphs. A subset of Tim Berners-Lee and Dan Connolly's Notation3
      (N3) language, it was defined by Dave Beckett, and is a superset of the minimal N-Triples
      format. Unlike full N3, Turtle doesn't go beyond RDF's graph model. SPARQL uses a similar N3
      subset to Turtle for its graph patterns, but using N3's brace syntax for delimiting subgraphs.
      Turtle was accepted as a first working draft by the World Wide Web Consortium (W3C) RDF Working
      Group on 9 August 2011.[1]  Turtle is popular among Semantic Web developers as a human-friendly
      alternative to RDF/XML. A significant proportion of RDF toolkits include Turtle parsing and
      serializing capability. Some examples are Redland, Sesame, Jena and RDFLib."^^xsd:string
Individual: c_is_Step
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasHumanReadableName  "relation c_is_Step"^^xsd:string,
 hasDescription  "links lifecycle step to a ordered step like step 1"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_is_Step"^^xsd:string
Individual: Step_1_mapping_process
Types:
Mapping_Process_Steps
Facts:
 has_process_details  mapping_process_description_lib_to_book,
 has_mapping_requirments  Book_to_Lib_Mapping_Requirements,
 has_version  OM2R_version_3,
 has_creator  Hendrik_Thomas,
 belongs_to_step  Mapping_Step,
 has_editing_priority  "1"^^xsd:int,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#Step_1_mapping_process"^^xsd:string,
 hasHumanReadableName  "Mapping process step 1"^^xsd:string,
 hasDescription  "Step 1 in the mapping process"^^xsd:string
Individual: c_hasApplicationThatUsesTheMapping
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasHumanReadableName  "relation c_hasApplicationThatUsesTheMapping"^^xsd:string,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#c_hasApplicationThatUsesTheMapping"^^xsd:string,
 hasDescription  "links a mapping to a particular application that uses the mapping"^^xsd:string
Individual: c_Domain
Types:
Domain
Facts:
 belongs_to_step  Identification_Step,
 hasDescription  "Domain of interest"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_Domain"^^xsd:string,
 has_editing_priority  "1"^^xsd:int,
 hasHumanReadableName  "Domain of interest"^^xsd:string
Individual: Demo_Mapping_URL
Types:
Source
Facts:
 belongs_to_step  Identification_Step,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#Demo_Mapping_URL"^^xsd:string,
 hasDescription  "URL to get the mapping for lib to book ontology"^^xsd:string,
 hasHumanReadableName  "URL to dlownload the mapping"^^xsd:string,
 has_editing_priority  "1"^^xsd:int,
 urlValue  "http://www.learninginformationlibrary.com/pool/booktolibmapping"^^xsd:string,
 lastCheckedDateAndTime  "2011-09-27T14:19:32"^^xsd:dateTime
Individual: Complete_Match
Types:
Matching_Scope
Facts:
 belongs_to_process_step  Step_2_matching_process,
 belongs_to_step  Matching_Step,
 belongs_to_process_step  Step_1_matching_process,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#Complete_Match"^^xsd:string,
 has_editing_priority  "1"^^xsd:int,
 hasDescription  "All ontology elements are addressed in the match"^^xsd:string,
 hasHumanReadableName  "Complete Match"^^xsd:string
Individual: Book
Types:
Domain
Facts:
```

```
 textValue  "This ontology represents element relevant for books, representing a personal knowledge
      collection of a researcher"^^xsd:string
Individual: Target_Ontology
Types:
Mapping_Targets
Facts:
 belongs_to_step  Identification_Step
Individual: c_Matching_Algorithm
Types:
Class_Representator,
Matching_Algorithm
Facts:
 has_version  OM2R_version_3,
 belongs_to_step  Matching_Step,
 hasHumanReadableName  "Matching_Algorithm"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_Matching_Algorithm"^^xsd:string,
 has_editing_priority  "100"^^xsd:int,
 hasDescription  "Specific algoritm used by the tool to find matching candidates"^^xsd:string
Individual: StringDistAlignment
Types:
Matching_Algorithm
Facts:
 has_source  StringDis_Source,
 compatible_to  Element_rdf_label,
 belongs_to_step  Matching_Step,
 hasUniqueIdentifier
      "ttp://www.modellmapping.org/om2rdemo.owl#fr.inrialpes.exmo.align.impl.method.StringDistAlignmen
      t"^^xsd:string,
 has_editing_priority  "10"^^xsd:int,
 hasDescription  "StringDistAlignment"^^xsd:string,
 hasHumanReadableName  "StringDistAlignment"^^xsd:string
Individual: Step_1_matching_process
Types:
Matching_Process_Steps
Facts:
 ElementTheAlgorymIsBasedOn  Element_rdf_label,
 has_matching_tool  Ontology_Alignment_API_Version_42,
 has_matching_scope  Complete_Match,
 has_matching_algorithm  StringDistAlignment,
 has_matching_method  Automatic_Matching,
 belongs_to_matching_process  Matching_process_between_books_and_library_ontology,
 has_matching_requirments  Matching_Requirments_83,
 is_Step  Step_1,
 hasAppliedThreshold  percent100,
 belongs_to_step  Matching_Step,
 has_editing_priority  "1"^^xsd:int,
 hasHumanReadableName  "First step in the matching process"^^xsd:string,
 hasDescription  "First step in the matching process"^^xsd:string,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#c_Step_1_matching_process"^^xsd:string
Individual: c_mapping
Types:
Mapping,
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasDescription  "Model element where the instance represent the mappings"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_mapping"^^xsd:anyURI,
 hasHumanReadableName  "Mapping"^^xsd:string,
 has_editing_priority  "2"^^xsd:int
Individual: LTM
Types:
Notation
Facts:
 compatible_to  Topic_Maps,
 belongs_to_step  Identification_Step,
 has_editing_priority  "1"^^xsd:int,
 hasHumanReadableName  "Linera Topic Maps Notation"^^xsd:string,
 hasHumanReadableName  "LTM"^^xsd:string,
 hasDescription  "LTM"^^xsd:string,
 hasUniqueIdentifier  "http://www.ontopia.net/download/ltm.html"^^xsd:string
Individual: c_hasHumanReadableName
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasHumanReadableName  "relation c_hasHumanReadableName"^^xsd:string,
 hasDescription    "om2r  embedded  meta-data  documentation  links  element  to  human  readable
      name"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_hasHumanReadableName"^^xsd:string
Individual: c_has_notation
```

```
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasHumanReadableName  "relation c_has_notation"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_has_notation"^^xsd:string,
 hasDescription  "links object to a particular notation"^^xsd:string
Individual: Alignment_API_Source
Types:
Source
Facts:
 belongs_to_step  Matching_Step,
 urlValue  "http://alignapi.gforge.inria.fr/"^^xsd:string,
 hasDescription  "Source to find the alignment API"^^xsd:string,
 has_editing_priority  "1"^^xsd:int,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#Alignment_API_Source"^^xsd:string,
 hasHumanReadableName  "Source for the Alignment API"^^xsd:string,
 lastCheckedDateAndTime  "2011-09-27T15:18:16"^^xsd:dateTime
Individual: c_matching_scope
Types:
Class_Representator,
Matching_Scope
Facts:
 belongs_to_step  Matching_Step,
 hasHumanReadableName  "Matching Scope"^^xsd:string,
 has_editing_priority  "9"^^xsd:int,
 hasDescription  "Defines the scope of the current matching process"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_matching_scope"^^xsd:string
Individual: c_has_matching_scope
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_has_matching_scope"^^xsd:string,
 hasDescription  "relation matching to a particular matching scope"^^xsd:string,
 hasHumanReadableName  "relation c_has_matching_scope"^^xsd:string
Individual: c_source_ontology
Types:
Ontology,
Class_Representator
Facts:
 has_version  OM2R_version_3,
 is_Target  Source_Ontology,
 hasHumanReadableName  "Source Ontology"^^xsd:string,
 hasDescription  "Represents ontologies that are the source of an mapping"^^xsd:string,
 has_editing_priority  "300"^^xsd:int,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_source_ontology"^^xsd:string
Individual: c_Model_Elements
Types:
Model_Elements,
Class_Representator
Facts:
 belongs_to_step  Mapping_Management_Step,
 belongs_to_step  Identification_Step,
 belongs_to_step  Mapping_Step,
 has_version  OM2R_version_3,
 belongs_to_step  Matching_Step,
 has_editing_priority  "1"^^xsd:int,
 hasHumanReadableName  "Model elements of the OM2R model"^^xsd:string,
 hasUniqueIdentifier  ""^^xsd:string,
 hasDescription  "Represents the general model elements of the OM2R"^^xsd:string
Individual: c_dcLastName
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasDescription  "links last name of a person"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_dcLastName"^^xsd:string,
 hasHumanReadableName  "relation c_dcLastName"^^xsd:string
Individual: c_Matching_Method
Types:
Matching_Method,
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasHumanReadableName  "Matching process"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_Matching_Method"^^xsd:string,
 has_editing_priority  "100"^^xsd:int,
 hasDescription  "General type of the matching process"^^xsd:string
Individual: Mafra
Types:
```

```
Matching_Tool
Facts:
 belongs_to_step  Matching_Step
Individual: c_target_ontology
Types:
Ontology,
Class_Representator
Facts:
 has_version  OM2R_version_3,
 is_Target  Target_Ontology,
 has_editing_priority  "200"^^xsd:int,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_target_ontology"^^xsd:string,
 hasDescription  "Represents an ontology that is the target of a mapping"^^xsd:string,
 hasHumanReadableName  "Target ontology"^^xsd:string
Individual: OWL_FULL
Types:
Formal_Language
Facts:
 belongs_to_step  Identification_Step,
 compatible_to  RDF_XML,
 compatible_to  N3,
 hasDescription  "OWL Full"^^xsd:string,
 hasUniqueIdentifier  "http://www.w3.org/TR/owl-ref/"^^xsd:string,
 has_editing_priority  "100"^^xsd:int,
 hasHumanReadableName  "OWL Full"^^xsd:string
Individual: c_has_matching_requirments
Types:
Class_Representator
Facts:
 hasHumanReadableName  "relation c_has_matching_requirments "^^xsd:string,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#c_has_matching_requirments"^^xsd:string,
 hasDescription  "matching linked to matching requirments"^^xsd:string
Individual: Trinity_Library_Searcher
Types:
ApplicationThatUseTheTheMapping
Facts:
 belongs_to_step  Mapping_Management_Step,
 has_source  Trinity_Searcher_Source,
 has_editing_priority  "1"^^xsd:int,
 hasDescription  "Library retrieval system used by Trinity College Library"^^xsd:string,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#Trinity_Library_Searcher"^^xsd:string,
 hasHumanReadableName  "Trinity Library Searcher System"^^xsd:string
Individual: c_Mapping_Requirements
Types:
Class_Representator,
Mapping_Requirements
Facts:
 has_version  OM2R_version_3,
 belongs_to_step  Mapping_Step,
 hasHumanReadableName  "Mapping requirements"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_Mapping_Requirements"^^xsd:string,
 hasDescription  "Defined specific requirments which need to be fullfiled in the mapping"^^xsd:string
Individual: c_belongs_to_step
Facts:
 has_editing_priority  "0"^^xsd:int,
 hasHumanReadableName  "relation c_belongs_to_step"^^xsd:string
Individual: c_Matching_Process_Steps
Types:
Matching_Process_Steps,
Class_Representator
Facts:
 belongs_to_step  Matching_Step,
 has_version  OM2R_version_3,
 has_editing_priority  "1"^^xsd:int,
 hasHumanReadableName  "Matching Steps"^^xsd:string,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#c_Matching_Process_Steps"^^xsd:string,
 hasDescription  "Represents matching steps"^^xsd:string
Individual: StringDis_Source
Types:
Source
Facts:
 belongs_to_step  Matching_Step,
 lastCheckedDateAndTime  "2011-09-27T15:13:58"^^xsd:dateTime,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#StringDis_Source"^^xsd:string,
 hasDescription  "URL to download the code for the String Dinstance Algorithm"^^xsd:string,
 has_editing_priority  "1"^^xsd:int,
 urlValue  "http://www.modellmapping.org/om2rdemo.owl#StringDis_Source"^^xsd:string,
 hasHumanReadableName  "URL to download the code for the String Dinstance Algorithm"^^xsd:string
```

```
Individual: OWL_2
Types:
Formal_Language
Facts:
 belongs_to_step  Identification_Step,
 compatible_to  RDF_XML,
 compatible_to  N3,
 has_editing_priority  "100"^^xsd:int,
 hasUniqueIdentifier  "http://www.w3.org/TR/owl2-overview/"^^xsd:string,
 hasDescription  "OWL 2"^^xsd:string,
 hasHumanReadableName  "OWL 2 Web Ontology Language Document Overview (Second Edition)"^^xsd:string,
 hasHumanReadableName  "*OWL 2"^^xsd:string
Individual: Automatic_Matching
Annotations:
rdfs:label "Automatic Matching"^^xsd:string,
rdfs:comment "Matching pairs are generated automatically"^^xsd:string
Types:
Matching_Method
Facts:
 compatible_to  Ontology_Alignment_API_Version_32,
 belongs_to_step  Matching_Step,
 compatible_to  Ontology_Alignment_API_Version_42,
 hasDescription  "Automated Matching process"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#Automatic_Matching"^^xsd:string,
 hasHumanReadableName  "automated matching"^^xsd:string,
 has_editing_priority  "100"^^xsd:int
Individual: c_dcAddress
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasDescription  "links a address to a person"^^xsd:string,
 hasHumanReadableName  "relation c_dcAddress"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_dcAddress"^^xsd:string
Individual: c_has_matching_algorithm
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasDescription  "links to tool to matching algorithm"^^xsd:string,
 hasHumanReadableName  "relation c_has_matching_algorithm "^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_has_matching_algorithm"^^xsd:string
Individual: OWL_Lite
Types:
Formal_Language
Facts:
 belongs_to_step  Identification_Step,
 compatible_to  RDF_XML,
 hasDescription  "OWL lite"^^xsd:string,
 hasUniqueIdentifier  "http://www.w3.org/TR/owl-ref/"^^xsd:string,
 has_editing_priority  "100"^^xsd:int,
 hasHumanReadableName  "OWL Lite"^^xsd:string
Individual: Book_to_Lib_Mapping_Requirements
Types:
Mapping_Requirements
Facts:
 belongs_to_step  Mapping_Step,
 belongs_to_process_step  Step_1_mapping_process,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#Book_to_Lib_Mapping_Requirements"^^xsd:string,
 has_editing_priority  "1"^^xsd:int,
 hasDescription  "Requirments for mapping between book and library ontology"^^xsd:string,
 textValue  "Mapping needs to be verified by experts to ensure quality"^^xsd:string,
 hasHumanReadableName     "Requirments   for   mapping   between   book   and   library   ontology
              DefaultRDFSDatatype(http://www.w3.org/2001/XMLSchema#string                   of
      [DefaultRDFSNamedClass(http://www.w3.org/2000/01/rdf-schema#Datatype)])"^^xsd:string
Individual: mapping_process_description_lib_to_book
Types:
Process_Description
Facts:
 belongs_to_process_step  Step_1_mapping_process,
 belongs_to_step  Mapping_Step,
 hasDescription  "Process description"^^xsd:string,
 textValue  "Candidates reviewed by experts"^^xsd:string,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#c_mapping_process_description_lib_to_book"^^xsd:strin
      g,
 hasHumanReadableName  "Process descriptiion of lib to book library"^^xsd:string
Individual: Demo_Library_Ontology
Types:
Ontology
```

```
Facts:
 date_of_mapping_creation  creation_data_2014-02-03,
 has_version  OM2R_version_3,
 has_formal_language  RDF_XML,
 has_notation  OWL_DL,
 has_domain  Library,
 is_Target  Target_Ontology,
 has_source  Demo_Ontology_Library_URL,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#Demo_Library_Ontology"^^xsd:string,
 hasHumanReadableName  "Library Ontology"^^xsd:string,
 hasDescription  "Ontology used in libraries"^^xsd:string,
 has_editing_priority  "2"^^xsd:int
Individual: Ontology_Alignment_API_Version_32
Types:
Matching_Tool
Facts:
 compatible_to  StrucSubsDistAlignment,
 has_version  OM2R_version_3,
 has_source  Alignment_API_Source,
 compatible_to  StringDistAlignment,
 belongs_to_step  Matching_Step,
 belongs_to_process_step  Step_1_matching_process,
 hasDescription  "Ontology Alignment API 4.2"^^xsd:string,
 hasUniqueIdentifier  "Ontology alignment API 3.2"^^xsd:string,
 has_editing_priority  "1"^^xsd:int,
 hasHumanReadableName  "Ontology Alignment API Version 4.2"^^xsd:string
Individual: c_OM2R_Modell
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3
Individual: Demo_Book_vs_Library_Mapping
Types:
Mapping
Facts:
 has_source  Demo_Mapping_URL,
 has_formal_language_mapping  Inria,
 is_Addressing  Demo_Library_Ontology,
 has_version  OM2R_version_3,
 date_of_mapping_creation  creation_data_2014-02-03,
 hasApplicationThatUsesTheMapping  Trinity_Library_Searcher,
 has_matching_method  Automatic_Matching,
 is_Addressing  Demo_Book_Ontology,
 has_notation  RDF_XML,
 hasChangeNotification  Trinity_Library_RSS,
 hasDescription  "Mapping betwen the book and library ontology"^^xsd:string,
 has_editing_priority  "1"^^xsd:int,
 hasHumanReadableName  "Demo Book vs Library Mapping"^^xsd:string,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#Demo_Book_vs_Library_Mapping"^^xsd:string
Individual: Manuel_Matching
Annotations:
rdfs:comment "Matching was conducted by human manually"^^xsd:string,
rdfs:label "Manuel Matching"^^xsd:string
Types:
Matching_Method
Facts:
 belongs_to_step  Matching_Step,
 hasDescription  "Manual Process"^^xsd:string,
 hasHumanReadableName  "http://www.modellmapping.org/om2rdemo.owl#Manuel_Matching"^^xsd:string,
 has_editing_priority  "100"^^xsd:int,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#Manuel_Matching"^^xsd:string
Individual: Demo_Ontology_Library_URL
Types:
Source
Facts:
 belongs_to_step  Identification_Step,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#Demo_Ontology_Library_URL"^^xsd:string,
 hasHumanReadableName  "URL to get the library ontology"^^xsd:string,
 hasDescription  "URL to get the library ontology"^^xsd:string,
 urlValue  "http://www.learninginformationlibrary.com/pool/library"^^xsd:string,
 textValue  "http://www.modellmapping.org/library.owl"^^xsd:string,
 lastCheckedDateAndTime  "2011-09-22T17:04:42"^^xsd:dateTime,
 has_editing_priority  "1"^^xsd:int
Individual: percent100
Types:
AppliedThreshold
Facts:
 belongs_to_process_step  Step_1_matching_process,
 belongs_to_step  Matching_Step,
```

```
 has_editing_priority  "1"^^xsd:int,
 hasDescription  "100% full matching required, e.g. identical names"^^xsd:string,
 hasHumanReadableName  "100%"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_AppliedThreshold"^^xsd:string
Individual: c_hasYear
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasDescription  "links a year to date"^^xsd:string,
 hasHumanReadableName  "relation c_hasYear"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_hasYear"^^xsd:string
Individual: c_hasMonth
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasDescription  "links month to a date"^^xsd:string,
 hasHumanReadableName  "relation c_hasMonth"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_hasMonth"^^xsd:string
Individual: c_Matching_Tool
Types:
Matching_Tool,
Class_Representator
Facts:
 has_version  OM2R_version_3,
 belongs_to_step  Matching_Step,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_Matching_Tool"^^xsd:string,
 has_editing_priority  "1"^^xsd:int,
 hasHumanReadableName  "Matching tools"^^xsd:string,
 hasDescription  "Tools that are used to identify mapping candidates = matchings"^^xsd:string
Individual: c_has_matching_tool
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasHumanReadableName  "relation c_has_matching_tool"^^xsd:string,
 hasDescription  "matching process linked to matching tool"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_has_matching_tool"^^xsd:string
Individual: c_OM2R_model_version
Types:
OM2R_model_version,
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasHumanReadableName  "OM2R model"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_OM2R_model_version"^^xsd:string,
 hasDescription  "Represents the model version of the OM2R"^^xsd:string
Individual: c_has_Language
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasDescription  "links to natural language, e.g English"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_has_Language"^^xsd:string,
 hasHumanReadableName  "relation has language"^^xsd:string
Individual: c_Matching_Process
Types:
Matching_Process
Facts:
 belongs_to_step  Matching_Step,
 hasDescription  "Represents a specific instance oif a matching procress"^^xsd:string,
 hasHumanReadableName  "Matching process"^^xsd:string,
 has_editing_priority  "50"^^xsd:int,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_Matching_Process"^^xsd:string
Individual: c_Mapping_Process
Types:
Mapping_Process,
Class_Representator
Facts:
 has_version  OM2R_version_3,
 belongs_to_step  Mapping_Step,
 has_editing_priority  "1"^^xsd:int,
 hasHumanReadableName  "Mapping process"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_Mapping_Process"^^xsd:string,
 hasDescription  "Represent the speciific mapping processes"^^xsd:string
Individual: c_Mapping_Creator
Types:
Class_Representator,
Mapping_Creator
Facts:
```

```
 belongs_to_process_step  Step_1_mapping_process,
 has_version  OM2R_version_3,
 belongs_to_step  Mapping_Step,
 hasDescription  "Details of the human creator of the mapping"^^xsd:string,
 has_editing_priority  "1"^^xsd:int,
 hasHumanReadableName  "Mapping Creator"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_Mapping_Creator"^^xsd:string
Individual: Mapping_Management_Step
Types:
Lifecycle_Process_Steps
Facts:
 is_Step  Step_4,
 hasHumanReadableName  "Management Process "^^xsd:string,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#Mapping_Management_Process_Step"^^xsd:string,
 hasDescription  "Management Process step of the ontology mapping lifecycle"^^xsd:string
Individual: c_notation
Types:
Class_Representator,
Notation
Facts:
 has_version  OM2R_version_3,
 belongs_to_step  Identification_Step,
 hasDescription  "Notation"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_notation"^^xsd:string,
 hasHumanReadableName  "Notation"^^xsd:string,
 has_editing_priority  "100"^^xsd:int
Individual: c_has_creator
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_has_creator"^^xsd:string,
 hasDescription  "links object to is creator"^^xsd:string,
 hasHumanReadableName  "relatiom has creator"^^xsd:string
Individual: c_Class_Placeholder
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_Class_Placeholder"^^xsd:string,
 has_editing_priority  "1"^^xsd:int,
 hasDescription   "This is a wor a round to enable relation to be added to a class in OWL
      DL"^^xsd:string,
 hasHumanReadableName  "Collect all instance that represents the class rather an instance of the
      class"^^xsd:string
Individual: Mixed_Matching
Annotations:
rdfs:comment "Matching was conducted by humans and automated processes"^^xsd:string,
rdfs:label "Mixed Matching"^^xsd:string
Types:
Matching_Method
Facts:
 compatible_to  Ontology_Alignment_API_Version_42,
 belongs_to_step  Matching_Step,
 compatible_to  Ontology_Alignment_API_Version_32,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#Mixed_Matching"^^xsd:string,
 hasDescription  "Process where automated and manual process is combined"^^xsd:string,
 hasHumanReadableName  "Mixed matching process"^^xsd:string,
 has_editing_priority  "100"^^xsd:int
Individual: c_Matching_Process_Details
Types:
Matching_Process_Details,
Class_Representator
Facts:
 belongs_to_step  Matching_Step,
 has_version  OM2R_version_3,
 hasHumanReadableName  "Matching process details"^^xsd:string,
 hasDescription  "Describes details of the applied matching process"^^xsd:string,
 has_editing_priority  "100"^^xsd:int,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#c_Matching_Process_Details"^^xsd:string
Individual: Demo_Ontology_Book_URL
Types:
Source
Facts:
 belongs_to_step  Identification_Step,
 has_editing_priority  "1"^^xsd:int,
 hasHumanReadableName  "URL to get the book ontology"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#Demo_Ontology_Book_URL"^^xsd:string,
 textValue  "http://www.modellmapping.org/book.owl"^^xsd:string,
```

```
 hasDescription  "URL to get the book "^^xsd:string,
 lastCheckedDateAndTime  "2011-09-14T17:04:39"^^xsd:dateTime,
 urlValue  "http://www.learninginformationlibrary.com/pool/book"^^xsd:string
Individual: c_hasDescription
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_hasDescription"^^xsd:string,
 hasHumanReadableName  "relation c_hasDescription"^^xsd:string,
 hasDescription     "om2r  embedded  meta-data  documentation  links  element  to  human  readable
      description"^^xsd:string
Individual: Source_Ontology
Types:
Mapping_Targets
Facts:
 belongs_to_step  Identification_Step
Individual: Ontology_Alignment_API_Version_42
Types:
Matching_Tool
Facts:
 belongs_to_step  Matching_Step,
 compatible_to  StringDistAlignment,
 has_source  Alignment_API_Source,
 has_version  OM2R_version_3,
 belongs_to_process_step  Step_1_matching_process,
 has_editing_priority  "1"^^xsd:int,
 hasDescription  "Ontology Alignment API 4.2"^^xsd:string,
 hasHumanReadableName  "Ontology Alignment API Version 4.2"^^xsd:string,
 hasUniqueIdentifier  "Ontology alignment API 4.2"^^xsd:string
Individual: c_hasTimeZone
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_hasTimeZone"^^xsd:string,
 hasDescription  "links time zone to a date"^^xsd:string,
 hasHumanReadableName  "relation c_hasTimeZone"^^xsd:string
Individual: c_has_source
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasHumanReadableName  "relation c_has_source"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_has_source"^^xsd:string,
 hasDescription  "links object to a particular source where to find the source code"^^xsd:string
Individual: Mapping_between_book_and_lib_ontology
Types:
Mapping_Process
Facts:
 has_mapping_process_step  Step_1_mapping_process,
 belongs_to_step  Mapping_Step,
 hasHumanReadableName  "Mapping_between_book_and_lib_ontology"^^xsd:string,
 has_editing_priority  "1"^^xsd:int,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#Mapping_between_book_and_lib_ontology"^^xsd:string,
 hasDescription  "Specific maping between_book_and_lib_ontology"^^xsd:string
Individual: Topic_Maps
Types:
Formal_Language
Facts:
 belongs_to_step  Identification_Step,
 compatible_to  LTM,
 compatible_to  XTM,
 hasUniqueIdentifier  "http://www.topicmaps.org/"^^xsd:string,
 has_editing_priority  "100"^^xsd:int,
 hasHumanReadableName  "Topic Maps"^^xsd:string,
 hasDescription  "Topic Maps"^^xsd:string
Individual: c_ApplicationThatUseTheMapping
Types:
Class_Representator,
ApplicationThatUseTheTheMapping
Facts:
 belongs_to_step  Mapping_Management_Step,
 has_version  OM2R_version_3,
 hasHumanReadableName  "Known Application that use the mapping"^^xsd:string,
 hasDescription  "Known Application that use the mapping, this knowledge is important to understand
      the implication of change to the mapping"^^xsd:string,
 has_editing_priority  "500"^^xsd:int,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#c_ApplicationThatUseTheMapping"^^xsd:string
```

```
Individual: c_has_editing_priority
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasDescription  "defnes a number which indicate a editing priority number"^^xsd:string,
 hasHumanReadableName  "relation c_has_editing_priority"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_has_editing_priority"^^xsd:string
Individual: creation_data_2014-02-03
Types:
Creation_Date
Facts:
 belongs_to_step  Mapping_Step,
 hasTimeZone  "Auckland"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#2014-02-03"^^xsd:string,
 hasHumanReadableName  "3rd Feb 2014"^^xsd:string,
 hasDay  "3"^^xsd:string,
 hasYear  "2014"^^xsd:string,
 hasMonth  "2"^^xsd:string
Individual: c_dcEmail
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasDescription  "links email to a person"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_dcEmail"^^xsd:string,
 hasHumanReadableName  "relation c_dcEmail"^^xsd:string
Individual: c_hasUniqueIdentifier
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasDescription   "om2r  embedded  meta-data  documentation  links  element  to  a  unique  identifier
      "^^xsd:string,
 hasHumanReadableName  "relation c_hasUniqueIdentifier"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_hasUniqueIdentifier"^^xsd:string
Individual: Hendrik_Thomas
Types:
Mapping_Creator
Facts:
 belongs_to_process_step  Step_1_mapping_process,
 belongs_to_step  Mapping_Step,
 has_editing_priority  "1"^^xsd:int,
 dcFirstName  "Hendrk"^^xsd:string,
 hasHumanReadableName  "Hendrik Thomas"^^xsd:string,
 hasDescription  "User Hendrik Thomas who created the mapping"^^xsd:string,
 dcLastName  "Thomas"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#Hendrik_Thomas"^^xsd:string,
 dcEmail  ""^^xsd:string
Individual: XTM
Types:
Notation
Facts:
 compatible_to  Topic_Maps,
 belongs_to_step  Identification_Step,
 has_editing_priority  "300"^^xsd:int,
 hasHumanReadableName  "XML Topic Maps"^^xsd:string,
 hasHumanReadableName  "*XTM"^^xsd:string,
 hasUniqueIdentifier  "http://www.topicmaps.org/xtm/1.0/"^^xsd:string,
 hasDescription  "XML Topic Maps"^^xsd:string
Individual: c_urlValue
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_urlValue"^^xsd:string,
 hasDescription  "links URL to a element"^^xsd:string,
 hasHumanReadableName  "relation c_urlValue"^^xsd:string
Individual: c_has_Domain
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasDescription  "links an object to its content domain, e.g. library"^^xsd:string,
 hasHumanReadableName  "relation domain"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_has_Domain"^^xsd:string
Individual: c_Language
Types:
Language,
Class_Representator
Facts:
```

```
 has_version  OM2R_version_3,
 has_editing_priority  "1"^^xsd:int,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_Language"^^xsd:string,
 hasHumanReadableName  "Language"^^xsd:string,
 hasDescription  "Language information"^^xsd:string
Individual: c_AppliedThreshold
Types:
AppliedThreshold
Facts:
 belongs_to_step  Matching_Step,
 has_editing_priority  "1"^^xsd:int,
 hasDescription  "Applied Threshold defines the level of corropspondes a matching needs to satisfay to
      be accept by the algorithm used"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_AppliedThreshold"^^xsd:string,
 hasHumanReadableName  "Applied Threshold"^^xsd:string
Individual: percent75
Types:
AppliedThreshold
Facts:
 belongs_to_process_step  Step_2_matching_process,
 belongs_to_step  Matching_Step,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_AppliedThreshold"^^xsd:string,
 hasHumanReadableName  "75%"^^xsd:string,
 hasDescription  "100% full matching required, e.g. identical names"^^xsd:string,
 has_editing_priority  "1"^^xsd:int
Individual: c_is_Addressing
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasDescription  "links a mapping to ontologies"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_is_Addressing"^^xsd:string,
 hasHumanReadableName  "relation c_is_Addressing"^^xsd:string
Individual: Mafra_Source
Types:
Source
Facts:
 belongs_to_step  Matching_Step,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#Mafra_URL"^^xsd:string,
 hasHumanReadableName  "Mafra"^^xsd:string,
 hasHumanReadableName  "MAFRA - A MApping FRAmework for Distributed Ontologies in the Semantic
      Web"^^xsd:string,
 has_editing_priority  "100"^^xsd:int,
 hasDescription  "Mafra tool"^^xsd:string,
 urlValue  "http://www.modellmapping.org/om2rdemo.owl#Mafra_Source"^^xsd:string,
 lastCheckedDateAndTime  "2014-05-15T15:03:34"^^xsd:dateTime
Individual: c_has_process_step
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasHumanReadableName  "relation c_has_process_step"^^xsd:string,
 hasDescription  "relation of lifecycle to lifecycle steps"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_has_process_step"^^xsd:string
Individual: c_has_formal_language_mapping
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasDescription  "links mapping to its formal mapping language used to represent the
      mapping"^^xsd:string,
 hasHumanReadableName  "relation c_has_formal_language_mapping"^^xsd:string,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#c_has_formal_language_mapping"^^xsd:string
Individual: Element_rdf_label
Types:
Element_the_matching_is_based_on
Facts:
 belongs_to_step  Step_1_matching_process,
 belongs_to_process_step  Step_1_matching_process,
 has_editing_priority  "21"^^xsd:int,
 hasHumanReadableName  "RDF Labe"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#Element_rdf_label"^^xsd:string,
 hasDescription  "RDF Label to name elements"^^xsd:string,
 textValue  "rdfs:label"^^xsd:string
Individual: c_Steps
Types:
Steps,
Class_Representator
Facts:
 has_version  OM2R_version_3,
```

```
 hasDescription  "Prepresents the order of steps in a process"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_Steps"^^xsd:string,
 hasHumanReadableName  "Process steps"^^xsd:string
Individual: Inria
Types:
Mapping_Formal_Language
Facts:
 belongs_to_step  Identification_Step,
 has_editing_priority  "100"^^xsd:int,
 hasHumanReadableName  "Inria"^^xsd:string,
 hasHumanReadableName  "*Ontology Alignment Format"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#Inria"^^xsd:string,
 hasDescription  "Format for representing ontology matchings"^^xsd:string
Individual: c_lastCheckedDateAndTime
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasDescription  "links a last check date to a element"^^xsd:string,
 hasHumanReadableName  "relation c_lastCheckedDateAndTime"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_lastCheckedDateAndTime"^^xsd:string
Individual: N3
Types:
Notation
Facts:
 compatible_to  OWL_Lite,
 compatible_to  OWL_2,
 belongs_to_step  Identification_Step,
 compatible_to  OWL_DL,
 compatible_to  OWL_FULL,
 hasHumanReadableName  "*N3"^^xsd:string,
 hasUniqueIdentifier  "http://www.w3.org/TeamSubmission/n3/"^^xsd:string,
 hasDescription  "N3"^^xsd:string,
 hasHumanReadableName  "Notation3 (N3): A readable RDF syntax"^^xsd:string,
 has_editing_priority  "1"^^xsd:int
Individual: c_Descriptors
Types:
Descriptors
Facts:
 belongs_to_step  Identification_Step,
 belongs_to_step  Mapping_Management_Step,
 belongs_to_step  Matching_Step,
 belongs_to_step  Mapping_Step,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_Descriptors"^^xsd:string,
 hasDescription  "Collects descriptors used in the model"^^xsd:string,
 hasHumanReadableName  "Descriptors used in the OM2R model"^^xsd:string,
 has_editing_priority  "1"^^xsd:int
Individual: c_textValue
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasDescription  "links text value to a element"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_textValue"^^xsd:string,
 hasHumanReadableName  "relation c_textValue"^^xsd:string
Individual: OM2R_version_1
Types:
OM2R_model_version
Facts:
 hasDescription  "OM2R version 1"^^xsd:string,
 hasHumanReadableName  "OM2R version 1"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#OM2R_version_1"^^xsd:string
Individual: OWL_DL
Annotations:
rdfs:comment "http://www.w3.org/2002/07/owl."^^xsd:string
Types:
Formal_Language
Facts:
 compatible_to  RDF_XML,
 compatible_to  Turtle,
 compatible_to  N3,
 belongs_to_step  Identification_Step,
 hasUniqueIdentifier  "http://www.w3.org/TR/owl-guide/"^^xsd:string,
 hasHumanReadableName  "OWL DL"^^xsd:string,
 hasDescription  "OWL DL"^^xsd:string,
 has_editing_priority  "100"^^xsd:int
Individual: c_Mapping_Process_Steps
Types:
Class_Representator,
Mapping_Process_Steps
Facts:
```

```
 belongs_to_step  Mapping_Step,
 has_version  OM2R_version_3,
 hasHumanReadableName  "Mapping process"^^xsd:string,
 has_editing_priority  "1"^^xsd:int,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_Mapping_Process_Steps"^^xsd:string,
 hasDescription  "Defined process steps of a given mapping"^^xsd:string
Individual: OM2R_version_2
Types:
OM2R_model_version
Facts:
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#OM2R_version_2"^^xsd:string,
 hasDescription  "OM2R version 2"^^xsd:string,
 hasHumanReadableName  "OM2R version 2"^^xsd:string
Individual: OM2R_version_3
Types:
OM2R_model_version
Facts:
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#OM2R_version_3"^^xsd:string,
 hasHumanReadableName  "OM2R version 3"^^xsd:string,
 hasDescription  "Version 3 of the model finalized June 2014"^^xsd:string
Individual: c_dateOfMappingCreation
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasDescription  "links a creation date to a mapping"^^xsd:string,
 hasHumanReadableName  "relation c_dateOfMappingCreation"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_dateOfMappingCreation"^^xsd:string
Individual: Trinity_Library_RSS
Types:
ChangeNotification
Facts:
 belongs_to_step  Mapping_Management_Step,
 has_source  Trinity_Library_RSS_source,
 hasHumanReadableName  "RSS feed of the Trinity Library"^^xsd:string,
 hasDescription  "New feed of the Trinity Library"^^xsd:string,
 hasUniqueIdentifier  ""^^xsd:string,
 has_editing_priority  "1"^^xsd:int
Individual: c_has_formal_language
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasHumanReadableName  "relation c_has_formal_language "^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_has_formal_language"^^xsd:string,
 hasDescription  "links representation to formal language it is defined in"^^xsd:string
Individual: c_Mapping_Targets
Types:
Mapping_Targets,
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasDescription  "Represents the target in a mapping process"^^xsd:string,
 hasHumanReadableName  "Targets of the Mapping process"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_Mapping_Targets"^^xsd:string
Individual: Mapping_Step
Annotations:
rdfs:comment  "Information  how  mappjng  candidates  have  been  identified  between  target  and  source
      ontology"^^xsd:string
Types:
Lifecycle_Process_Steps
Facts:
 is_Step  Step_3,
 hasHumanReadableName  "Mapping Proces step"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#Mapping_Process_Step"^^xsd:string,
 hasDescription  "Mapping Process step of the ontology mapping lifecycle"^^xsd:string
Individual: Trinity_Library_RSS_source
Types:
Source
Facts:
 belongs_to_step  Mapping_Management_Step,
 lastCheckedDateAndTime  "2014-05-06T19:21:34"^^xsd:dateTime,
 hasDescription  "Trinity Library RSS"^^xsd:string,
 has_editing_priority  "1"^^xsd:int,
 urlValue  "http://trinituCollege.library.ie/newfeed"^^xsd:string,
 hasHumanReadableName  "Trinity Library RSS"^^xsd:string,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#Trinity_Library_RSS_source"^^xsd:string
Individual: Matching_process_between_books_and_library_ontology
Types:
Matching_Process,
```

```
Class_Representator
Facts:
 has_process_step  Step_2_matching_process,
 has_process_step  Step_1_matching_process,
 belongs_to_step  Matching_Step,
 has_editing_priority  "50"^^xsd:int,
 hasDescription     "Matching   process   used   to   created   matcings   for   the   lib   and   book
      ontology"^^xsd:string,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#Matching_between_books_and_library_ontology"^^xsd:str
      ing,
 hasHumanReadableName    "Matching   process   used   to   created   matcings   for   the   lib   and   book
      ontology"^^xsd:string
Individual: Matching_Requirments_83
Types:
Matching_Requirments
Facts:
 belongs_to_step  Matching_Step,
 belongs_to_process_step  Step_1_matching_process,
 textValue  "Matching must be fast"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_Matching_Requirments"^^xsd:string,
 has_editing_priority  "1"^^xsd:int,
 hasDescription  "Represents matching requirements"^^xsd:string,
 hasHumanReadableName  "Matching Requirements"^^xsd:string
Individual: c_Ontolooy_Mapping_Lifecycle_Entities
Types:
Ontolooy_Mapping_Lifecycle_Entities
Facts:
 has_version  OM2R_version_3,
 has_editing_priority  "1"^^xsd:int,
 hasHumanReadableName  "Ontolooy_Mapping_Lifecycle_Entities"@en,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#c_Ontolooy_Mapping_Lifecycle_Entities"^^xsd:anyURI,
 hasDescription   "Represent  the  relevant  the  ontologies  and  mappings  which  are  addressed  in  the
      lifecycel"^^xsd:string
Individual: Identification_Step
Annotations:
rdfs:comment "Information about the ontologies which are addressed in this mapping"^^xsd:string
Types:
Lifecycle_Process_Steps
Facts:
 is_Step  Step_1,
 hasDescription  "Identification Process step of the ontology mapping lifecycle"^^xsd:string,
 hasHumanReadableName  "Identificatinon Process Step"^^xsd:string,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#Identification_Process_Step"^^xsd:string
Individual: Element_rdfs_class
Types:
Element_the_matching_is_based_on
Facts:
 belongs_to_process_step  Step_2_matching_process,
 hasHumanReadableName  ""^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#Element_rdfs_class"^^xsd:string,
 hasDescription  ""^^xsd:string
Individual: c_Element_the_matching_is_based_on
Types:
Element_the_matching_is_based_on,
Class_Representator
Facts:
 belongs_to_step  Matching_Step,
 has_version  OM2R_version_3,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#c_Element_the_matching_is_based_on"^^xsd:string,
 hasHumanReadableName  "The Element the Matching Process is based on"^^xsd:string,
 has_editing_priority  "1"^^xsd:int,
 hasDescription  "Represens the elements the matching algorithm is based on"^^xsd:string
Individual: Trinity_Searcher_Source
Types:
Source
Facts:
 belongs_to_step  Mapping_Management_Step,
 hasHumanReadableName  "Trinity Library Searcher"^^xsd:string,
 lastCheckedDateAndTime  "2014-05-06T15:57:02"^^xsd:dateTime,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#Trinity_Searcher"^^xsd:string,
 urlValue  "www.trinity.org/search"^^xsd:string
Individual: c_formal_languages
Types:
Class_Representator,
Formal_Language
Facts:
 belongs_to_step  Identification_Step,
```

```
 has_version  OM2R_version_3,
 hasHumanReadableName  "Formal Language"^^xsd:string,
 has_editing_priority  "200"^^xsd:int,
 hasDescription  "Formal Language"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_formal_languages"^^xsd:string
Individual: c_Mapping_Details
Types:
Class_Representator,
Mapping_Details
Facts:
 has_version  OM2R_version_3,
 belongs_to_step  Mapping_Step,
 has_editing_priority  "4"^^xsd:int,
 hasHumanReadableName  "Mapping process"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#c_Mapping_Details"^^xsd:string,
 hasDescription  "Represent the mapping process"^^xsd:string
Individual: StrucSubsDistAlignment_Source
Types:
Source
Facts:
 belongs_to_step  Matching_Step,
 hasDescription  "Url to get the code for the Struc Subs Dist Alignment"^^xsd:string,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#StrucSubsDistAlignment_Source"^^xsd:string,
 hasHumanReadableName  "Url to get the code for the Struc Subs Dist Alignment"^^xsd:string,
 lastCheckedDateAndTime  "2011-09-27T15:33:55"^^xsd:dateTime,
 has_editing_priority  "1"^^xsd:int,
 urlValue  "http://www.modellmapping.org/om2rdemo.owl#StrucSubsDistAlignment_Source"^^xsd:string
Individual: Partial_Match
Annotations:
rdfs:comment ""^^xsd:string
Types:
Matching_Scope
Facts:
 belongs_to_step  Matching_Step,
 has_editing_priority  "1"^^xsd:int,
 hasDescription   "Partial matching which means not all ontology elements are addressed in the
      matching"^^xsd:string,
 hasHumanReadableName  "Partical Matching"^^xsd:string,
 hasUniqueIdentifier  "http://www.modellmapping.org/om2rdemo.owl#Partial_Match"^^xsd:string
Individual: c_belongs_to_process_step
Types:
Class_Representator
Facts:
 has_version  OM2R_version_3,
 hasUniqueIdentifier
      "http://www.modellmapping.org/om2rdemo.owl#c_belongs_to_process_step"^^xsd:string,
 hasHumanReadableName  "relation_belongs_to_process_step"^^xsd:string,
 hasDescription  "Relation of a process step to a process"^^xsd:string
```