Christophe Debruyne, Brian Walshe, and Declan O'Sullivan. Towards a project centric metadata model and lifecycle for ontology mapping governance. In Gabriele Anderst-Kotsis and Maria Indrawan-Santiago, editors, Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services, Brussels, Belgium, December 11-13, 2015, pages 50:1–50:10. ACM, 2015

Towards a Metadata Model and Lifecycle for Ontology Mapping Governance

Christophe Debruyne
Trinity College Dublin
Dublin 2, Ireland
debruync@scss.tcd.ie

Brian Walshe
Trinity College Dublin
Dublin 2, Ireland
bwalshe@scss.tcd.ie

Declan O'Sullivan
Trinity College Dublin
Dublin 2, Ireland
declan.osullivan@scss.tcd.ie

ABSTRACT

Ontology matching and mapping is concerned with discovering correspondences between two ontologies to a mapping that enable applications to relate, interlink or integrate data. The construction of such mappings is not trivial as they are created to serve a purpose and result from collaboration. Current ontology-mapping metadata formats only capture a glimpse of the mapping construction process and focus on the exchange of mappings and some limited properties to facilitate reuse and discovery. For mapping governance to be possible, we argue that a suitable metadata model âĂŞ presented in this paper âĂŞ needs to capture all aspects from the ontology mapping lifecycle; from the inception of a project to the execution of these mappings. This allows one to formulate queries that not only would facilitate the discovery and reuse, but also queries that allow one to govern the ontology mapping projects and render the construction processes more transparent and traceable.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; K.4.3 [**Computers and Society**]: Organizational Impacts—*Computer-supported collaborative work*

General Terms

Management, Documentation

Keywords

Ontology Alignment, Provenance, Metadata Management

1. INTRODUCTION

Ontologies are commonly defined as "a [formal] explicit specification of a [shared] conceptualization" [8] and constitute the key resources for realizing a Semantic Web [1]. They provide a vocabulary or description of a domain of interest in some formalism – on the Semantic Web often RDF Schema or the Web Ontology Language (OWL) – that computerized agents are able to process. These ontologies are developed to meet certain semantic interoperability requirements ranging from integration, to publication of data on the Web as Linked Data. Since ontologies are developed for particular purposes in domains of interests that may overlap, one faces the problem of semantic heterogeneity: different representations of the same or similar concepts, relations and instances. On the Linked Data Web, for instance, one may observe that both DBpedia and the LMDB share concepts such as "actor" and "director", albeit represented differently.

Euzenat and Shvaiko stated that this problem is typically tackled in two steps: detecting the correspondences between the different ontologies and interpreting these correspondences to create an executable mapping with respect to the application needs (e.g., data translation vs. query answering) [19]. The field of ontology matching has been around for over a decade and many important contributions have been made. Given the field's maturity, Shvaiko and Euzenat – considered authorities in the domain of ontology matching – have reflected on the question as to whether the field still has challenges to overcome. In [19], for instance, they formulated several problems that still need to be addressed ranging from large scale evaluation techniques and efficiency of matching algorithms, to user involvement and the socio-technical aspects of ontology matching.

One of the challenges they formulated is concerned with the infrastructure and support for alignment management [19]: "[t]he challenge 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."

It is interesting in the above quote to observe that management activities are artifact-centric (i.e., focused around the alignments) and only concerned with storing and sharing. Indeed, as we will show in the following section, related work is often limited to (annotations of) the produced artifacts. We argue that proper ontology mapping infrastructures should take a "project" centric view and generate metadata and artifacts as the ontology mapping project progresses (from the requirements specification of such a project and discovery of reusable alignments to the creation of an alignment and an executable mapping thereof), as each subsequent phase in the ontology mapping lifecycle is the result of agreements between the community of stakeholders for which this mapping is intended. A project-centric view would allow one to formulate questions such as "What were the most debated aspects in a particular ontology mapping project?", as we would have access to the interactions that led to these decisions.

In this paper, we propose to view the creation of ontology mappings as a data governance activity, which implies that management processes such as checking the status or the integrity of ontology mappings should be supported. As ontology mappings are – and should be – published on the (Linked Data) Web as RDF, we are challenged with an interesting tension field between treating certain aspects of creating ontology mappings as integrity constraints in our governance setting and the Open World Assumption adopted by the Web Ontology Language (OWL). One example is requiring properties to be total, which is difficult to simulate in OWL. Our proposal consists of an ontology capturing the different stages of an ontology mapping's lifecycle and the methods adopted for quality assurance in our setting.

The remainder of this paper is organized as follows: Section 2 provides an overview of related work on ontology metadata vocabularies and ontology mapping lifecycles; Section 2 also presents our observation that a project-centric view on ontology mapping construction would allow one to capture in more detail the different aspects of ontology mapping construction and would also enable us to support ontology mapping governance as one would be able to ask questions about the project and not only about the generated artifacts; Section 3 presents our refined ontology mapping lifecycle and ontology mapping metadata vocabulary to support ontology mapping governance; Section 4 sheds some light on the tension field created by adopting Semantic Web technologies (and the assumptions that go with these) and the constraints (or governance rules) that ontology mapping projects should âĂŞ in our opinion âĂŞ comply with; Section 5 then provides some insights on the ontology engineering process; and we conclude the contributions of this paper and present our future directions in Section 6.

2. RELATED WORK

2.1 Ontology Mappinng Metadata

Metadata is key in allowing users and systems to select adequate alignments based on several criteria. Several metadata vocabularies (or ontologies) have been proposed in literature.

The Alignment Format (AF) [5], for instance, proposed a simple vocabulary to represent simple correspondences – correspondences between entities – between two ontologies. One can capture the ontologies that were matched, their formats, the matching method adopted, etc. To capture more detail, however, one has to extend the ontology. An example of such an extension is the Expressive and Declarative Ontology Alignment Language (EDOAL) [2] provides an expressive and declarative ontology alignment language based on AF. With EDOAL, one is also able to represent complex correspondences.

Both AF and EDOAL are used to represent alignments, which one can use to render executable mappings in, for instance, SWRL. All metadata connected to these artefacts do provide documentation that aids in discovery and reuse, but does not reveal much about the processes that lead to the creation of these alignments. As Thomas et al. noted, they are focussed on the representation of alignment correspondences and not on the creation and management related metadata [24]. These authors also observed that the retrieval and efficiency of ontoloyg mappng discovery can be improved by using a semantic moel of metadata descring the ontology, the mapping features and lifecycle information [23]. They thus proposed OM2R [24], a metadata model for ontology mapping reuse with predicates to represent the different stages in the ontology mapping lifecycle.

Though OM2R does make the distinction between management processes (and related metadata) and the resulting artefacts – mappings can be represented as AF – the metadata mostly focuses on the choice of the ontologies, the creation of candidate correspondences (and the associated tools) and the creation of "confirmed mappings", which are the selected and refined candidate correspondences that will constitute the alignment. Though OM2R provides quite a few predicate that lack in current representation formats such as scope and requirements of an ontology matching project, it has a fairly simplistic representation of the activities and generated artefacts. The artefacts such as scopes and requirements are futhermore captured "as is", and the whole process of reaching the necessary agreements that lead to these artefacts and to subsequent stages in the mapping process are lost.

We argue that in order to formulate richer queries such as the requirements that were the most difficult to agree upon or the most discussed correspondences, we should treat these artefacts as first class citizens that are generated and used by these activities. All these activities are furthmore informed by agents, both human as well as software. Moreover, the interactions between the human stakeholders provide valuable information on how and why decisions were made – e.g., as a discussion thread – in an adequate way such that aforementioned queries are possible to formulate.

2.2 Ontology Mapping Lifecycles

Very few ontology mapping lifecycle models are proposed in literature and those that do focus on the matching and mapping processes. The whole mapping process is mostly artifact-centric; focusing on the alignments (represented for instance with EDOAL) and their renderings in, for instance, SWRL.

The OISIN [16] process prescribed the processes and decisions to be made for creating ontology mappings. The process starts with the discovery of ontologies for the creation of a new mapping. It stops with the management phase to contain processes for sharing, altering and integrating mappings. As the model is focused on the creation of new mappings, the (partial) reuse of existing alignments are not explicit in the model. The model is furthermore not iterative; the use of a rendered mapping that might inform the stakeholder does not trigger any processes. This feedback loop has usually been taken into account in ontology engineering methodologies such as OTKM [21].

OISIN was extended to incorporate this feedback loop to support ontology mapping reuse in [23]. Having most likely drawn inspiration from ontology engineering, the authors furthermore proposed a meta-process to capture the details of an ontology mapping's lifecycle. The work proposed in their study has a few limitations. First, it – again – considers reuse, sharing and publication as separate management activities not integrated in a broader workflow. It also starts from the discovery of ontologies that need to be mapped and is artifact centric; there are ontologies, alignments or mappings and the metadata capturing details of a mapping's lifecycle. Metadata captured in the latter is focused on rendering the discovery and reuse of mappings more efficient.

The "context" of an mapping, or better yet, the context of creating that mapping provides valuable information to better interpret the correspondences in alignments [4]. Both OISIN and its extension treated do not specifically treat the context of an ontology mapping (the purpose, scope, requirements, etc.). Appropriate lifecycle models should thus encorporate activities for an ontology mapping project's inception stages (amongst others). Treating the creation of an ontology mapping as a project, one can again draw inspiration from existing ontology engineering methods. By taking a project-centric stance, we can perceive an mapping engineering project as generating alignments, mappings, etc., but also as generating scopes, requirements that will be used as guides and even as evaluation criteria. From a conceptual point of view, it is more sensible to formalize "requirement part of mapping project" and "mapping project resulting in mapping" instead of "requirement part of mapping".

In order to propose a ontology mapping governance framework, we argue that a project-centric view must be adopted that takes into account all phases of such a project and in which all the activities by the stakeholders are adequately recorded such that the ontology mapping creation process is traceable and transparant. The necessiates both a refined lifecycle model and vocaculary, which we will present in the next section.

3. OVERVIEW

In the previous section, we argued one has to adopt a project-centric perspective to capture all stages of an ontology mapping's construction and capture the interactions between all the stakeholders that lead to a rendered mapping. Such information would allow one not only to formulate queries to look for existing mappings, but also to formulate questions such as "What are the most debated correspondences?" or "Which requirements have not been met yet in an ontology mapping project?"

In this section we first propose a novel ontology mapping lifecycle that takes into account all stages of such a project. This model will actually prescribe the different stages of a project. We then proceed with presenting a vocabulary – based on the provenance ontology – to capture the different stages. Though the adoption of ontology languages allows one to use SPARQL to formulate these questions and even enables some lightweight reasoning, we will elaborate on how we deal with a closer world assumption for ontology mapping governance activities and the open world assumption adopted on the Semantic Web.

3.1 Ontology Mapping Lifecycle

Our ontology mapping lifecycle is based on the work presented in [16, 23], but takes a project-centric view that allows us to encorporate reuse activities as part of a project's workflow. It furthermore encorporates activities related to project initiation and adopts the terminology adopted by Euzenat and Shaiko, in particular the different between ontology alignment and mapping. Our lifecycle is depicted in Figure 1 and will be described below.

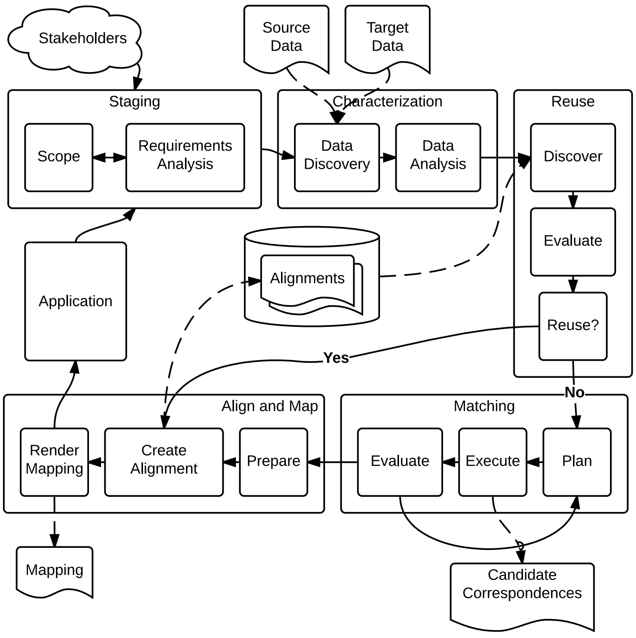

Figure 1: Ontology Mapping Lifecycle

Stage. Ontology mappings are built for a particular purpose – be it a data integration task or a Linked Data mashup – and therefore a community of stakeholders. We do assume that most ontology mapping project involve more than one stakeholder. More tech savvy people configuring the matchers and domain experts evaluating, from a conceptual point of view, the generated correspondences is a very plausible scenario. Staging involves identifying the community of stakeholders, agreeing on the scope of the project and listing the requirements. Both scope and requirements will be used to guide and assess the (intermediate) outcomes of the project. Requirements are particularly important as an adequate governance platform should at least encourage the stakeholders to reflect on and report on how the outcomes have met these requirements.

Characterize. Though they might be known beforehand, characterization is concerned with the discovery and analysis of ontologies to be mapped. As in [23], the objective of characterization is "to analyze the addressed ontologies with their respect to their amenability for mappings in order to identify difficulties that may be involved in underatking a mapping." The outcome of characterization are traces of the discovery ana analysis activities as well as a selection of a source and target ontology to be mapped.

Reuse. Given our project-centric view on the creation of ontology mappings, we can now integrate reuse activities as part of the workflow. Given our selection from the characterization activity, one should look and assess existing alignments. Requirements will be a valuable tool to assess the ontologies and these should be explicitely captured whenever applicable. There are three outcomes: i) an existing alignment is fit for immediate reuse and can be rendered into a mapping for the project; ii) an existing alignment is not immediately fit for reuse, but can serve as a basis for a new alignment; iii) no existing alignment can be reused and one should be made from scratch. We consider the third case as a special case of the second as it can be regarded as the creation of a new alignemnt starting from an empty base.

Match. Match is concerned with the planning of, execution and evaluation of the applying ontology matchers. Planning is concerned on agreeing on a set of matcher configurations; a combination of a matcher, parameters and additional resources to be provided as input. Examples of such resources are the alignments considered as a basis to create a new alignment. When the community agrees on a set of matcher configuration, the candidate correspondences are generated. Candidate correspondences are a set of correspondences that do not yet constitute an alignment. Candidate correspondences are first evaluated for their fitness before an alignment is created.

Align and Map. The activities in this phase of the lifecycle are concerned with the creation of an alignment and a mapping. Euzenat and Shvaiko's defines an alignment as [6]: "a set of correspondences between two [...] ontologies [...]. The alignment is the output of the matching process." Though matching is defined as the process of finding relationships or correspondences between entities of different ontologies and could thus imply fully automatic, manual or semi-automatic processes, we make a distinction between the correspondences created by applying matchers and the manual refinement of these correspondences for the creation of a suitable alignment as producing two different artifacts. Planning entails agreeing on how the candidate correspondences have to be amended to create an alignment fit for this project's purpose. Examples could be agreeing on removing all correspondences related to certain concepts, as they are not within the scope of the project or defining a strategy for amending the relationships that hold between entities. Based on the requirements, the community also agrees on the formalism that should be adopted for rendering the mapping. The alignment, or versions thereof, is then stored in an alignment repository. The mapping, which as a directed version of an alignment [6], is the artefact that software agents can execute.

Application. The mappings rendered in the previous phase are to be used to support the application foreseen by the community of stakeholders. The application of the rendered mapping might provide new insights to the community or even break over time, e.g., when the structure of data sources changes. Both the application of a mapping and external factors that influence the application trigger not interactions within the community; should a new (version of) a mapping be created? Some of these activities can be supported in a semi-automatic manner, such as monitoring changes in ontologies to start new discussions.

3.2 Metadata Model

The creation of metadata that documents the complete ontology mapping lifecycle is key to enable reuse [23]. We already noted that the state of the art is mostly concerned with the resulting artifacts (i.e., the alignments and mappings). The creation of an ontology mapping is, however, a goal-oriented, collaborative undertaking. This undertaking not only results in a aforementioned artifacts, but also in a community of stakeholders and several other documents that are used throughout the process such as the project scope and a list of requirements. In this section, we will present or metadata model that will capture these aspects, allowing one to not only support queries to facilitate reuse, but also to formulate queries to govern the ontology mapping construction processes.

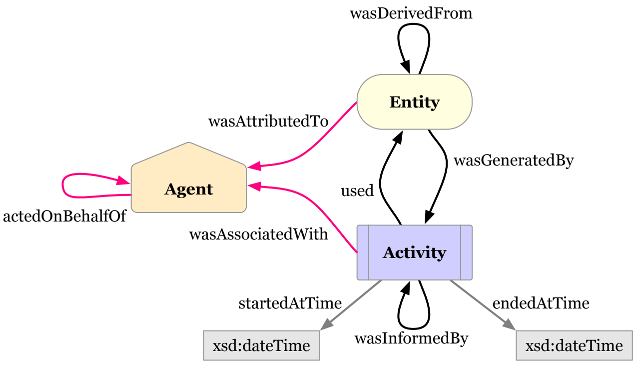

Figure 2: Core concepts and relations in PROV-O. Image from [15].

The metadata model we developed – available as an OWL2 ontology – extends the PROV-O [15], a W3C Recommendation for representing and exchanging provenance information as RDF. Provenance information provides insights on a resource's origin, wuch as the who created that resource, whent it was modified or how it was created [26]. Provenance, as stated in [10], is key in evaluating the quality and trusting information on the Web. PROV-O's core concepts and relations (shown in Figure 2) provide a good starting point for describing the activities and intermeidate artefacts that lead to the realization of an ontology mapping:

We will now present our metadata model. We start of by describing a concept to represent ontology mapping projects and continue with the activities and entities that relate to each of the lifecycles represented in Figure 1. We assume that most provenance information that can be captured with PROV-O are a given; i.e., we will not explicitely state all aspects of provenance of provenance such as start and end dates and times. Concepts and relations in our ontology will be prefixed with "gm".

3.2.1 Representing Ontology Mapping Projects

We introduced the concept Project to represent ontology mapping projects. According to PROV-O, a prov:Plan is "an entity that represents a set of actions or steps intended by one or more agents to achieve some goals". Since we have described the phases and activities associated with such a project, prov:Plan is a suitable supertype for Project. The "implementation" of prov:Plan is, however, left to be extended by applications [15]. We therefore associate our Project with the different activities that project go through.

Figure 3 depicts our Project and its relation with sub-

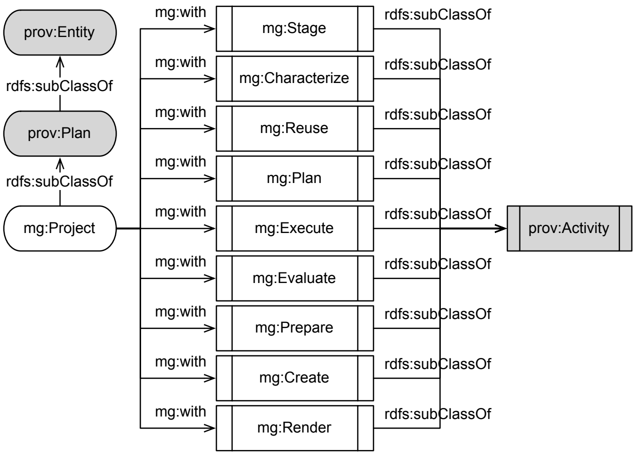

Figure 3: Project and the related activities.

classes of prov:Activity related to the creation of ontology mappings. Note that some of the activities correspond to phases in our lifecycle (such as staging), but others are more fine grained (such as mg:Plan, mg:Execute and mg:Evaluate that correspond with the execution phase.

The ontology developed for mapping governance is kept pretty lightweight; role and type hierarchies, domain and range declarations, and so forth. Since OWL adopts the OWA, some of our constraints have to be covered by the collaborative environment (see Section 4). In this environment, we need:

- Each activity to be related with exactly one Project.
- If an project is related with an instance of one of the later activities, then that project must be related with at least one instance of the prior activities. For example, if a mg:Project is mg:with a mg:Characterize, then that mg:Project must also be mg:with a mg:Stage.
- Some activities, such as the planification of execution tasks, can be informed by different activities. Those conditions will be described in the following subsections.

3.2.2 Stage

The activity mg:Stage – depicted in Figure 4 – is the first activity in an ontology mapping lifecycle where the community of stakeholders decides on both the scope of the project and its requirements. Both the scope and the requirements will be used in subsequent stages and activities to validate (intermediate) results. The Stage activity produced three types of artefact: a scope, a community and requirements. All three artefacts are disjoint subtypes of prov:Entity. A mg:Requirement is a documented representation of a condition that must be met or possessed by the alignment or rendered mapping to solve a problem or achieve an objective.[1] Each Stage must result in at least one requirement, and each requirement belongs to exactly one instance of Staging. The mg:Scope represents the features that will characterize the alignment and rendered mapping. Every staging activity results in exactly one scope and each instance of Scope is related to at most one instance of Staging. A mg:Community represents the group of human agents (represented as instances of prov:Person) who will contribute to the development of an ontology alignment and mapping. At the end of this activity, the community is defined. Each community has at least one prov:Person related to it. In order for an ontology mapping project to be collaborative, however, that number must be greater.

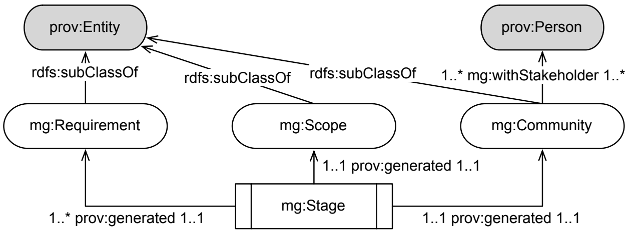

Figure 4: Modelling mg:Stage.

3.2.3 Characterize

The activity mg:Characterize (see Figure 5) is concerned with the discovery of suitable data sources, if not known beforehand, and the assessment of these data sources for their amenability for being mapped. As stated by [23], the goal of the latter is to identify difficulties that may be involved in undertaking the mapping, e.g., based on the form and quality of the ontology. Characterize uses the requirements that were generated in a mg:Stage. All requirements have to be used – in one way or another – by the end of mg:Characterize. This could be during discovery or evaluation. One way of doing so is explicitly referring to the requirements in a discussion. So, given a mg:Project p, the set of mg:Requirement generated by the mg:Stage of p must be equal to the set of mg:Requirement used by the mg:Characterize of p. Characterize furthermore considers two or more ontologies, which are captured with a special predicate that is a subproperty of prov:used. At the end of Characterize, four artefacts are generated (all of which are disjoint subclasses of prov:Entity):

- The mg:DiscoveryReport that summarized – in some way – the process of looking for suitable ontologies to be mapped.
- The mg:OntologyAnalysis, in which the process of assessing and analysing the considered ontologies are – in some way – summarized.
- The mg:Selection containing the two ontologies that will be mapped. The source and target ontologies off course have to be present as considered ontologies of the Characterization activity.
- The mg:MappingDecision, which captures whether or not the stakeholder group decides to go forth with the mapping project.

---

[1] This definition is based on the IEEE Glossary of Software Engineering Technology's definition of "software requirement".

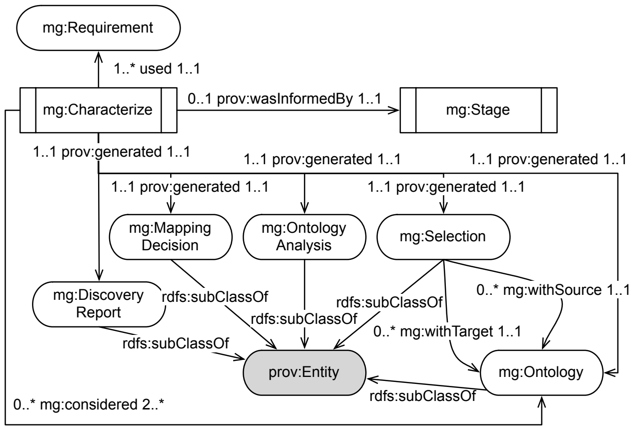

Figure 5: Modelling mg:Characterize.

3.2.4 Reuse

Reuse is the activity of discovering reusable alignments based on the requirements, the source ontology and the target ontology. There are three possible outcomes: either no reusable alignment is found, an alignment is found that could serve as the basis for creating a new alignment, or an alignment can be reused in its current form. Depending on the project, however, the alignment to be reused might need to be refined and rendered in a different format. Note that in the lifecycle, the first two possibilities are captured by the "no". When going from the reuse decision to the planification of the matching task; one can reuse a set of correspondences in an existing mapping as initial input. In the case that no reusable mapping is found, that set is considered to be empty. The reuse activity is informed by the characterization activity and uses the ontology analysis to look for reusable mappings found in alignment repositories. Looking for reusable alignments can be supported with Alignment Search Engines. This activity finally results in the creation of a Reuse Evaluation and a Reuse Decision.

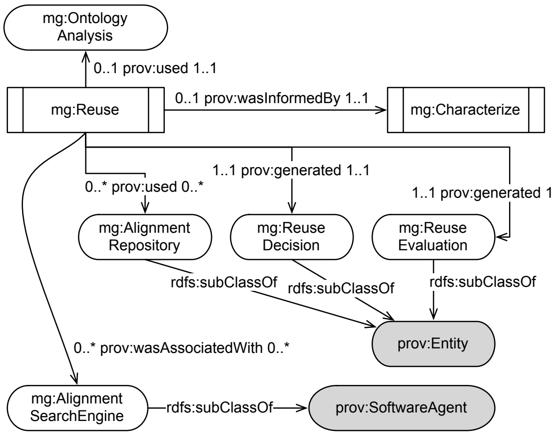

Figure 6: Modelling mg:Reuse.

Important to note is that the reuse activity may rely on many alignment repositories while searching for reusable ontology alignments. The reuse activity does not have to rely on such repositories, as they may not be available. Similarly reuse may rely on one or more alignment search engines. Important, however, is that by the time an ontology mapping project reaches the end of the reuse activity, the ontology analysis used by that activity is the one produced by the characterization activity of that same project.

3.2.5 Plan, Execute and Evaluate

Plan is the activity of planning the matching execution. In this activity, the configuration of the matchers are discussed taking into account the requirements and the ontology analysis. Planning is preceded by the reuse or matching evaluation activities and results in one or more matcher configurations, as depicted in Figure 7. Again, the requirements used in this activity should be the same as the requirements generated by the mg:Stage activity of that same project. Similarly, the reuse evaluation and ontology analysis should be the same of the ontology mapping project.

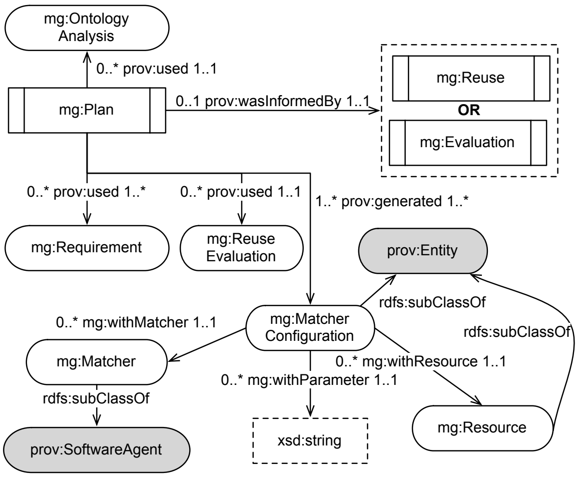

Figure 7: Modelling mg:Plan.

Both instances of mg:Reuse and mg:Evaluate activities can inform at most one plan activity, and plan activities have to be informed by one. We will later see that mg:Evaluate will either inform the the start of creating an alignment or a new planning activity. Matcher configurations can be reused by Plan activities. They are identified by the combination of a matcher, parameter and resource. This actually means that there should be no two matcher configurations with different URIs having the same combinations of aforementioned parameters.

Execution (see Figure 8) is the activity of applying the matchers and their configurations to the two ontologies for creating a set of candidate correspondences. The activity should use the matchers and configurations that have been decided upon in the planning activity. The ontologies used by this activity have to be to the source and target ontologies of the selection. Candidate correspondences are a representation of the set of relations holding, or supposed to hold according to a (combination of) particular matching algorithm(s), between entities of different ontologies. Candidate correspondences can be represented using EDOAL (see Section 2.1). The set of matchers associated with this activity should be equal to the set of matchers referred to the matcher configurations used by that activity.

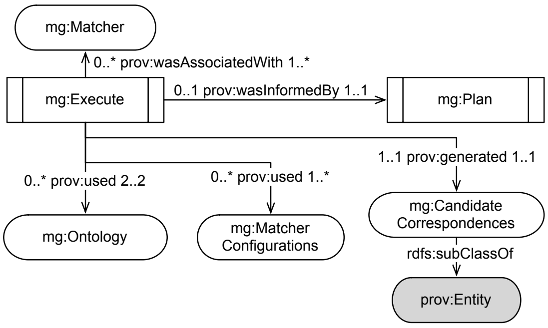

Figure 8: Modelling mg:Execute.

Evaluate is the process of reflecting on and evaluating the generated candidate correspondences and decide whether to proceed with the creation of an alignment or get back to planning the execution. The evaluation activity uses the candidate correspondences that have been generated by the execution activity preceding it.

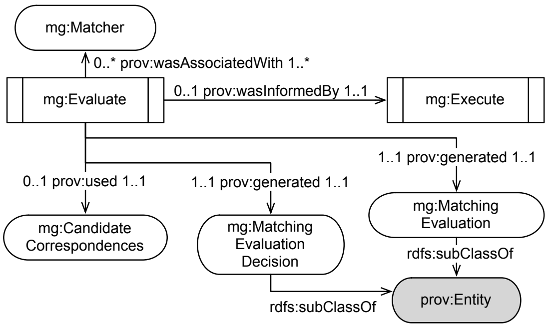

Figure 9: Modelling mg:Evaluate.

3.2.6 *Prepare, Create and Render*

Prepare is the activity concerned with planning the creation of an alignment and the generation (or "rendering") of a mapping. During preparation, the community of stakeholders create an Alignment Plan. This plan contains the actions that need to be undertaken during the creation of the alignment. Examples could be manual refinement.

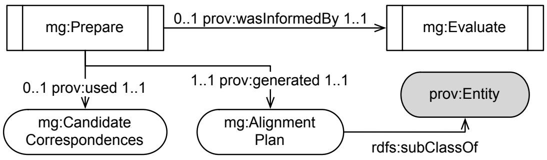

Figure 10: Modelling mg:Prepare.

The process of creating an alignment based on the candidate correspondences that were deemed to be an adequate starting point. Using the alignment plan created in the Prepare activity, the community of stakeholders select, refine and create – where necessary – the correspondences that will constitute the alignment.

As stated previously, Euzenat and Shvaiko define an alignment as: "a set of correspondences between two [...] ontologies [...]. The alignment is the output of the matching process." We note that this corresponds with what we call the set of candidate correspondences. In our workflow, the alignment is the result of selecting, refining (a subset of) those candidate correspondences that will constitute an alignment. There is thus a human "intervention". This fits with our broader aim of regarding the problem of ontology mapping from the perspective of a project rather than from the perspective of an artifact.

The create activity relies on the candidate correspondences and alignment plan generated by the Execute and Prepare activities of the same ontology mapping process that immediately precede that create activity.

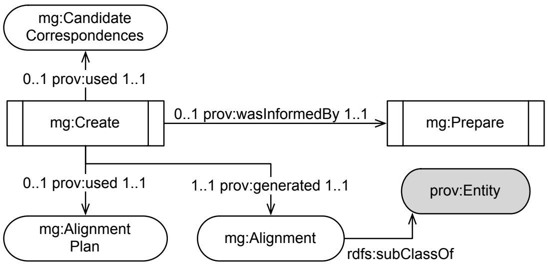

Figure 11: Modelling mg:Create.

Render is the activity of creating a mapping from an ontology alignment. A mapping is the oriented, or directed, version of an alignment: it maps the entities of one ontology to at most one entity of another ontology [6]. The creation of such a mapping should be fairly straightforward and result in a machine executable artifact fit for the purpose defined by scope and requirements created in the staging activity.

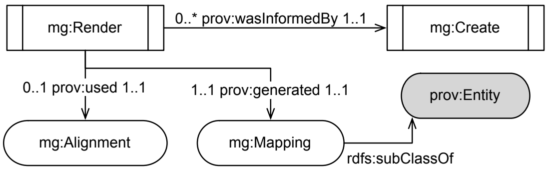

Figure 12: Modelling mg:Render.

One can notice that the same gm:Create activity can trigger zero or more gm:Render activities. This is to accommodate the need to render mappings with different formalisms using the same alignment. An instance of mapping has to be related to at least one instance of rendering. When reusing alignments in their interiority, a community can wish to render the same mapping.

4. MAPPING GOVERNANCE

According to Wikipedia, data governance is defined as[2]: “a control that ensures that the data entry by an operations team member or by an automated process meets precise standards, such as a business rule, a data definition and data integrity constraints in the data model. The data governor uses data quality monitoring against production data to communicate errors in data back to operational team members, or to the technical support team, for corrective action.”

Though we have proposed an ontology for capturing the ontology mapping construction process that facilitates the publication of that information as Linked Data on the Web, we are now faced with an interesting tension field: the Open World Assumption adopted on the Semantic Web and integrity constraints that mapping projects have to comply with in order to ensure a certain quality.

For example. all our activities must be related to a project via mg:with. Each instance of mg:Stage is associated with exactly one instance of mg:Project. One could assume that providing the following general axiom could solve this problem: $Stage \sqcap \neg(= 1 with^{-}.Project) \sqsubseteq \bot$. This axiom, however, will not work because of the OWA. The fact that a certain instance of mg:Stage has no explicit relationship with a mg:Project in a knowledge base does not mean that that relationship does not exist. One also needs to be careful not to create axioms that would lead to the cration of unnamed individuals (i.e., blank nodes). Similarly, the constraint that a mg:Project can only be related to a mg:Charaterize if that same mg:Project is already related to a mg:Stage cannot be modelled with general axioms such as: $\exists with.Characterize \sqsubseteq \exists with.Stage \sqsubseteq \bot$.

To tackle this problem, we adopt an approach inspired by [22]; formulating our integrity constraints that should hold from a governance constraint as a set of SPARQL ASK and SELECT queries. ASK queries are used to ask the system whether there are project violating certain conditions. SELECT queries are used to list the projects that have certain issues that need to be fixed. This set of queries can then be applied to other knowledge bases adopting the ontology. By applying an OWL 2 reasoner on a knowledge base, we can use the following two SPARQL queries to detect project that violate aforementioned constraints (namespaces omitted):

```
SELECT ?stage WHERE {
  ?s a mg:Stage.
  NOT EXISTS { ?project mg:with ?s. }
}

SELECT ?project WHERE {
  ?p a mg:Project.
  ?p mg:with ?c.
  ?c a mg:Characterize.
  NOT EXISTS { ?p mg:with ?s. ?s a mg:Stage. }
}
```

5. APPLICATION

Proofs of concepts – queries?

6. DISCUSSION

Plan-P [7]
[12]
[25]
[13] [17]
[9]
[14] [3]
[11]
[20]
[18]

7. CONCLUSIONS AND FUTURE WORK

Background
Objective
Methods
Results
Conclusions

Acknowledgements

This study is supported by the Science Foundation Ireland (Grant 13/RC/2106) as part of the ADAPT Centre for Digital Content Platform Research (`http://www.adaptcentre.ie/`) at Trinity College Dublin.

8. REFERENCES

[1] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, 2001.

[2] J. David, J. Euzenat, F. Scharffe, and C. T. dos Santos. The alignment API 4.0. *Semantic Web*, 2(1):3–10, 2011.

[3] C. Debruyne, T. Tran, and R. Meersman. Grounding ontologies with social processes and natural language. *J. Data Semantics*, 2(2-3):89–118, 2013.

[4] J. C. dos Reis, C. Pruski, and C. Reynaud-Delaître. State-of-the-art on mapping maintenance and challenges towards a fully automatic approach. *Expert Syst. Appl.*, 42(3):1465–1478, 2015.

[5] J. Euzenat. An API for ontology alignment. In S. A. McIlraith, D. Plexousakis, and F. van Harmelen, editors, *The Semantic Web - ISWC 2004: Third International Semantic Web Conference,Hiroshima, Japan, November 7-11, 2004. Proceedings*, volume 3298 of *Lecture Notes in Computer Science*, pages 698–712. Springer, 2004.

[6] J. Euzenat and P. Shvaiko. *Ontology matching.* Springer, 2007.

[7] D. Garijo and Y. Gil. Augmenting prov with plans in p-plan: Scientific processes as linked data. In *Second International Workshop on Linked Science: Tackling Big Data (LISC), held in conjunction with the International Semantic Web Conference (ISWC)*, Boston, MA, 2012.

[8] T. R. Gruber. Toward principles for the design of ontologies used for knowledge sharing? *Int. J. Hum.-Comput. Stud.*, 43(5-6):907–928, 1995.

[9] T. A. Halpin. Object-role modeling. In L. Liu and M. T. Özsu, editors, *Encyclopedia of Database Systems*, pages 1941–1946. Springer US, 2009.

[10] O. Hartig and J. Zhao. Publishing and consuming provenance metadata on the web of linked data. In D. L. McGuinness, J. Michaelis, and L. Moreau, editors, *Provenance and Annotation of Data and*

---

[2] `https://en.wikipedia.org/wiki/Data_governance`, last accessed August 4, 2015.

Processes - Third International Provenance and Annotation Workshop, IPAW 2010, Troy, NY, USA, June 15-16, 2010. Revised Selected Papers, volume 6378 of *Lecture Notes in Computer Science*, pages 78–90. Springer, 2010.

[11] M. Hert, G. Reif, and H. C. Gall. A comparison of rdb-to-rdf mapping languages. In C. Ghidini, A. N. Ngomo, S. N. Lindstaedt, and T. Pellegrini, editors, *Proceedings the 7th International Conference on Semantic Systems, I-SEMANTICS 2011, Graz, Austria, September 7-9, 2011*, ACM International Conference Proceeding Series, pages 25–32. ACM, 2011.

[12] V. Khatri and C. V. Brown. Designing data governance. *Commun. ACM*, 53(1):148–152, Jan. 2010.

[13] M. C. A. Klein, D. Fensel, A. Kiryakov, and D. Ognyanov. Ontology versioning and change detection on the web. In A. Gómez-Pérez and V. R. Benjamins, editors, *Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web, 13th International Conference, EKAW 2002, Siguenza, Spain, October 1-4, 2002, Proceedings*, volume 2473 of *Lecture Notes in Computer Science*, pages 197–212. Springer, 2002.

[14] W. Kunz and H. W. J. Rittel. Information science: On the structure of its problems. *Information Storage and Retrieval*, 8(2):95–98, 1972.

[15] D. McGuinness, T. Lebo, and S. Sahoo. PROV-O: The PROV ontology. W3C Recommendation, W3C, Apr. 2013. http://www.w3.org/TR/2013/REC-prov-o-20130430/.

[16] D. O'Sullivan, V. Wade, and D. Lewis. Understanding as we roam. *IEEE Internet Computing*, 11(2):26–33, 2007.

[17] P. Plessers and O. D. Troyer. Ontology change detection using a version log. In Y. Gil, E. Motta, V. R. Benjamins, and M. A. Musen, editors, *The Semantic Web - ISWC 2005, 4th International Semantic Web Conference, ISWC 2005, Galway, Ireland, November 6-10, 2005, Proceedings*, volume 3729 of *Lecture Notes in Computer Science*, pages 578–592. Springer, 2005.

[18] R. Shosha, C. Debruyne, and D. O'Sullivan. Towards an adaptive tool and method for collaborative ontology mapping. In I. Ciuciu, H. Panetto, C. Debruyne, A. Aubry, P. Bollen, R. Valencia-García, A. Mishra, A. Fensel, and F. Ferri, editors, *On the Move to Meaningful Internet Systems: OTM 2015 Workshops*, volume 9416 of *Lecture Notes in Computer Science*, pages 319–328. Springer, 2015.

[19] P. Shvaiko and J. Euzenat. Ontology matching: State of the art and future challenges. *IEEE Trans. Knowl. Data Eng.*, 25(1):158–176, 2013.

[20] M. Solanki. DIO: A pattern for capturing the intents underlying designs. In E. Blomqvist, P. Hitzler, A. Krisnadhi, T. Narock, and M. Solanki, editors, *Proceedings of the 6th Workshop on Ontology and Semantic Web Patterns (WOP 2015) co-located with the 14th International Semantic Web Conference (ISWC 2015), Bethlehem, Pensylvania, USA, October 11, 2015.*, volume 1461 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2015.

[21] Y. Sure, S. Staab, and R. Studer. Ontology engineering methodology. In S. Staab and R. Studer, editors, *Handbook on Ontologies*, International Handbooks on Information Systems, pages 135–152. Springer, 2009.

[22] J. Tao, E. Sirin, J. Bao, and D. L. McGuinness. Extending OWL with integrity constraints. In V. Haarslev, D. Toman, and G. E. Weddell, editors, *Proceedings of the 23rd International Workshop on Description Logics (DL 2010), Waterloo, Ontario, Canada, May 4-7, 2010*, volume 573 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2010.

[23] H. Thomas, R. Brennan, and D. O'Sullivan. Moom - A prototype framework for management of ontology mappings. In *25th IEEE International Conference on Advanced Information Networking and Applications, AINA 2011, Biopolis, Singapore, March 22-25, 2011*, pages 548–555. IEEE Computer Society, 2011.

[24] H. Thomas, R. Brennan, and D. O'Sullivan. Using the OM2R meta-data model for ontology mapping reuse for the ontology alignment challenge - a case study. In P. Shvaiko, J. Euzenat, A. Kementsietsidis, M. Mao, N. F. Noy, and H. Stuckenschmidt, editors, *Proceedings of the 7th International Workshop on Ontology Matching, Boston, MA, USA, November 11, 2012*, volume 946 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2012.

[25] K. Weber, B. Otto, and H. Österle. One size does not fit all—a contingency approach to data governance. *J. Data and Information Quality*, 1(1):4:1–4:27, June 2009.

[26] J. Zhao and O. Hartig. Towards interoperable provenance publication on the linked data web. In C. Bizer, T. Heath, T. Berners-Lee, and M. Hausenblas, editors, *WWW2012 Workshop on Linked Data on the Web, Lyon, France, 16 April, 2012*, volume 937 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2012.