Semantic Web 0 (2021) 1–0
IOS Press

Modular Ontology Modeling

Cogan Shimizu [a] Karl Hammar [b] Pascal Hitzler [a]
[a] *Department of Computer Science, Kansas State University, USA*
[b] *Department of Computer Science, Jönköping University, Sweden*

Abstract. Reusing ontologies for new purposes, or adapting them to new use-cases, is frequently difficult. In our experiences, we have found this to be the case for several reasons: (i) differing representational granularity in ontologies and in use-cases, (ii) lacking conceptual clarity in potentially reusable ontologies, (iii) lack and difficulty of adherence to good modeling principles, and (iv) a lack of reuse emphasis and process support available in ontology engineering tooling. In order to address these concerns, we have developed the Modular Ontology Modeling (MOMo) methodology, and its supporting tooling infrastructure, CoModIDE (the *Comprehensive Modular Ontology IDE* – "commodity"). MOMo builds on the established eXtreme Design methodology, and like it emphasizes modular development and design pattern reuse; but crucially adds the extensive use of graphical schema diagrams, and tooling that support them, as vehicles for knowledge elicitation from experts. In this paper, we present the MOMo workflow in detail, and describe several experiences in executing it. We provide a thorough and rigorous evaluation of CoModIDE in its role of supporting the MOMo methodology's graphical modeling paradigm. We find that CoModIDE significantly improves approachability of such a paradigm, and that it displays a high usability.

Keywords: keywords

1. Introduction

Over the last two decades, ontologies have seen widespread use for a variety of purposes. Some of them, such as the Gene Ontology [1], have found significant use by third parties. However, the majority of ontologies have seen hardly any re-use outside the use cases for which they were originally designed [2].

It behooves us to ask why this is the case, in particular, since the heavy re-use of ontologies was part of the original conception for the Semantic Web field. Indeed, many use cases have high topic overlap, so that a re-use of ontologies on similar topics should, in principle, lower development cost. However, according to our experience, it is often much easier to develop a new ontology from scratch, than it is to try to re-use and adapt an existing ontology. We can observe that this sentiment is likely shared by many others, as the new development of an ontology so often seems to be preferred over adapting an existing one.

We posit, based on our experience, that four of the major issues preventing wide-spread re-use are (i) differing representational granularity, (ii) lack of conceptual clarity in many ontologies, (iii) lack and difficulty of adherence to established good modeling principles, and (iv) lack of re-use emphasis and process support in available ontology engineering tooling. We explain these aspects in more detail in the following. As a remedy for these issues, we propose tool-supported *modularization*, in a specific sense which we also explain in detail.

Representational granularity refers to modeling choices which determine the level of detail to be included in the ontology, and thus in the data (knowledge) graph. As an example, one model may simply refer to temperatures at specific space-time locations. Another model may also record an uncertainty interval. A third model may also record information about the measurement instrument, while a fourth may furthermore record calibration data for said instrument. Another example may be population figures for cities; the values are frequently estimated through the use of statistical models. That is, depending on the data and which statistical model was used, different figures would be calculated.

Note that a fine-grained ontology can be populated with coarse-granularity data; the converse is not true. If a use case requires fine-granularity data, a coarse-grained ontology is essentially useless. On the other hand, using a fine-grained ontology for a use case that requires only coarse granularity data is unwieldy due to (possibly massively) increased size of ontology and data graph.

Even more problematically, is that two use cases may differ in granularity in different ways in different parts of the data, respectively, ontology. That is, the level of abstraction is not uniform across the data. For example, one use case may call for details on the statistical models underlying population data, but not for measurement instruments for temperatures, whereas another use case may only need estimated population figures, but require calibration data for temperature measurements. Essentially, this means that attempting to re-use a traditional ontology may require modifying it in very different ways in different parts of the ontology. An additional complication is that ontologies are traditionally presented as monolithic entities and it is often hard to determine where exactly to apply such a change in granularity.

Conceptual clarity is a rather elusive concept that certainly has a strong subjective component. By this, we mean that an ontology should be designed and presented in such a way that it "makes sense" to domain experts, without too much difficulty. While presentation and documentation do play a major role, it is equally important to have intuitive naming conventions for ontological entities and, in particular, a structural organization (i.e., a schema for a data graph) which is meaningful for the domain expert.

We can briefly illustrate this using an example from the OAEI[1] Conference benchmark ontologies [3, 4]. One lists "author of paper" and "author of student paper" as two distinct subclasses of "person." This raises the question: why is "author of student paper" not a subclass of "author of paper" (apart from subclassing both as "person" which we will discuss in the next paragraph). In another ontology in this collection, "author" is a subclass of "user", and "author" itself has exactly two subclasses: "author, who is not a reviewer" and "co-author" – which is hardly intuitive.

[1] For more information on the Ontology Alignment Evaluation Initiative, see http://oaei.ontologymatching.org/.

By definition, an ontology with high conceptual clarity will be much easier to re-use, simply because it is much easier to understand the ontology in the first place. Thus, a key quest for ontology research is to develop ontology modeling methodologies which make it easier to produce ontologies with high conceptual clarity.

That *following already established good modeling principles* makes an ontology easier to understand and re-use, should go without saying. However, good modeling principles are not simply a checklist that can easily be followed. Even simple cases, such as the recommendation to not have perdurants and endurants[2] together in subclass relationships (in the example above, author should not be a subclass of person; rather, authorship is a *role* of the person) are commonly not followed in existing ontologies. At the current stage of research, "good modeling" appears to largely be a function of modeling experience and more of an art, than a science, which has not been condensed well enough into tangible insights that can easily be written up in a tutorial or textbook.

A further issue is that even in cases where the aforementioned re-use challenges are manageable, implementing and subsequently maintaining re-use in practice is problematic due to *limited support for re-use in available tooling*. Once a reusable ontology resource has been located, preparing it for re-use – which can be done in many ways, e.g., by copying the entire design, by using *owl:Imports*, by copying individual definitions, by locally subsuming the remote ontology entities, etc. – is time-consuming and error-prone (especially when several resources are re-used).

Furthermore, through ontology re-use, the ontologist commits to a design and logic built by a third party. As the resulting ontology evolves, keeping track of the provenance of re-used ontological resources and their locally instantiated representations may become important, e.g., to resolve design conflicts resulting from differing requirements, or to keep up-to-date with the evolution of the re-used ontology. Such state-keeping is decidedly non-trivial without appropriate tool support.

[2] These are ontological terms; a perdurant means "an entity that only exists partially at any given point in time" and and endurant means "an entity that can be observed as a complete concept, regardless of the point time."

Processes and tools should be sought that make it possible to leverage modeling experience by seasoned experts, without actually requiring their direct involvement. This was one of the original ideas behind ontology re-use which, unfortunately, did not quite work out that well, for reasons including those mentioned above. Our modularization approach, however, together with the systematic utilization of ontology design patterns, and our accompanying tools, gives us a means to address this issue.

The notion of *module* has taken on a variety of meanings in the Semantic Web community [5–7]. For our purposes, a *module* is a part of the ontology (i.e., a subset of the ontology axioms) which captures a key notion together with its key attributes. For example, an event module may contain, other than an Event class, also relations and classes designed for the representation of the event's place, time, and participants. On the other hand, a module for a cooking recipe may encompass relations and classes for recording ingredients and their amounts, time and equipment required, and so on. A module is thus as much a technical entity, in the sense of a defined part of an ontology, as well as a conceptual entity, in the sense that it should encompass different classes (and relationships between them) which "naturally" (from the perspective of domain experts) belong together. Modules may overlap. They may be nested. They provide an organization of an ontology as an interconnected collection of modules, each of which resonates with the corresponding part of domain conceptualization by the experts.

Note that modules, in this sense, indicate a departure from a more traditional perspective on ontologies, where they are often viewed as enhanced taxonomies. From our perspective, the occurrence of subclass relationships within an ontology is rather coincidental, but is certainly not a key guiding principle for modeling or ontology organization. As we will see, modules make it possible to approach ontology modeling in a divide-and-conquer fashion; first, by modeling one module at a time, and then connecting them.[3]

Modules furthermore provide an easy way of avoiding the hassle of dealing with ontologies that are large and monolithic: understanding an ontology amounts to understanding each of its modules, and then their interconnections. This, at the same time, provides a recipe for documentation which resonates with domain experts' conceptualizations (which were captured by means of the modules), and thus makes the documentation and ontology easier to understand. Additionally, using modules facilitates modification, and thus adapting an ontology to a new purpose, as a module is much more easily replaced by a new module with, for instance, higher granularity, because the module inherently identifies where changes should be localized.

The systematic use of *ontology design patterns* [10, 11] is another central aspect of our approach, as many of their promises resonate with the issues that our approach is addressing [12]. An ontology design pattern is a generic solution to a recurring ontology modeling problem. To give an example, a "Trajectory" pattern would be a partial ontology that can be used to record "trajectories," such as the route of a ship or piece of cargo. If well-designed, this pattern may, with only minor and easy modifications, be suitable to be used as a template for trajectory modules within many ontologies. It must be noted that patterns are not one-size-fits-all solutions. For example, the Trajectory pattern from [13], which we have found to be highly versatile, assumes a discretized recording of a trajectory (as a time-sequence of locations), however it would not account for recording of a trajectory as, say, a set of equations.

In our approach, well-designed ontology design patterns, provided as templates to the ontology modelers, make it easier to follow already established good modeling principles, as the patterns themselves will already reflect them [14]. When a module is to be modeled, within our process there will always be a check whether some already existing ontology design pattern is suitable to be adapted for the purpose. Modules, as such, are often derived from patterns as templates.

The principles and key aspects laid out above are tied together in a clearly defined modular ontology modeling process which is laid out below, and which is a refinement – with some changes of emphasis – of the eXtreme Design methodology [15]. It is furthermore supported by a set of tools developed for support of this process, the CoModIDE plug-in to Protégé, and which we will discuss in detail below. Also central to our approach is that it is usually a collaborative process with a (small) team that jointly has the required domain, data and ontology engineering expertise, and that the actual modeling work utilizes schema

[3] Other divide and conquer approaches have also recently been proposed [8, 9], and while they seem to be compatible with ours, exact relationships still need to be established.

diagrams as the central artifact for modeling, discussion, and documentation.

This paper is structured as follows. Section 2 describes our related work – this covers precursor methods, the eXtreme Design methodology, and overviews of concepts fundamental to our approach. Section 3 describes our modular ontology modeling process in detail. Section 4 presents CoModIDE as a tool for supporting the development of modular ontologies through a graphical modeling paradigm, as well as a rigorous evaluation of its effectiveness and usability. Section 5 describes additional, supporting infrastructure for this process. Then, in Section 6, we provide experiential data on executing this process. Finally, in Section 7, we conclude.

This paper is significantly extends [16] and summarizes several other workshop and conference papers: [17], [18], and [19].

2. Related Work

2.1. Ontology Engineering Methods

The ideas underpinning the Modular Ontology Modeling methodology build on years of prior ontology engineering research, covering organizational, process, and technological concerns that impact the quality of an ontology development process and its results.

The METHONTOLOGY methodology is presented by Férnandez et al. in [20]. It is one of the earlier attempts to develop a development method specifically for ontology engineering processes (prior methods often include ontology engineering as a subdiscipline within knowledge management, conflating the ontology-specific issues with other more general types of issues). Férnandez et al. suggest, based largely on the authors' own experiences of ontology engineering, an ontology lifecycle consisting of six sequential work phases or *stages*: *Specification*, *Conceptualisation*, *Formalisation*, *Integration*, *Implementation*, and *Maintenance*. Supporting these stages are a set of support activities: *Planification*, *Acquiring knowledge*, *Documenting*, and *Evaluating*.

The On-To-Knowledge Methodology (OTKM) [21] is, similarly to METHONTOLOGY, a methodology for ontology engineering that covers the big steps, but leaves out the detailed specifics. OTKM is framed as covering both ontology engineering and a larger perspective on knowledge management and knowledge processes, but it heavily emphasises the ontology development activities and tasks (in [21] denoted the *Knowledge Meta Process*). OTKM emphasises initial collaboration between domain experts and ontology engineers in the *Kick-off* phase. In the subsequent *Refinement* phase an ontology engineer formalises the initial semi-formal model into a real ontology on their own, without aid of a domain expert. In subsequent *Evaluation*, both technical and user-focused aspects of the knowledge based system in which the ontology is used, are evaluated. Finally, the *Application and Evolution* phase concerns the deployment of said knowledge based system, and the organisational challenges associated with maintenance responsibilities.

DILIGENT, by Pinto et al. [22], is an abbreviation for *Distributed, Loosely-Controlled and Evolving Engineering of Ontologies*, and is a method aimed at guiding ontology engineering processes in a distributed Semantic Web setting. The method emphasises decentralised work processes and ontology usage, domain expert involvement, and ontology evolution management. This distributed development process is formalised into five activities: *build*, *local adaptation*, *analysis*, *revision*, and *local update*. The authors show how Rhetorical Structure Theory [23] can be used as a framework to constrain design discussions in a distributed ontology engineering setting, guiding the design process.

In all three of these well-established methods, the process steps that are defined are rather coarse-grained. They give guidance on overall activities that need to be performed in constructing an ontology, but more fine-grained guidance (e.g., how to solve common modeling problems, how to represent particular designs on concept or axiom level, or how to work around limitations in the representation language) is not included. It is instead assumed that the reader is familiar with such specifics of constructing an ontology. This lack of guidance arguably is a contributor to the three issues preventing re-use, discussed in Section 1.

2.2. Ontology Design Patterns

Ontology Design Patterns (ODPs) were introduced at around the same time independently by Gangemi [11] and Blomqvist and Sandkuhl [10], as potential solutions to the drawbacks of classic methods described above. The former defines such patterns by way of the characteristics that they display, including examples such as *"[an ODP] is a template to represent, and possibly solve, a modelling problem"* [11, p. 267] and *"[an ODP] can/should be used to describe a 'best practice' of modelling"* [11, p. 268]. The latter describes ODPs as generic descriptions of recurring constructs in ontologies, which can be used to construct components or modules of an ontology. Both approaches emphasise that patterns, in order to be easily reusable, need to include not only textual descriptions of the modelling issue or best practice, but also some formal ontology language encoding of the proposed solution. The documentation portion of the pattern should be structured and contain those fields or slots that are required for finding and using the pattern.

A substantial body of work has been developed based on this idea, by a sizable distributed research community[4]. Key contributions include the eXtreme Design methodology (detailed in Section 2.3) and several other pattern-based ontology engineering methods (Section 2.4). The majority of work on ODPs has been based on the use of miniature OWL ontologies as the formal pattern encoding, but there are several examples of other encodings, the most prominent of which are OPPL [24] and more recently OTTR [9].

MOMo extends on those methods, but also incorporates results from our past work on how to document ODPs [25–27], how to implement ODP support tooling [28] and how to instantiate patterns into modules by "stamping out copies" [14].

2.3. eXtreme Design

The eXtreme Design (XD) methodology [15] was originally proposed as a reaction to previous waterfall-oriented methods (e.g., some of those discussed above). XD instead borrows from agile software engineering methods, emphasizing a divide-and-conquer approach to problem-solving, early or continuous deployment rather than a "one-shot" process, and early and frequent refactoring as the ontology grows. Crucially, XD is built on reusing of ontological best practices via ODPs.

The XD method consists of a number of tasks, as illustrated in Figure 1. The first two tasks deal with establishing a project context (i.e., introducing initial terminology and obtaining an overview of the problem) and collecting initial requirements in the form of a prioritized list of user stories (describing the required functionality in layman's terms). These steps are performed by the whole XD team together with the customer, who is familiar with the domain and who understands the required functionalities of the resulting ontology. The later steps of the process are performed in pairs of two developers (these steps are in the figure enclosed in the large box). They begin by selecting the top prioritised user story that has not yet been handled, and transform that story into a set of requirements in the form of competency questions (data queries), contextual statements (invariants), and reasoning requirements. Customer involvement at this stage is required to ensure that the user story has been properly understood and that the elicited requirements are correctly understood.

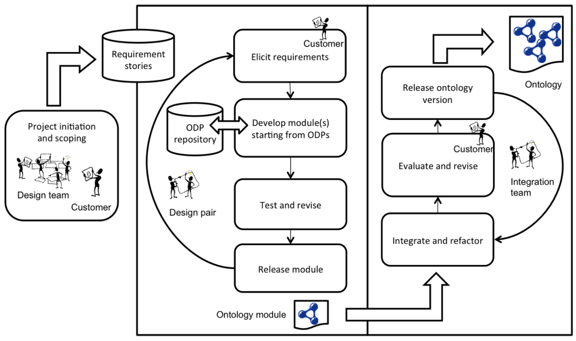

Fig. 1. eXtreme Design method overview, from [15].

The development pair then selects one or a small set of interdependent competency questions for modelling. They attempt to match these against a known ODP, possibly from a designated ODP library. The ODP is adapted and integrated into the ontology module under development (or, if this iteration covers the first requirements associated with a given user story, a new module is created from it). The module is tested against the selected requirements to ensure that it covers them properly. If that is the case, then the next set of requirements from the same user story is selected, a pattern is found, adapted, and integrated, and so on. Once all requirements associated with one user story have been handled, the module is released by the pair and integrated with the ontology developed by the other pairs in the development team. The integration may be performed either by the development pair themselves, or by a specifically designated integration pair.

XD has been evaluated experimentally and observationally, with results indicating that the method con-

[4]https://ontologydesignpatterns.org

tributes to reduced error rates in ontologies [29, 30], increased coverage of project requirements [29], and that pattern usage is perceived as useful and helpful by inexperienced users [29–31]. However, results also indicate that there are pitfalls associated with a possibility of over-dependence on ODP designs, as noted in [31].

2.4. Other Pattern-based Methods

SAMOD [32], or *Simplified Agile Methodology for Ontology Development*, is a recently developed methodology that builds on and borrows from test-driven and agile methods (in particular eXtreme Design). SAMOD emphasises the use of tests to confirm that the developed ontology is consistent with requirements, and prescribes that the developer construct three types of such tests: *model tests*, *data tests*, and *query tests*. The method prescribes a light-weight three-step process broadly mirroring XD, i.e., consisting of (1) constructing an ontology module as a partial solution to the development scenario (including tests), (2) merging that new module into the main branch ontology, (3) refactoring as needed. After each of these steps, all the tests defined for the module and/or main branch ontology are executed, and development is halted until all tests are passed.

Hammar [33] presents a set of proposed improvements to the XD methodology under the umbrella label "XD 1.1". These include (1) a set of roles and role-specific responsibilities in an XD project, (2) suggestions on how to select and implement other forms of ontology re-use in XD than just patterns (e.g., *import*, *remote references*, *slicing*, *partial cloning*), and (3) a project adaptation questionnaire supporting XD projects in adapting the process to their particular development context (e.g., team cohesion, distribution, skill level, domain knowledge, etc).

XD, SAMOD, and XD 1.1 emphasize the needs for suitable support tooling for, e.g., finding suitable ODPs, instantiating those ODPs into an ontology, and executing tests across the ontology or parts of it. In developing MOMo and the CoModIDE platform, we propose and develop solutions to two additional support tooling needs: that of *intuitive and accessible graphical modeling*, and that of a *curated high-quality pattern library*.

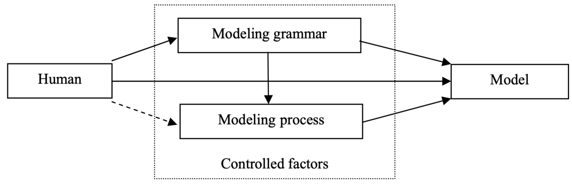

Fig. 2. Factors affecting conceptual modeling, from [34].

2.5. Graphical Conceptual Modelling

[34] proposes three factors (see Figure 2) that influence the construction of a conceptual model, such as an ontology; namely, the *person* doing the modeling (both their experience and know-how, and their interpretation of the world, of the modeling task, and of model quality in general), the *modeling grammar* (primarily its expressive power/completeness and its clarity), and the *modeling process* (including both initial conceptualisation and subsequent formal model-making). Crucially, only the latter two factors can feasibly be controlled in academic studies. Research in this space tends to focus on one or the other of these factors, i.e., studying the characteristics of a modeling language *or* a modeling process. Our work on CoModIDE straddles this divide: employing graphical modeling techniques reduces the grammar available from standard OWL to those fragments of OWL that can be represented intuitively in graphical format; employing design patterns affects the modeling process.

Graphical conceptual modeling approaches have been extensively explored and evaluated in fields such as database modeling, software engineering, business process modeling, etc. Studying model grammar, [35] compares EER notation with an early UML-like notation from a comprehensibility point-of-view. This work observes that restrictions are easier to understand in a notation where they are displayed coupled to the types they apply to, rather than the relations they range over. [36] proposes a quality model for EER diagrams that can also extend to UML. Some of the quality criteria in this model, that are relevant in graphical modeling of OWL ontologies, include *minimality* (i.e., avoiding duplication of elements), *expressiveness* (i.e., displaying all of the required elements), and *simplicity* (displaying no more than the required elements).

[37] studies the usability of UML, and reports that users perceive UML class diagrams (closest in intended use to ontology visualizations) to be less easy-to-use than other types of UML diagrams; in partic-

ular, relationship multiplicities (i.e., cardinalities) are considered frustrating by several subjects. UML displays such multiplicities by numeric notation on the end of connecting lines between classes. [38] analyses UML and argues that while it is a useful tool in a design phase, it is overly complex and as a consequence, suffers from redundancies, overlaps, and breaks in uniformity. [38] also cautions against using difficult-to-read and -interpret adornments on graphical models, as UML allows.

Various approaches have been developed for presenting ontologies visually and enabling their development through a graphical modeling interface, the most prominent of which is probably *VOWL*, the *Visual Notation for OWL Ontologies* [39], and its implementation viewer/editor WebVOWL [40, 41]. VOWL employs a force-directed graph layout (reducing the number of crossing lines, increasing legibility) and explicitly focuses on usability for users less familiar with ontologies. As a consequence of this, VOWL renders certain structures in a way that, while not formally consistent with the underlying semantics, supports comprehensibility; for instance, datatype nodes and `owl:Thing` nodes are duplicated across the canvas, so that the model does not implode into a tight cluster around such often used nodes. It has been evaluated over several user studies with users ranging from laymen to more experienced ontologists, with results indicating good comprehensibility. CoModIDE has taken influence from VOWL, e.g., in how we render datatype nodes. However, in a collaborative editing environment in which the graphical layout of nodes and edges needs to remain consistent for all users, and relatively stable over time, we find the force-directed graph structure (which changes continuously as entities are added/removed) to be unsuitable.

For such collaborative modeling use cases, the commercial offering *Grafo*[5] offers a very attractive feature set, combining the usability of a VOWL-like notation with stable positioning, and collaborative editing features. Crucially, however, Grafo does not support pattern-based modular modeling or import statements, and only supports RDFS semantics, and as a web-hosted service, does not allow for customizations or plugins that would support such a modeling paradigm.

CoModIDE is partially based on the Protégé plugin *OWLAx*, as presented in [42]. *OWLAx* plugin supports one-way translation from graphical schema diagrams drawn by the user, into OWL ontology classes and properties; however, it does not render such constructs back into a graphical form. There is thus no way of continually maintaining and developing an ontology using only OWLAx. There is also no support for design pattern re-use in this tool.

[5]https://gra.fo

3. The Modular Ontology Modeling Methodology

Modular Ontology Modeling (MOMo[6]) consists of a well-defined process, together with the utilization of specific components that support the process. In this part of the paper, we lay out the key components, namely schema diagrams, our approach to OWL axiomatization, ontology design patterns, and the concept of modules already mentioned previously, as well as the process which ties them together. In section 4 and 5, we discuss our supporting tools and infrastructure, however they should be considered just one possible instantiation of the more general MOMo methodology. Indeed, most of the first part of the MOMo process is, in our experience, best done in analog mode, armed with whiteboards, flip-charts and a suitable modeling team.

3.1. The Modeling Team

Team composition is of critical importance for establishing a strong, versatile modular ontology. Different perspectives are very helpful, as long as the group does not lose focus. Arrival at a consensus model between all parties which constitutes a synthesis of different perspectives is key, and such a consensus is much more likely to be suitable to accommodate future use cases and modifications. It is therefore advisable to have more than one domain expert with overlapping expertise, and more than one ontology engineer on the team. Based on our experiences, three types of participants are needed in order to have a team that can establish a good modular ontology: domain experts, ontology engineers, and data scientists. Of course some people may be able to fill more than one role. An overall team size of 6-12 people appears to be ideal, based on our experiences (noted in Section 6). Meetings with the whole team will be required, but in the MOMo process

[6]Momo is the protagonist in the 1973 fantasy novel “Momo” by Michael Ende. The antagonists are Men in Grey that cause people to waste time.

most of the work will fall on the ontology engineers between the meetings.

1. The *domain experts* should primarily bring a deep knowledge of the relevant subject area(s) and of the use case scenario(s). Ideally, they should also be aware of perspectives taken by other domain experts in order to avoid overspecialization of the model.
2. The *ontology engineers* should be familiar with the MOMo process, supporting tools, and relevant standards (in particular, OWL), and guide the meetings. Their role is to capture the discussions, resulting in (draft) schema diagrams which are then further discussed and refined by the team. Between team meetings, they will also work out detailed documentation of what has been discussed, which, in turn, will be used as prompts in following modeling sessions. At least one of the ontology engineers should have a deep understanding of the logical underpinnings of OWL.
3. The *data scientists* should bring a detailed understanding of the actual data that is relevant to the use case(s) and will or may be utilized (e.g., integrated by means of the ontology as overarching schema). Their role is to make sure that the model does not deviate in an incompatible way from the actual data that is available.

3.2. Schema Diagrams

Schema diagrams are a primary tool of the MOMo process. In particular, they are the visual vehicle used to coalesce team discussions into a draft model and used centrally in the documentation. This diagram-based approach is also reflected in our tools, which we will present in sections 4–5.

Let us first explain what we do – and do not – mean by schema diagram, and we use Figure 3 as an example,[7] which depicts the Provenance module from the Enslaved Ontology [43].

Our schema diagrams are labeled graphs that indicate OWL entities and their (possible) relationships. Nodes can be labeled by (1) classes (EntityWithProvenance – rectangular, orange, solid border), (2) modules (Agent, PersonRecord, ProvenanceActivity – rectangular, light blue, dashed border), (3)

[7]The schema diagrams in this paper were produced with yEd, available from https://www.yworks.com/products/yed/.

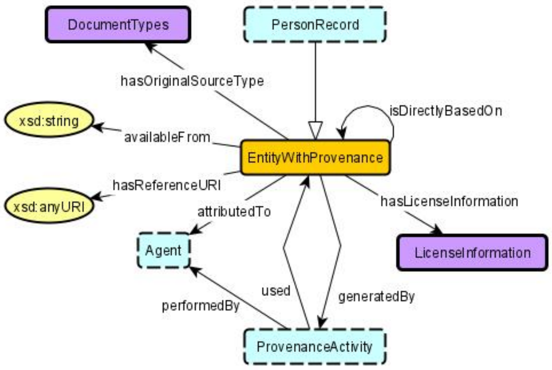

Fig. 3. Schema diagram for the Provenance module from the Enslaved Ontology [43]. It is based on the Provenance pattern from [18], which in turn is based on the core of PROV-O [44].

controlled vocabularies (DocumentTypes, LicenseInformation – rectangular, purple, solid border), (4) datatypes (xsd:string, xsd:anyURI – oval, yellow, solid border). Arrows can be white-headed without label, indicating a subclass relationship (the arrow between PersonRecord and EntityWithProvenance) or can be labeled with the name of the property, which could be a data or an object property, which is identified by the target of the arrow, which may be a datatype.

Indication of a module in a diagram means that instead of the node (the light blue, dashed border), there may be a complex model in its very own right, which would be discussed, depicted, and documented separately. For example, PersonRecord in the Enslaved Ontology is a complex module with several sub-modules. The diagram in Figure 3 "collapses" this into a single node, in order to emphasize what is essential for the Provenance module. Controlled vocabularies are predefined sets of IRIs with a specific meaning that is documented externally (i.e., not captured in the ontology itself). A typical example would be IRIs for physical units like meter or gram or, as in our example diagram, IRIs for specific copyright licences, such as CC-BY-SA. Datatypes would be the concrete datatypes allowed in OWL. This type of schema diagram underlays a study on automatic schema diagram creation from OWL files [17].

Note that our schema diagrams do *not* directly indicate what the underlying OWL axioms are. A (labeled) arrow only indicates that a property could *typically* be used between nodes of the indicated types. It does not indicated any of functionality, existential or universal restriction, etc. It also does not indicate any specific domain or range axioms or use of logical connectives, such as conjunction or disjunction. In the end, the ontology will consist of a set of OWL axioms (i.e., a concrete axiomatization will be done), but these are created rather late in the process. During team modeling, simple diagrams help to focus on the essentials and are intuitively accessible even for participants with no background in ontology engineering. The ontology engineers, however, should keep in mind that logical axioms are needed eventually and that the diagrams alone remain highly ambiguous.

3.3. OWL Axioms

As already mentioned, OWL axioms are the key constituents of an ontology as a data artifact, although in our experience quality documentation is of at least the same importance. As has been laid out elsewhere [45], axiomatizations can have different interpretations, and while they can, for example, be used for performing deductive reasoning, this is not their main role as part of the MOMo approach. Rather, for our purposes axioms serve to *disambiguate* meaning, for a human user of the ontology. As such, they can also be understood as a way to disambiguate the schema diagram, as appropriate (e.g., by labeling a property functional, by declaring domain and range restrictions).

As such, we recommend a rather complete axiomatization, as long as it does not force an overly specific reading on the ontology. We usually use the checklist from the OWLAx tool [42] to axiomatize with simple axioms. More complex axioms, in particular those that span more than two nodes in a diagram, can be added conventionally or by means of the ROWLTab Protégé plug-in [46, 47]. We also utilize what we call *structural tautologies* which are axioms that are in fact tautologies such as $A \sqsubseteq \geqslant 0R.B$, to indicate that individuals in classes A and B *may* have an R relation between them, and that this would be a typical usage of the property R.

3.4. Ontology Design Patterns

As already mentioned, Ontology Design Patterns (ODPs) have originated in the early 2000s as reusable solutions to frequently occurring ontology design problems. Most ODPs can currently be found on the ontologydesignpatterns.org portal, and they appear to be of very varied quality both in terms of their design and documentation, and following a variety of different design principles. While they proved to be useful for the community [12], as part of MOMo, we reimagine ontology design patterns and their use.

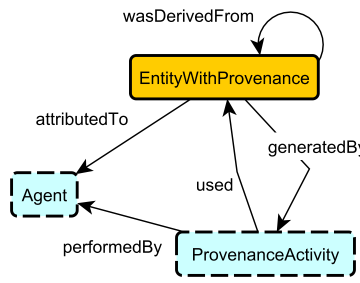

Fig. 4. Schema diagram of the MODL Provenance ODP. It is based on the core of PROV-O [44]

Most importantly, rather than working with a crowdsourced collection of ODPs, there seems to be a significant advantage in working with a well-curated *library* of ontology design patterns that are developed with a similar mindset, and expressed and documented in a uniform way. A first version of such a library is the Modular Ontology Design Library (MODL) [18], which contains patterns that we have frequently found to be useful in the recent past. We furthermore utilize the Ontology Pattern Language (OPLa) [26, 27] which is an annotation language using OWL that makes it possible to work with ODPs (and modules) in a programmatic way.

As an example, a schema diagram for the MODL Provenance pattern is provided in Figure 4. In MOMo, the pattern would be used as a *template* in the sense that it serves as a blueprint, usually for a module – such as the Provenance module depicted in Figure 3 in the resulting ontology. That is, the pattern can be modified, simplified, extended at will, but usually both the schema diagram and the axioms of the ODP will still be reflected and in some way recognizable in the module. The resulting ontology will also use OPLa to capture the information that the resulting module has re-used an ODP as a template.

3.5. Modules

An (ontology) *module* is a part of an ontology which captures a key notion, and its key relations to other notions. An example that was already discussed is given in Figure 3. A module may sometimes consist of a central class together with relations (properties) to other

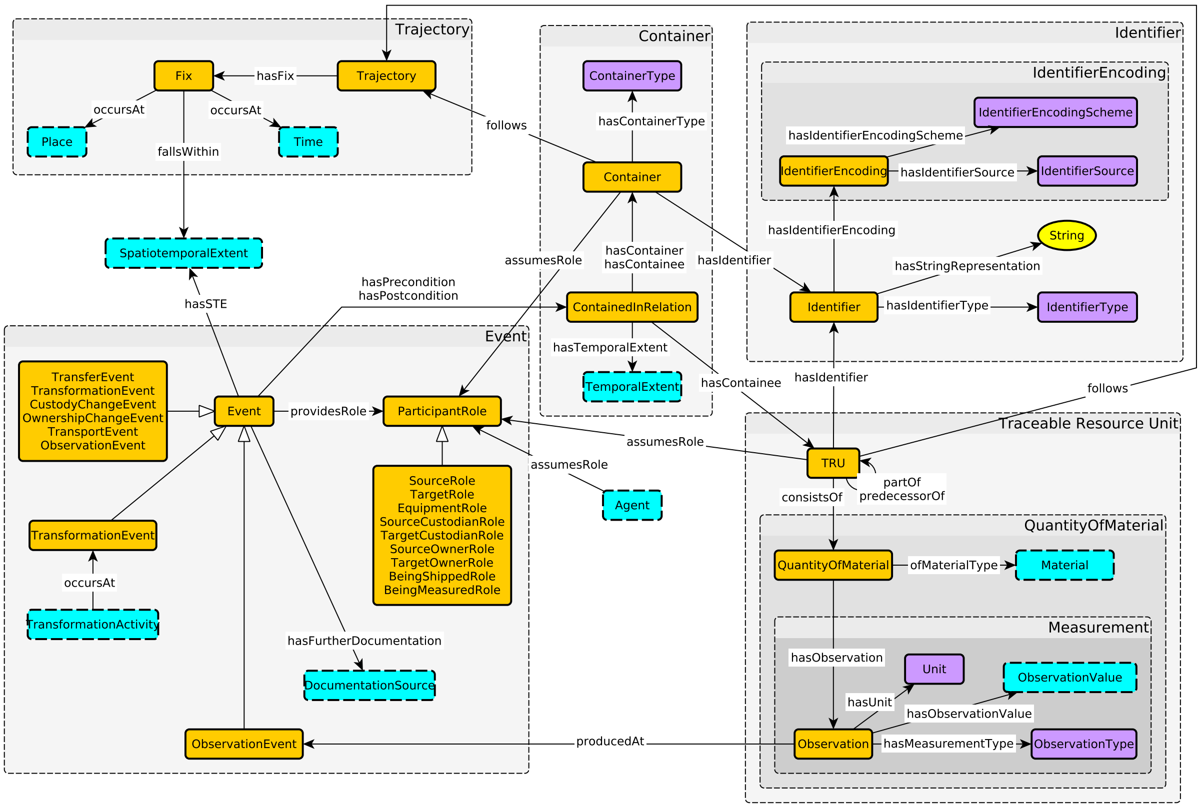

Fig. 5. Schema diagram of a supply chain ontology currently under development by the authors.

classes, modules, controlled vocabularies or datatypes, but can sometimes also be of a more complex structure.

Modules can be overlapping, or nested, and there is often no clear-cut answer to the question whether a particular property and class does or does not belong to a module. E.g., in context of Figure 3, the PersonRecord class could reasonably be considered to be outside the module. Likewise, the EntityWithProvenance class may or may not be considered part of the PersonRecord module. The latter may depend on the question how "central" provenance for person records is, in the application context of the ontology. In this sense, ontology modules are ambiguous in their delineation, just as the human concepts they are based on.

As a data artifact, though, i.e., in the OWL file of the ontology, we will use the above-mentioned Ontology Pattern Language OPLa to identify modules, i.e. the ontology engineers will have to make an assessment how to delineate each module in this case. OPLa will furthermore be used to identify ODPs (if any) which were used as templates for a module.

Finally, an ontology's modules will drive the documentation, which will usually discuss each module in turn, with separate schema diagrams, axioms, examples and explanations, and will only at the very end discuss the overall ontology which is essentially a composition of the modules. In a diagram that encompasses several modules, the modules can be identified visually using frames or boxes around sets of nodes and arrows. An example for this is given in figure 5. Several modules are identified by grey boxes in this diagram, including nested modules such as on the lower right.

3.6. The MOMo Workflow

We now describe the Modular Ontology Modeling workflow that we have been applying and refining over the past few years. It borrows significantly from the eXtreme Design approach described in Section 2.3, but has an emphasis on modularization, systematic use of schema diagrams, and late-stage OWL generation.

First, we briefly list the steps of the workflow, and then discuss each of them in more detail. A walk-through tutorial for the approach can be found in [48].

1. Describe use cases and gather possible data sources.
2. Gather competency questions.
3. Identify key notions for the domain to be modeled.
4. Identify existing ontology design patterns to be used.
5. Create schema diagrams for modules.
6. Set up documentation and determine axioms for each module.
7. Create ontology schema diagram from the module schema diagrams.
8. Add axioms spanning more than one module.
9. Reflect on entity naming and all axioms.
10. Create OWL file(s).

This workflow is not necessarily a strict sequence, and work on later steps may cause reverting to an earlier step for modifications. Sometimes subsequent steps are done together, e.g., 4 and 5, or 7 and 8.

Steps 1 through 4 can usually be done through a few shorter one-hour teleconferences (or meetings), the number of which depends a lot on the group dynamics and prior experience of the participants. This sequence would usually also include a brief tutorial on the modeling process. If some of the participants already have a rather clear conception of the use cases and data sources, then 2 or 3 one-hour calls would often suffice.

In our experience, synchronous engagement (in the sense of longer meetings) of the modeling team usually cannot be avoided for step 5. Ideally, they would be conducted through in-person meetings, which for efficiency should usually be set up for 2 to 3 subsequent days. Online meetings can also be almost as effective, but for this we recommend several, at least 3, subsequent half-day sessions about 4-5 hours in length.

Steps 6 to 10 are mostly up to the ontology engineers at the team, however they would request feedback and correctness checks from the data and domain experts. This can be done asynchronously, but depending on preference could also include some brief teleconferences (or meetings).

Design an ontology that can be used as part of a “recipe discovery” website. The ontology shall be set up such that content from existing recipe websites can be mapped into it (i.e. the ontology will be populated with data from the recipe websites). On the discovery website, detailed graph-queries (using the ontology) shall produce links to recipes from different recipe websites as results. The ontology should be extendable towards incorporation of additional external data, e.g., nutritional information about ingredients or detailed information about cooking equipment.

Fig. 6. Example use case description, taken from [48].

3.6.1. Describe use cases and gather possible data sources

As the first step, the use case, i.e., the problem to be addressed, should be described. The output description can be very brief, e.g., a paragraph of text, and it does not necessarily have to be very crisp. In fact it may describe a set of related use cases rather than one specific use case, and it may include future extensions which are currently out of scope. Setting up a use case description in this way alerts the modeling team to the fact that the goal is to arrive at a modular ontology that is extensible and re-useable for adjacent but different purposes.

An example for such a use case description can be found in Figure 6. In this particular case, the possible data sources would be a set of different recipe websites such as allrecipes.com.

3.6.2. Gather competency questions

Competency questions are examples for queries of interest, expressed in natural language, that should be answerable from the data graph with which the ontology would be populated. Competency questions help to refine the use case scenario, and can also aid as a sanity check on the adequacy of the data sources for the use case. While the competency questions can often be gathered during work on the use case description, it is sometimes also helpful to collect them from potential future users. For example, for an ontology on the history of the slave trade [43], professionals, school children, and some members of the general public were asked to provide competency questions. A few examples are provided in Figure 7. We found that in many cases, 10-12 sufficiently different competency questions will be enough.

Who were the godparents of my great-great grandmother, Beatriz of the Ambaca nation, baptized at São José church in Rio de Janeiro on April 12, 1840?

Who did Thomas Jefferson enslave at Monticello?

I am researching an enslaved person named Mohammed who was a new arrival from West Africa in Charleston in 1776. Is there data about what slave ship he might have been on?

Fig. 7. Example competency questions, taken from [43].

Recipe	RecipeName	RecipeInstructions
TimeInterval	QuantityOfFood	Quantity
Equipment	FoodType	Difficultylevel
RecipeClassification	NutritionalInfo	Source

Fig. 8. Example for key notions in the scenario of Figure 6, taken from [48].

3.6.3. *Identify key notions for the domain to be modeled*

This is a central step which sets the stage for the actual modeling work in step 5. The main idea is that each of the identified key notions will become a module, however, during modeling, some closely related notions may also become combined into a single module. It is also possible that at a later stage is is realized that a key notion had been forgotten, which is easily corrected by adding the new key notion to the previous list.

The key notion are determined by the modeling team, by taking into consideration the use case description, the possible data sources, and the competency questions from the previous steps. An example for key notions, for the recipe scenario from Figure 6, is give in Figure 8.

3.6.4. *Identify existing ontology design patterns to be used*

In MOMo, we utilize pattern libraries such as MODL. For each of the key notions identified in the previous step, we thus attempt to find a pattern from the library which seems close enough or modifiable, so that it can serve as a template for a first draft of a corresponding module. For example, for source, it seems reasonable to use the Provenance pattern depicted in Figure 4. MODL also has patterns for quantities.

For some key notions there may be different reasonable choices for a pattern. For example, Recipe may be understood as a Document, a Plan, or a Process. In this case the modeling team should consult the use case and the competency questions to select a pattern that seems to be a good overall fit.

In some cases, there will be no pattern in the library which can reasonably be used as as template. This is of course fine, it just means that the module will have to be developed from scratch.

3.6.5. *Create schema diagrams for modules*

This step usually requires synchronous work sessions by the modeling team, led by the ontology engineers. The key notions are looked at in isolation, one at a time, although of course the ontology engineers should simultaneously keep an eye on basic compatibility between the draft modules. The modeling order is also important. It often helps to delay the more complicated, involved or controversial modules, and focus first on modules that appear to be relatively clear or derivable from an existing pattern. It is also helpful to begin with notions that are most central to the use case.

A typical modeling session could begin with a discussion what pattern may be most suitable to use as a template (thus overlapping with step 4). Or it could start with the domain experts attempting to explain the key notion, and its main aspects, to the ontology engineers. The ontology engineers would query about details of the notion, and also about available data, until they can come up with a draft schema diagram which can serve as a *prompt*.

Indeed, the idea of prompting with schema diagrams is in our experience a very helpful one for these modeling sessions. A prompt in this sense does not have to be exact or even close in terms of the eventual solution. Rather, the diagram used as prompt reflects an attempt by the ontology engineer based on his current (and often naturally) limited understanding of the key notion. Usually, such a prompt will prompt(!) the domain and data experts to point out the deficiencies of the prompt diagram, thus making it possible to refine or modify it, or to completely reject it and come up with a new one. Discussions around the prompts also sometimes expose disagreements between the different domain experts in the team, in which case the goal is to find a consensus solution. It is important, though, that the ontology engineers attempt to keep the discussion focused on mostly the notion currently modeled.

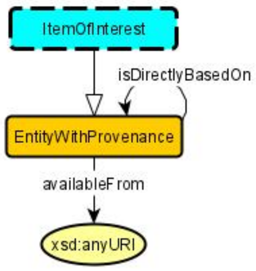

Fig. 9. A minimalistic provenance module based on the MODL Provenance pattern shown in Figure 4.

Ontology engineers leading the modeling should also keep in mind that schema diagrams are highly ambiguous. This is important for several reasons.

For instance, some critique by a domain expert may be based on an unintended interpretation of the diagram. When appropriate, the ontology engineers should therefore explain the meaning of the diagram in natural language terms, such as "there is one *hasChild* arrow leading from the *Person* class to itself, but this does not necessarily mean that a person can be their own child." It is sometimes indeed helpful to keep this in mind when creating schema diagrams; in the example just given, the diagram could have two *Person* classes depicted, with the *hasChild* arrow pointing from one of them to the other. Good namings of classes and properties in the diagram will also help to avoid unintended interpretations.

Furthermore, eventually (see the next step) the ontology engineers will have to convert the schema diagrams into a formal model which will no longer be ambiguous. The ontology engineers should therefore be aware that they need to understand how to interpret the diagram in the same way as the domain experts. This can usually be done by asking the domain experts – during this step or a subsequent one – concrete questions about the intended meaning, e.g., whether a person can have several children, or at most one, etc.

It is of course possible that a module may use a pattern as a template, but will end up to being a highly simplified version of the pattern. E.g., the provenance module depicted in Figure 9 was derived from the pattern depicted in Figure 4.

3.6.6. *Set up documentation and determine axioms for each module*

We consider the documentation to be a primary part of an ontology: In the end, an OWL file alone, in particular if sizeable, is really hard to understand, and it will mostly be humans who will deal with the ontology when it is populated or re-used. In MOMo, creation of the documentation is in fact an integral part of the modeling process, and the documentation is a primary vehicle for communication with the domain and data experts in order to polish the model draft.

MOMo documentations – see [49] for an example – discuss each of the modules in turn, and for each module, a schema diagram is given together with the formal OWL axioms (and possible additional axioms not expressible in OWL) that will eventually be part of the OWL file. Since the documentation is meant for human consumption, we prefer to use a concise formal representation of axioms, usually using description logic syntax or rules, together with an additional listing of the axioms in a natural language representation.

Domain and data experts can be asked specific questions, as mentioned above, to determine the most suitable axioms. Sometimes, the choice of axiom appears to be arbitrary, but would have direct bearing on the data graph. An example for this would be whether the property *availableFrom* in Figure 9 should be declared functional. Indeed, if declared functional, then any *EntityWithProvenance* can have at most one URI it is available from. This may or may not be desired in terms of data or use case, or it may simply be a choice that has to be made by the modeling team in order to disambiguate how the model shall be used.

In our experience, using axioms that only contain two classes and one property suffices to express an overwhelming majority of the desired logical theory. We are thus utilizing the relatively short list of 17 such axioms that was determined for support in the OWLAx Protégé plug-in [42] and that can also be found discussed in [48]. More complex axioms can of course also be added as required. Axioms can often also be derived from the patterns used as templates.

We would like to mention, in particular, two types of axioms that we found very helpful. One of them are *structural tautologies* which we have already discussed in Section 3.3. The other are *scoped* domain (respectively, range) axioms (introduced as the *class-oriented strategy* in [50]).

Scoped domain (resp., range) axioms differ from unscoped or global ones in that they make the domain (resp., range) contingent on the range (resp., domain). In formal terms, a domain axiom is of the form

$\exists R.\top \sqsubseteq B$, which indicates that the global domain of R is B. The scoped version is $\exists R.A \sqsubseteq B$, i.e., in this case the domain of R falls into B only if the range of R falls into A. The situation for range is similar: Global range is $\top \sqsubseteq \forall R.B$, indicating that the global range of R is B, while the scoped version is $A \sqsubseteq \forall R.B$, which states that the range of R falls into B only if the domain falls into A.

Using scoped versions of domain and range helps to avoid making overly general domain or range axioms. E.g., if you specify two global domains for a property R, then the domain would in fact amount to a conjunction of the two domains given. In the scoped case this is avoided, if the corresponding ranges are, for example, disjoint.

To give an example, consider the two scoped domain axioms $\exists$providesRole.WhiteChessPlayerRole $\sqsubseteq$ ChessGame and $\exists$providesRole.EmployeeRole $\sqsubseteq$ Organization. These two axioms are scoped domain axioms for providesRole, however they would not interfere. The same could not be reasonably stated using global domain axioms.

We generally recommend to use scoped versions of domain and range axioms – and, likewise, for functionality, inversefunctionality, and cardinality axioms – instead of the global versions. It makes the axioms easier to re-use, and avoids overly general axioms which may be undesirable in a different context.

3.6.7. Create ontology schema diagram from the module schema diagrams, and add axioms spanning more than one module

A combined schema diagram, see Figure 5 for an example, can be produced from the diagrams for the individual modules, In our experience, it is best to focus on understandability of the diagram [17]. The following guidelines should be applied with caution – exceptions at the right places may sometimes be helpful.

- Arrange key classes in columns and rows.
- Prefer vertical or horizontal arrows; this will automatically happen if classes are arranged in columns and rows.
- Avoid sub-class arrows: We have found that subclass arrows can sometimes be confusing for readers that are not intimately familiar with the formal logical meaning of them. E.g., in Figure 5, SourceRole is a subclass of ParticipantRole, which means that a container may assume SourceRole. However the diagram does not show a direct arrow from Container to the box containing SourceRole, and this in some cases makes the diagram harder to understand, in particular if there is an abundance of sub-class relationships.
- Prefer straight arrows.
- Avoid arrow crossings; if they are needed, make them near perpendicular.
- Use "module" boxes (light blue with dashed border) to refer to distant parts of the diagram to avoid cluttering the diagram with too many arrows.
- Avoid partial overlap of module groupings (grey boxes) in the diagram, even if modules are in fact overlapping.
- Break any guideline if it makes the diagram easier to understand.

The schema diagram for the entire ontology should then also be perused for additional axioms that may span more than one module. These axioms will often be rather complex, but they can often be expressed as rules. For complex axioms, rules are preferable over OWL axioms since they are easier for humans to understand and create [47]; the ROWLtab Protégé plugin [46] can for example be used to convert many of these rules into OWL.

3.6.8. Reflect on entity naming and all axioms

Good names for ontology entities, in particular classes and properties, are very helpful to make an ontology easier to understand and therefore to re-use. We use a mix of established practice, common sense, and our own naming conventions which have proven to be useful. We list the most important ones in the following.

- The entity names (i.e., the last part of the URI, after the namespace) should be descriptive. Avoid the encoding of meaning in earlier parts of the URI. An exception would be concrete datatypes such as xsd:string.
- Begin class names and controlled vocabulary names with uppercase letters, and properties (as well as individuals and datatypes) with lowercase letters.
- Use CamelCase for enhanced readability of composite entity names. E.g., use AgentRole rather than Agentrole, and use hasQuantityValue rather than hasquantityvalue.
- Use singular class names, e.g., Person instead of Persons.
- Use class names that are specific, and that help to avoid common misunderstandings. For example,

use ActorRole instead of Actor, to avoid accidental subClassing with Person.

- Whenever possible, use directional property names, and in particular avoid using nouns as property names. E.g., use hasQuantityValue instead of quantityvalue. The inverse property could then be consistently names as quantityValueOf. Other examples would be providesAgentRole and assumesAgentRole.
- Make particularly careful choices concerning property names, and that they are consistent with the domain and range axioms chosen. E.g., a hasName property should probably never have a domain (other than owl:Thing), as many things can indeed have names.

It is helpful to keep these conventions in mind from the very start. However, during actual modeling sessions, it is often better to focus more on the structure of the schema diagram that is being designed, and to delay a discussion on most appropriate names for ontology entities. These can be relatively easily changed during the documentation phase.

3.6.9. Create OWL file(s)

Creation of the OWL file can be done using CoModIDE (discussed below). The work could be done in parallel with writing up the documentation; however we describe it as the last point in order to emphasize that most of the work on a modular ontology is done conceptually, using discussions, diagrams, and documentation; and that the formal model, in form of an OWL file, is really only the final step in the creation.

For the sake of future maintainability, the generated OWL file should incorporate OPLa annotations that identify modules and their provenance; such annotations are created by CoModIDE.

4. CoModIDE

CoModIDE is intended to simplify MOMo-based ontology engineering projects. Per the MOMo methodology, initial modeling rarely needs to (or should) make use of the full set of language constructs that OWL 2 provides; instead, at these early stages of the process, work is typically carried out graphically – whether that be on whiteboards, in vector drawing software, or even on paper. This limits the modeling constructs to those that can be expressed intuitively using graphical notations, i.e., schema diagrams[8], as discussed above.

Per MOMo, the formalization of the developed solution into an OWL ontology is carried out after-the-fact, by a designated ontologist with extensive knowledge of both the language and applicable tooling. However, this comes at a cost, both in terms of hours expended, and in terms of the risk of incorrect interpretations of the previously drawn graphical representations (the OWL standard does not define a graphical notation syntax, so such representations are sometimes ambiguous). CoModIDE intends to reduce costs by bridging this gap, by providing tooling that supports both user-friendly schema diagram composition (using both ODP-based modules and "free-hand" modeling of classes and relationships), and direct OWL file generation.

4.1. Design and Features

The design criteria for CoModIDE, derived from the requirements discussed above, are as follows:

- CoModIDE should support visual-first ontology engineering, based on a graph representation of classes, properties, and datatypes. This graphical rendering of an ontology built using CoModIDE should be consistent across restarts, machines, and operating system or Protégé versions.
- CoModIDE should support the type of OWL 2 constructs that can be easily and intuitively understood when rendered as a schema diagram. To model more advanced constructs (unions and intersections in property domains or ranges, the property subsumption hierarchy, property chains, etc), the user can drop back into the standard Protégé tabs.
- CoModIDE should embed an ODP repository. Each included ODP should be free-standing and completely documented. There should be no external dependency on anything outside of the user's machine[9]. If the user wishes, they should be able to load a separately downloaded ODP

[8] We find that the size of partial solutions users typically develop fit on a medium-sized whiteboard; but whether this is a naturally manageable size for humans to operate with, or whether it is the result of constraints of or conditioning to the available tooling, i.e., the size of the whiteboards often mounted in conference rooms, we cannot say.

[9] Our experience indicates that while our target users are generally enthusiastic about the idea of reusing design patterns, they are

repository, to replace or complement the built-in one.
- CoModIDE should support simple composition of ODPs; patterns should snap together like Lego blocks, ideally with potential connection points between the patterns lighting up while dragging compatible patterns. The resulting ontology modules should maintain their coherence and be treated like modules in a consistent manner across restarts, machines, etc. A pattern or ontology interface concept will need be developed to support this.

CoModIDE is developed as a plugin to the versatile and well-established Protégé ontology engineering environment. The plugin provides three Protégé views, and a tab that hosts these views (see Figure 10). The *schema editor* view provides an a graphical overview of an ontology's structure, including the classes in the ontology, their subclass relations, and the object and datatype properties in the ontology that relate these classes to one another and to datatypes. All of these entities can be manipulated graphically through dragging and dropping. The *pattern library* view provides a built-in copy of the MODL ontology design pattern library [18], sourced from various projects and from the ODP community wiki[10]. A user can drag and drop design patterns from the pattern library onto the canvas to instantiate those patterns as modules in their ontology. The *configuration* view lets the user configure the behavior of the other CoModIDE views and their components. For a detailed description, we refer the reader to the video walkthrough on the CoModIDE webpage[11]. We also invite the reader to download and install CoModIDE themselves, from that same site.

When a pattern is dragged onto the canvas, the constructs in that pattern are copied into the ontology (optionally having their IRIs updated to correspond with the target ontology namespace), but they are also annotated using the OPLa vocabulary, to indicate 1) that they belong to a certain pattern-based module, and 2) what pattern that module implements. In this way module provenance is maintained, and modules can be manipulated (folded, unfolded, removed, annotated) as needed.

quickly turned off of the idea when they are faced with patterns that lack documentation or that exhibit link rot.

[10]http://ontologydesignpatterns.org/

[11]https://comodide.com

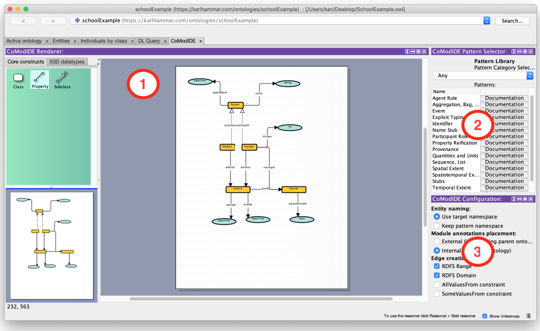

Fig. 10. CoModIDE User Interface featuring 1) the schema editor, 2) the pattern library, and 3) the configuration view.

4.2. Evaluation Method

We have evaluated CoModIDE through a four-step experimental setup, consisting of: a survey to collect subject background data (familiarity with ontology languages and tools), two modeling tasks, and a follow-up survey to collect information on the usability of both Protégé and CoModIDE. The tasks were designed to emulate a MOMo process, where a conceptual design is developed and agreed upon by whiteboard prototyping, and a developer is then assigned to formalizing the resulting whiteboard schema diagram into an OWL ontology. Our experimental hypotheses were defined as follows:

H1. When using CoModIDE, a user takes less time to produce correct and reasonable output, than when using Protege.

H2. A user will find CoModIDE to have a higher SUS score than when using Protege alone.

During each of the modeling tasks, participants were asked to generate a *reasonable* and *correct* OWL file for the provided schema diagram. In order to prevent a learning effect, the two tasks utilized two different schema diagrams. To prevent bias arising from differences in task complexity, counterbalancing was employed (such that half the users performed the first task with standard Protégé and the second task with CoModIDE, and half did the opposite). The correctness of the developed OWL files, and the time taken to complete each tasks, were recorded (the latter was however, for practical reasons, limited to 20 minutes per task).

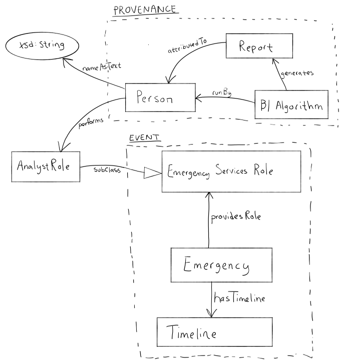

Fig. 11. Task A Schema Diagram

The following sections provide a brief overview of each the steps. The source material for the entire experiment is available online.[12]

4.2.1. *Introductory Tutorial*

When recruiting our participants for this evaluation, we did not place any requirements on ontology modeling familiarity. However, to establish a shared baseline knowledge of foundational modeling concepts (such as one would assume participants would have in the MOMo scenario we try to emulate, see above), we provided a 10 minute tutorial on ontologies, classes, properties, domains, and ranges. The slides used for this tutorial may be found online with the rest of the experiment's source materials.

4.2.2. a priori *Survey*

The purpose of the *a priori* survey was to collect information relating to the participants base level familiarity with topics related to knowledge modeling, to be used as control variables in later analysis. We used a 5-point Likert scale for rating the accuracy of the following statements.

CV1. I have done ontology modeling before.

CV2. I am familiar with Ontology Design Patterns.

CV3. I am familiar with Manchester Syntax.

CV4. I am familiar with Description Logics.

CV5. I am familiar with Protégé.

Finally, we asked the participants to describe their relationship to the test leader, (e.g. student, colleague, same research lab, not familiar).

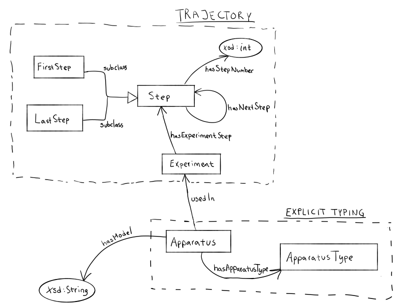

Fig. 12. Task B Schema Diagram

4.2.3. *Modeling Task A*

In Task A, participants were to develop an ontology to model how an analyst might generate reports about an ongoing emergency. The scenario identified two design patterns to use:

- **Provenance**: to track who made a report and how;
- **Event**: to capture the notion of an emergency.

Figure 11 shows how these patterns are instantiated and connected together. Overall the schema diagram contains seven concepts, one datatype, one subclass relation, one data property, and six object properties.

4.2.4. *Modeling Task B*

In Task B, participants were to develop an ontology to capture the steps of an experiment. The scenario identified two design patterns to use:

- **Trajectory**: to track the order of the steps;
- **Explicit Typing**: to easily model different types of apparatus.

Figure 12 shows how these patterns are instantiated and connected together. Overall, the schema diagram contains six concepts, two datatypes, two subclass relations, two data properties, and four object properties (one of which is a self-loop).

4.2.5. a posteriori *Survey*

The *a posteriori* survey included the SUS evaluations for both Protégé and CoModIDE. The SUS is a very common "quick and dirty," yet reliable tool for measuring the usability of a system. It consists of ten questions, the answers to which are used to compute a total usability score of 0–100. Additional information on the SUS and its included questions can be found online.[13]

[12] http://urn.kb.se/resolve?urn=urn:nbn:se:hj:diva-47887

[13] https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html

Table 1
Mean, standard deviation, relative standard deviation, and median responses to *a priori* statements

	mean	σ	relative σ	median
CV1	3.05	1.75	57 %	3
CV2	3.05	1.32	43 %	3
CV3	2.33	1.56	67 %	1
CV4	2.81	1.33	47 %	3
CV5	2.95	1.63	55 %	3

Additionally, we inquire about CoModIDE-specific features. These statements are also rated using a Likert scale. However, we do not use this data in our evaluation, except to inform our future work.. Finally, we requested any free-text comments on CoModIDE's features.

4.3. Results

4.3.1. Participant Pool Composition

Of the 21 subjects, 12 reported some degree of familiarity with the authors, while 9 reported no such connection. In terms of self-reported ontology engineering familiarity, the responses are as detailed in Table 1. It should be observed that responses vary widely, with a relative standard deviation (σ/mean) of 43–67 %.

4.3.2. Metric Evaluation

We define our two metrics as follows:

- **Time Taken:** number of minutes, rounded to the nearest whole minute and capped at 20 minutes due to practical limitations, taken to complete a task;
- **Correctness of Output** is a discrete measure that corresponds to the structural accuracy of the output. That is, 2 points were awarded to structurally accurate OWL files; 1 point for a borderline case (e.g one or two incorrect linkages, or missing a domain statement but including the range); and 0 points for any other output.

For these metrics, we generate simple statistics that describe the data, per modeling task. Tables 2a and 2b show the mean, standard deviation, and median for the Time Taken and Correctness of Output, respectively.

In addition, we examine the impact of our control variables (CV). This analysis is important, as it provides context for representation or bias in our data set. These are reported in Table 2c. CV1-CV5 correspond exactly to those questions asked during the *a priori* Survey, as described in Section 4.2. For each CV, we calculated the bivariate correlation between the sample data and the self-reported data in the survey. We believe that this is a reasonable measure of impact on effect, as our limited sample size is not amenable to partitioning. That is, the partitions (as based on responses in the *a priori* survey) could have been tested pair-wise for statistical significance. Unfortunately, the partitions would have been too small to conduct proper statistical testing. However, we do caution that correlation effects are strongly impacted by sample size.

We analyze the SUS scores in the same manner. Table 4 presents the mean, standard deviation, and median of the data set. The maximum score while using the scale is a 100. Table 2d presents our observed correlations with our control variables.

Finally, we compare the each metric for one tool against the other. That is, we want to know if our results are statistically significant—that as the statistics suggest in Table 2, CoModIDE does indeed perform better for both metrics and the SUS evaluation. To do so, we calculate the probability p that the samples from each dataset come from different underlying distributions. A common tool, and the tool we employ here, is the Paired (two-tailed) T-Test—noting that it is reasonable to assume that the underlying data are normally distributed, as well as powerful tool for analyzing datasets of limited size. The threshold for indicating confidence that the difference is significant is generally taken to be $p < 0.05$. Table 3 summarizes these results.

4.3.3. Free-text Responses

18 of the 21 subjects opted to leave free-text comments. We applied fragment-based qualitative coding and analysis on these comments. I.e., we split the comments apart per the line breaks entered by the subjects, we read through the fragments and generated a simple category scheme, and we then re-read the fragments and applied these categories to the fragments (allowing at most one category per fragment) [51, 52]. The subjects left between 1–6 fragments each for a total of 49 fragments for analysis, of which 37 were coded, as detailed in Table 5.

Of the 18 participants who left comments, 3 left comments containing no codable fragments; these either commented upon the subjects own performance in the experiment, which is covered in the aforementioned completion metrics, or were simple statements of fact

Table 2

Summary of statistics comparing Protege and CoModIDE.

	mean	σ	median
Protégé	17.44	3.67	20.0
CoModIDE	13.94	4.22	13.5

(a) Mean, standard deviation, and median *time taken* to complete each modeling task.

	mean	σ	median
Protégé	0.50	0.71	0.0
CoModIDE	1.33	0.77	1.5

(b) Mean, standard deviation, and median *correctness of output* for each modeling task.

	CV1	CV2	CV3	CV4	CV5
TT (P)	-0.61	-0.18	-0.38	-0.58	-0.62
Cor. (P)	0.50	0.20	0.35	0.51	0.35
TT (C)	0.02	-0.34	-0.28	-0.06	0.01
Cor. (C)	-0.30	0.00	-0.12	-0.33	-0.30

(c) Correlations control variables (CV) on the Time Taken (TT) and Correctness of Output (Cor.) for both tools Protégé (P) and CoModIDE (C).

	CV1	CV2	CV3	CV4	CV5
SUS (P)	0.70	0.52	0.64	0.73	0.64
SUS (C)	-0.34	-0.05	-0.08	-0.29	-0.39

(d) Correlations with control variables (CV) on the SUS scores for both tools Protégé (P) and CoModIDE (C).

(e.g., *"In order to connect two classes I drew a connecting line"*).

Table 3

Significance of results.

Time Taken	Correctness	SUS Evaluation
$p \approx 0.025 < 0.05$	$p \approx 0.009 < 0.01$	$p \approx 0.0003 < 0.001$

Table 4

Mean, standard deviation, and median SUS score for each tool. The maximum score is 100.

	mean	σ	median
Protégé	36.67	22.11	35.00
CoModIDE	73.33	16.80	76.25

Table 5

Free text comment fragments per category

Code	Fragment #
Graph layout	4
Dragging & dropping	6
Feature requests	5
Bugs	8
Modeling problems	5
Value/preference statements	9

4.4. Discussion

4.4.1. Participant Pool Composition

The data indicates no correlation (bivariate correlation $< \pm 0.1$) between the subjects' reported author familiarity, and their reported SUS scores, such as would have been the case if the subjects who knew the authors were biased. The high relative standard deviation for a priori knowledge level responses indicates that our subjects are rather diverse in their skill levels. As discussed below, this variation is fortunate as it allows us to compare the performance of more or less experienced users.

4.4.2. Metric Evaluation

Before we can determine if our results confirm H1 and H2, we must first examine the correlations between our results and the control variables gathered in the *a priori* survey. In this context, we find it reasonable to use these thresholds for a correlation $|r|$: 0-0.19 very weak, 0.20-0.39 weak, 0.40-0.59 moderate, 0.60-0.79 strong, 0.80-1.00 very strong.

As shown in Table 2c, the metric *time taken* when using Protégé is negatively correlated with each CV. The *correctness* metric is positively correlated with each CV. This is unsurprising and reasonable; it indicates that familiarity with the ontology modeling, related concepts, and Protégé improves (shortens) time taken to complete a modeling task and improves the correctness of the output. However, for the metrics pertaining to CoModIDE, there are only very weak and three weak correlations with the CVs. We may construe this to mean that performance when using CoModIDE, with respect to our metrics, is largely agnostic to our control variables.

To confirm H1, we look at the metrics separately. *Time taken* is reported better for CoModIDE in both mean and median. When comparing the underlying data, we achieve $p \approx 0.025 < 0.05$. Next, in comparing the *correctness* metric from Table 2b, CoModIDE again outperforms Protégé in both mean and median. When comparing the underlying data, we achieve a statistical significance of $p \approx 0.009 < 0.01$. With these together, we reject the null hypothesis and confirm H1.

This is particularly interesting; given the above analysis of CV correlations where we see no (or very weak) correlations between prior ontology modeling familiarity and CoModIDE modeling results, and the confirmation of H1, that CoModIDE users perform better than Protégé users, we have a strong indicator that we have achieved increased *approachability*.

When comparing the SUS score evaluations, we see that the usability of Protégé is strongly influenced by familiarity with ontology modeling and familiarity with Protégé itself. The magnitude of the correlation suggests that newcomers to Protege do not find it very usable. CoModIDE, on the other hand is weakly, negatively correlated along the CV. This suggests that switching to a graphical modeling paradigm may take some adjusting.

However, we still see that the SUS scores for CoModIDE have a greater mean, tighter σ, and greater median, achieving a very strong statistical significance $p \approx 0.0003 < 0.001$. Thus, we may reject the null hypothesis and confirm H2.

As such, by confirming H1 and H2, we may say that CoModIDE improves the approachability of ontology engineering, especially for those not familiar with ontology modeling—with respect to our participant pool. However, we suspect that our results are generalizable, due to the strength of the statistical significance (Table 3) and participant pool composition (Section 4.3.1).

4.4.3. Free-text Responses

The fragments summarized in Table 5 paints a quite coherent picture of the subjects' perceived advantages and shortcomings of CoModIDE, as follows:

- *Graph layout:* The layout of the included MODL patterns, when dropped on the canvas, is too cramped and several classes or properties overlap, which reduces tooling usability.
- *Dragging and dropping:* Dragging classes was hit-and-miss; this often caused users to create new properties between classes, not move them.
- *Feature requests:* Pressing the "enter" key should accept and close the entity renaming window. Zooming is requested, and an auto-layout button.
- *Bugs:* Entity renaming is buggy when entities with similar names exist.
- *Modeling problems:* Self-links/loops cannot easily be modeled.
- *Value/preference statements:* Users really appreciate the graphical modeling paradigm offered, e.g., *"Much easier to use the GUI to develop ontologies"*, *"Moreover, I find this system to be way more intuitive than Protégé"*, *"CoModIDE was intuitive to learn and use, despite never working with it before."*

We note that the there is a near-unanimous consensus among the subjects that graphical modeling is intuitive and helpful. When users are critical of the CoModIDE software, these criticisms are typically aimed at specific and quite shallow bugs or UI features that are lacking. The only consistent criticism of the modeling method itself relates to the difficulty in constructing self-links (i.e., properties that have the same class as domain and range).

4.5. CoModIDE 2.0

Since the evaluation, we have made plenty of progress on improving CoModIDE in significant ways. Aside from bug fixes and general quality of life improvements (i.e. versions 1.1.1 and 1.1.2) addressing many of the free-text responses in Section 4.4.3, we have implemented additional key aspects of the MOMo methodology. In particular, they are as follows.

- **Modules** are now directly supported. The grey boxes, as shown in Figure 5, can now be created by highlighting a group of connected classes and datatypes and pressing 'G'. These new nodes can be folded into a single cell in order to simplify large or complex diagrams. Outgoing edges are maintained from the collapsed node. Newly instantiated patterns (i.e. those dragged-and-dropped from the pattern library) appear pre-grouped into modules.

- **OPLa Annotations** are added whenever modules are created directly, and are properly retained when dragged-and-dropped from the pattern library. In particular, the isNativeTo and reusesPatternAsTemplate properties are currently supported. This generally subsumes the functionality of [19].
- The **Systematic Axiomatization** process from MOMo is now directly supported. By clicking on a named edge on the canvas, the user can now customize exactly what the edge represents in the "Edge Inspection Tool." The list offers the human readable labels for the list of axioms generally used in the MOMo workflow, and described in Section 3.6.6.

We have also added functionality to assist in navigating a complex pattern space through the notion of interfaces. That is, categorizing patterns based on the roles that they may play. For example, a more general ontology may call for some pattern that satisfies a spatial extent modeling requirement. To borrow from software engineering terms, one could imagine several different implementations of a 'spatial extent" interface.

In addition, we have added simple, manual alignment to external ontologies. More information on this upper alignment tool for CoModIDE can be found in [53].

In order to improve the extensibility of the platform, we have reworked the overarching conceptual framework for functionality in CoModIDE. Functionality is now categorized into so-called toolkits which communicate through a newly implemented message bus. This allows for a relatively straightforward integration process for external developers.

It is also important to recall that CoModIDE is not just a development platform, but a tool that enables research into ontology engineering. To that point, we have implemented an opt-in telemetry agent that collects and sends anonymized usage characteristics back to the developers. This records session lengths, clicks, and other such metrics that give us insight on how ontologies are authored in a graphical environment.

5. Additional Infrastructure and Resources

5.1. The Modular Ontology Design Library (MODL)

The Modular Ontology Design Library (MODL) is both an artifact and a framework for creating collections of ontology design patterns [18]. MODL is a method for establishing a well-curated and well-documented collection of ODPs that is structured using OPLa. This allows for a queryable interface when the MODL is very large, or if the MODL is integrated into other tooling infrastructure. For example, CoModIDE uses OPLa annotations to structure, define, and relate patterns in its internal MODL, as described in Section 4.1.

5.2. OPLa Annotator

The OPLa Annotator [19] is a standalone plugin for Protégé. This plug-in allows for the guided creation of opla:isNativeTo annotations on ontological entities of an OWL file. While this particular functionality is subsumed in CoModIDE, it does not require the graphical canvas or the creation of modules, and can be a quicker option when the imposed graphical organization is not desired or required.

5.3. ROWLTab

ROWLTab [47] is another standalone plugin for Protégé. It is based on the premise that some ontology users, and frequently non-ontologists, find conceptualizing knowledge through rules to be more convenient. This plugin allows the user to enter SWRL rules which will then, when applicable, be converted into equivalent OWL axioms. An extension to this plug-in, detailed in [54], allows for existential rules.

5.4. SDOnt

SDOnt [17] is an early tool for generating schema diagrams from OWL files. Unlike other visual OWL generators, SDOnt does not give a strictly disambiguous diagram. Instead, it generates schema diagrams in the style that has been described in Section 3.2 based on the TBox of the input OWL file. This program only requires Java to run and can be run on any OWL ontology; although, as with any graph visualization, it tends to work best with smaller schemas.

6. Experiences

In this section, we briefly detail some of our experiences executing the MOMo workflow, in particular the interesting differences across domains, project requirements, and virtual meeting requirements.[14]

6.1. Pattern Modeling at VoCamps

VoCamps have a rich history in the Semantic Web community.[15] These workshops began as a way to create community-driven taxonomies, but have since evolved to include the development of ontology design patterns and modules for particular domains. VoCamps are very similar to the MOMo methodology, but generally focus on the first few steps, those relating to the use case, requirements, competency questions, and the identification and development of ontology design patterns.

Overall, the challenges that we encountered generally concerned preparation. Overcoming the obstacles of unclear or fuzzy requirements and desired outcomes resulted in our more structured, iterative schema diagrammatic approach as described earlier in this paper. We also observed that there is much difficulty in finding a balance between developing a general patterns that cut across domains and one that immediately satisfies a specific application. One approach is to develop the module, and then work backwards from there to create a more general pattern (e.g. [55]). In other VoCamps we also found great value in trying to find common modeling needs across different projects and working together to find a pattern that can be easily differentiated to individual use cases; an example can be found in [56]. The major difference between these approaches is the team composition. The former was one ontologist and two chemists on a relatively small project. The latter was a collaboration between three different teams on a high profile, nationally funded initiative.

6.2. The GeoLink Modular Oceanography Ontology

The GeoLink project generally concerned the fusion of heterogeneous data sources pertaining to the ocean science domain [57]. The resulting ontology represents our conceptualization of modular ontology in its infancy. Its development process served as the framework for the MOMo workflow. Additionally, an important takeaway from this project was recognizing the difference between the needs of data owners, producers, and consumers. In particular, a data model needs to be sufficiently complex in order to be useful, but also relatively straightforward to populate and comprehend at varying levels of abstraction. Thus we developed and utilized the concept of shortcuts, which allow for more complex model to be simplified for the purposes of data population. Some examples can be found in [49, 58].

6.3. The Enslaved Ontology

The Enslaved Project[16] has been an extremely formative experience for the MOMo methodology. The overarching motivation and deliverable is described in [43]. Here, we describe a number of the experiences that directly impacted our ontology engineering process. A fully indepth description of our experiences is also reported in [43].

The project team is large with many different roles. It, perhaps, comes as no surprise that it was important to understand and agree on the different roles and responsibilities held by participating members. However, this is not altogether obvious, as it is unfortunately difficult to predict which expertise and knowledge is required during modeling. As such, clear delineation of roles, responsibilities, and communication of domain expertise are necessary for a successful modeling session.

In terms of modeling, the project was interesting for a number of reasons. The first was that, being centered around historical data and the interpretations of historians, the ground truth was elusive. As such, we took a records-based approach where provenance directly modeled and a first-class citizen. This could not be directly addressed by existing patterns, and needed directly modeled from scratch. During this process, we followed the MOMo principles for using schema diagrams as the primary conceptual vehicle; in doing so, we were able to narrow down the records-based approach by slowly matching our conceptualization to how the digital historians view their own data.

As part of the overarching project, the ontology is to be used as a schema for a knowledge graph that will be persisted using Wikibase[17] software. This means that we needed to build a way to map between the con-

[14] The virtual meeting requirements are largely an artifact of the lockdown measures in response to the COVID-19 pandemic (2020), but serve to inform any virtual, collaborative modeling sessions moving forward.

[15] See https://vocamp.org/.

[16] See https://enslaved.org/.

[17] See https://wikiba.se/.

ceptual schema represented by the Enslaved Ontology and the underlying Wikibase schema for the data. The major difference lays in the notion of provenance. The mapping is reported in [59]. This effort indicates that with some effort initial modeling can be done such that the mapping becomes natural.

Finally, we have demonstrated the evolvability of a modular ontology by completely replacing the original Place module with a much more sophisticated version that more closely resembles the data reality, as well as improving the user experience in navigating the knowledge graph.

6.4. The Modular Ontology for Space Weather Research

This preliminary modular ontology can be found in [56]. The development of this modular ontology is still ongoing, but while executing MOMo, we have identified additional conceptual and developmental tools. As discussed in Step 1 (Section 3.6.1), an important aspect of modeling is also understanding exactly the needs a knowledge graph is to support. As such, we have found value in adapting certain mechanisms from software engineering. By graphically modeling the interactions between users and the knowledge graph, such as through flow charts, data flow diagrams, and other such artifacts, we found it easier to understand and differentiate use case scenarios for the knowledge graph. We additionally, in line with experiences outlined in [47] and in combination with the ExplicitTyping pattern from MODL, utilized existential rules to more transparently encode domain knowledge into the knowledge graph.

6.5. Food Traceability

The Modular Ontology for Grain Supply Chain Tracing is under active development, and in doing so we are executing MOMo. The particular experience that we wish to report here pertains to our transition from in person modeling to virtual meetings. This transition really drove home that the schema diagram, as the primary conceptual vehicle for communication between ontology engineers and domain experts, is very effective. In particular, it solidified the iterative process of ontology engineers interrogating domain experts and then drafting (or adjusting) schema diagrams to represent that knowledge, iterating as necessary. Axiomatization only occurs after the majority of the top-level conceptual modeling has been completed satisfactorily.

6.6. The RealEstateCore Ontology

The RealEstateCore ontology is developed by a consortium of academic organisations, software developers, and real estate owners, to support data integration and reasoning in smart building scenarios [60]. It covers domains such as building component topology (closely related to CAD/BIM disciplines and models), description and operation of building management systems (e.g., HVAC, access control, lighting, etc.), newer internet-connected IoT-type sensor equipment, and the administrative processes related to the operation of a building. At the time of writing, RealEstateCore has been selected as the standard ontology for, and is being deployed within, several large property owners in northern Europe, including Vasakronan, YIT, Brunswick, Entra, Akademiska Hus, and Castellum; for a total deployment size of well north of 10 million square meters. Additionally, through a collaboration with Microsoft, RealEstateCore has been ported to the Digital Twins Definition Language, where it is the recommended smart building ontology for the Azure Digital Twins PaaS product.[18]

In building RealEstateCore, the development process has been shaped by the need to rapidly reach a state where the ontology can be deployed and used by domain experts and software developers with limited ontology engineering experience. For this to work, we have found great benefit in adhering to a set of development practices mirroring those proposed in MOMo:

- Development has been directly based on the businesses' use cases and existing data sources.
- Modularity has been employed as a way of encapsulating complexity and dividing the domain into manageable parts.
- Established model designs from the field have been reused as far as possible, reducing both development time and barrier-to-entry for domain experts familiar with established practices in the field.
- White-board prototyping using schema diagrams has been used extensively in knowledge elicitation with domain experts, prior to creating formal OWL representations.

[18] https://github.com/Azure/opendigitaltwins-building

Some tensions in the modularity and reuse aspects were observed, since existing reusable ontologies (or other industry standard models) typically have not been developed with modularity or ease-of-modularization in mind. Consequently, the modules that we have developed for RealEstateCore have been substantially larger than the ones proposed by MOMo, typically encompassing entire sub-domains of the problem space, rather than individual key concepts. This design has also been approachable to our domain experts, since their overall conceptual view of the problem has not had to be divided into too many small building blocks. We have however found that the finer-granularity guidance provided by ODPs has also been very helpful in modeling individual problems. We have thus employed ODP-based design consistently (using MOMo-style template-based ODP instantiation), but we have not encapsulated ODP instantiations into modules. As the ontology grows and evolves, its maintainability would benefit if modularity annotations (e.g., using OPLa) were added to it – we hope to develop this in a future iteration.

We found the use of pattern-based schema diagrams for prototyping to be an extraordinarily powerful method of fleshing out ideas and developing a developer/expert consensus. For any major feature development, we have worked by developing joint conceptual models at hackathon-style events, based on traditional whiteboards, which an ontology expert has thereafter taken and turned into OWL files, in line with the MOMo methodology. For visualizing and double-checking the whole or partial designs with domain experts, CoModIDE and the leading visualizer WebVOWL have both been used; while WebVOWL was more suitable when viewing the whole ontology, CoModIDE fared well for many smaller parts of the ontology, e.g., those developed as result of a single whiteboard session.

We must however also caution that we experienced some cases of over-dependence on design patterns; e.g., in cases where strict adherence to pattern design in some detailed partial solution might be detrimental to the larger-granularity design. One example was the use of the *Explicit Typing* ODP from the MODL library [18]; the use of which (until it was replaced) complicated the overall class subsumption hierarchy. For further discussion on the lessons learned during this work, see [61].

7. Conclusion

The re-use of ontologies for new purposes, or adapting them to new use-cases, is frequently very difficult. In our experiences, we have found this to be the case for several reasons: (i) differing representational granularity, (ii) lack of conceptual clarity of the ontology design, (iii) adhering to good modeling principles, and (iv) a lack of re-use emphasis and process support available in ontology engineering tooling. In order to address these concerns, we have developed the Modular Ontology Modeling (MOMo) workflow and supporting tooling infrastructure, CoModIDE (The Comprehensive Modular Ontology Integrated Development Environment – "commodity").

In this paper, we have presented the MOMo workflow in detail, from introducing the schema diagram as the primary conceptual vehicle for communicating between ontology engineers and domain experts, to presenting several experiences in executing the workflow across many distinct domains with different use cases and data requirements.

We have also shown how the CoModIDE platform allows ontology engineers, irrespective of previous knowledge level, to develop ontologies more correctly and more quickly, than by using standard Protégé; that CoModIDE has a higher usability (SUS score) than standard Protégé; and that the CoModIDE issues that concern users primarily derive from shallow bugs as opposed to methodological or modeling issues. Taken together, this implies that the modular graphical ontology engineering paradigm is a viable way for supporting the MOMo workflow.

7.1. Future Work

From here, there are still many avenues of investigation remaining, pertaining to both the MOMo workflow and CoModIDE.

Regarding the workflow, we will continue to execute the workflow in new domains and observe differences in experiences. Currently, we are examining how to better incorporate spatially-explicit modeling techniques. In addition, we wish to further explore how schema diagrams may represent distinctly different semantics, such as ShEx [62], SHACL [63], rather than OWL.

We also foresee the continued development of the platform. As mentioned in Section 4.5, we have improved its internal structure so that it may support bundled pieces of functionality. In particular, we will develop such toolkits for supporting holistic ontology engineering projects, going beyond just the modeling process. This will include the incorporation of ontology alignment systems so that CoModIDE may export automatic alignments alongside the designed deliverable, and the incorporation of recommendation software, perhaps based on input seed data. Further, we see a route for automatic documentation in the style of our own technical reports. Finally, we wish to examine collected telemetry data in order to analyse how users develop ontologies in a graphical modeling paradigm.

Acknowledgements. Cogan Shimizu and Pascal Hitzler acknowledge partial support from the financial assistance award 70NANB19H094 from U.S. Department of Commerce, National Institute of Standards and Technology and the National Science Foundation under Grant No. 2033521 and the Andrew W. Mellon Foundation. The authors acknolwedge the helpful input from Adila Krisnadhi and MD Kamruzzaman Sarker; and development support from Sulogna Chowdhury, Abhilekha Dalal, Riley Mueller, and Lu Zhou.

References

[1] Gene Ontology Consortium: The Gene Ontology (GO) database and informatics resource, *Nucleic Acids Research* **32**(Database–Issue) (2004), 258–261. doi:10.1093/nar/gkh036.

[2] P. Hitzler, Semantic Web: A Review of the Field, *Communications of the ACM* (2021), to appear.

[3] A. Algergawy, D. Faria, A. Ferrara, I. Fundulaki, I. Harrow, S. Hertling, E. Jiménez-Ruiz, N. Karam, A. Khiat, P. Lambrix, H. Li, S. Montanelli, H. Paulheim, C. Pesquita, T. Saveta, P. Shvaiko, A. Splendiani, É. Thiéblin, C. Trojahn, J. Vatascinová, O. Zamazal and L. Zhou, Results of the Ontology Alignment Evaluation Initiative 2019, in: *Proceedings of the 14th International Workshop on Ontology Matching co-located with the 18th International Semantic Web Conference (ISWC 2019), Auckland, New Zealand, October 26, 2019*, P. Shvaiko, J. Euzenat, E. Jiménez-Ruiz, O. Hassanzadeh and C. Trojahn, eds, CEUR Workshop Proceedings, Vol. 2536, CEUR-WS.org, 2019, pp. 46–85.

[4] O. Zamazal and V. Svátek, The Ten-Year OntoFarm and its Fertilization within the Onto-Sphere, *J. Web Semant.* **43** (2017), 46–53. doi:10.1016/j.websem.2017.01.001.

[5] A. Algergawy, S. Babalou, F. Klan and B. König-Ries, Ontology Modularization with OAPT, *J. Data Semant.* **9**(2) (2020), 53–83. doi:10.1007/s13740-020-00114-7.

[6] B.C. Grau, I. Horrocks, Y. Kazakov and U. Sattler, Extracting Modules from Ontologies: A Logic-Based Approach, in: *Modular Ontologies: Concepts, Theories and Techniques for Knowledge Modularization*, H. Stuckenschmidt, C. Parent and S. Spaccapietra, eds, Lecture Notes in Computer Science, Vol. 5445, Springer, 2009, pp. 159–186. doi:10.1007/978-3-642-01907-4_8.

[7] H. Stuckenschmidt, C. Parent and S. Spaccapietra (eds), *Modular Ontologies: Concepts, Theories and Techniques for Knowledge Modularization*, Lecture Notes in Computer Science, Vol. 5445, Springer, 2009. ISBN 978-3-642-01906-7. doi:10.1007/978-3-642-01907-4.

[8] D. Osumi-Sutherland, M. Courtot, J.P. Balhoff and C.J. Mungall, Dead simple OWL design patterns, *J. Biomed. Semant.* **8**(1) (2017), 18:1–18:7. doi:10.1186/s13326-017-0126-0.

[9] M.G. Skjæveland, D.P. Lupp, L.H. Karlsen and H. Forssell, Practical Ontology Pattern Instantiation, Discovery, and Maintenance with Reasonable Ontology Templates, in: *The Semantic Web - ISWC 2018 - 17th International Semantic Web Conference, Monterey, CA, USA, October 8-12, 2018, Proceedings, Part I*, D. Vrandecic, K. Bontcheva, M.C. Suárez-Figueroa, V. Presutti, I. Celino, M. Sabou, L. Kaffee and E. Simperl, eds, Lecture Notes in Computer Science, Vol. 11136, Springer, 2018, pp. 477–494. doi:10.1007/978-3-030-00671-6_28.

[10] E. Blomqvist and K. Sandkuhl, Patterns in Ontology Engineering: Classification of Ontology Patterns, in: *ICEIS 2005, Proceedings of the Seventh International Conference on Enterprise Information Systems, Miami, USA, May 25-28, 2005*, C. Chen, J. Filipe, I. Seruca and J. Cordeiro, eds, 2005, pp. 413–416.

[11] A. Gangemi, Ontology Design Patterns for Semantic Web Content, in: *The Semantic Web - ISWC 2005, 4th International Semantic Web Conference, ISWC 2005, Galway, Ireland, November 6-10, 2005, Proceedings*, Y. Gil, E. Motta, V.R. Benjamins and M.A. Musen, eds, Lecture Notes in Computer Science, Vol. 3729, Springer, 2005, pp. 262–276. doi:10.1007/11574620_21.

[12] P. Hitzler, A. Gangemi, K. Janowicz, A. Krisnadhi and V. Presutti (eds), *Ontology Engineering with Ontology Design Patterns – Foundations and Applications*, Studies on the Semantic Web, Vol. 25, IOS Press, 2016.

[13] Y. Hu, K. Janowicz, D. Carral, S. Scheider, W. Kuhn, G. Berg-Cross, P. Hitzler, M. Dean and D. Kolas, A geo-ontology design pattern for semantic trajectories, in: *International Conference on Spatial Information Theory*, T. Tenbrink, J. Stell, A. Galton and Z. Wood, eds, Springer, 2013, pp. 438–456.

[14] K. Hammar and V. Presutti, Template-Based Content ODP Instantiation, in: *Advances in Ontology Design and Patterns [revised and extended versions of the papers presented at the 7th edition of the Workshop on Ontology and Semantic Web Patterns, WOP@ISWC 2016, Kobe, Japan, 18th October 2016]*, K. Hammar, P. Hitzler, A. Krisnadhi, A. Lawrynowicz, A.G. Nuzzolese and M. Solanki, eds, Studies on the Semantic Web, Vol. 32, IOS Press, 2016, pp. 1–13. doi:10.3233/978-1-61499-826-6-1.

[15] E. Blomqvist, K. Hammar and V. Presutti, Engineering Ontologies with Patterns - The eXtreme Design Methodology, in: *Ontology Engineering with Ontology Design Patterns - Foundations and Applications*, P. Hitzler, A. Gangemi, K. Janowicz, A. Krisnadhi and V. Presutti, eds, Studies on the Semantic

Web, Vol. 25, IOS Press, 2016, pp. 23–50. doi:10.3233/978-1-61499-676-7-23.

[16] C. Shimizu, K. Hammar and P. Hitzler, Modular Graphical Ontology Engineering Evaluated, in: *The Semantic Web - 17th International Conference, ESWC 2020, Heraklion, Crete, Greece, May 31-June 4, 2020, Proceedings*, A. Harth, S. Kirrane, A.N. Ngomo, H. Paulheim, A. Rula, A.L. Gentile, P. Haase and M. Cochez, eds, Lecture Notes in Computer Science, Vol. 12123, Springer, 2020, pp. 20–35. doi:10.1007/978-3-030-49461-2_2.

[17] C. Shimizu, A. Eberhart, N. Karima, Q. Hirt, A. Krisnadhi and P. Hitzler, A Method for Automatically Generating Schema Diagrams for OWL Ontologies, in: *Knowledge Graphs and Semantic Web - First Iberoamerican Conference, KGSWC 2019, Villa Clara, Cuba, June 23-30, 2019, Proceedings*, B. Villazón-Terrazas and Y. Hidalgo-Delgado, eds, Communications in Computer and Information Science, Vol. 1029, Springer, 2019, pp. 149–161. doi:10.1007/978-3-030-21395-4_11.

[18] C. Shimizu, Q. Hirt and P. Hitzler, MODL: A Modular Ontology Design Library, in: *WOP@ISWC*, CEUR Workshop Proceedings, Vol. 2459, CEUR-WS.org, 2019, pp. 47–58.

[19] C. Shimizu, Q. Hirt and P. Hitzler, A Protégé Plug-In for Annotating OWL Ontologies with OPLa, in: *The Semantic Web: ESWC 2018 Satellite Events - ESWC 2018 Satellite Events, Heraklion, Crete, Greece, June 3-7, 2018, Revised Selected Papers*, A. Gangemi, A.L. Gentile, A.G. Nuzzolese, S. Rudolph, M. Maleshkova, H. Paulheim, J.Z. Pan and M. Alam, eds, Lecture Notes in Computer Science, Vol. 11155, Springer, 2018, pp. 23–27. doi:10.1007/978-3-319-98192-5_5.

[20] M. Fernández-López, A. Gómez-Pérez and N. Juristo, METHONTOLOGY: From Ontological Art Towards Ontological Engineering, Technical Report, SS-97-06, American Association for Artificial Intelligence, 1997.

[21] Y. Sure, S. Staab and R. Studer, On-To-Knowledge Methodology (OTKM), in: *Handbook on Ontologies*, S. Staab and R. Studer, eds, International Handbooks on Information Systems, Springer Berlin Heidelberg, 2004, pp. 117–132.

[22] H.S. Pinto, S. Staab, C. Tempich and Y. Sure, Distributed Engineering of Ontologies (DILIGENT), in: *Semantic Web and Peer-to-Peer: Decentralized Management and Exchange of Knowledge and Information*, S. Staab and H. Stuckenschmidt, eds, Springer, 2006, pp. 303–322.

[23] W.C. Mann and S.A. Thompson, *Rhetorical Structure Theory: A Theory of Text Organization*, Information Sciences Institute, University of Southern California, Los Angeles, 1987.

[24] M. Egaña, E. Antezana and R. Stevens, Transforming the Axiomisation of Ontologies: The Ontology Pre-Processor Language, in: *Proceedings of the Fourth OWLED Workshop on OWL: Experiences and Directions*, K. Clark and P.F. Patel-Schneider, eds, CEUR Workshop Proceedings, 2008.

[25] N. Karima, K. Hammar and P. Hitzler, How to Document Ontology Design Patterns, in: *Advances in Ontology Design and Patterns*, K. Hammar, P. Hitzler, A. Lawrynowicz, A. Krisnadhi, A. Nuzzolese and M. Solanki, eds, Studies on the Semantic Web, Vol. 32, IOS Press, Amsterdam, 2017, pp. 15–28.

[26] Q. Hirt, C. Shimizu and P. Hitzler, Extensions to the Ontology Design Pattern Representation Language, in: *WOP@ISWC*, CEUR Workshop Proceedings, Vol. 2459, CEUR-WS.org, 2019, pp. 76–75.

[27] P. Hitzler, A. Gangemi, K. Janowicz, A.A. Krisnadhi and V. Presutti, Towards a Simple but Useful Ontology Design Pattern Representation Language, in: *Proceedings of the 8th Workshop on Ontology Design and Patterns (WOP 2017) co-located with the 16th International Semantic Web Conference (ISWC 2017), Vienna, Austria, October 21, 2017.*, E. Blomqvist, Ó. Corcho, M. Horridge, D. Carral and R. Hoekstra, eds, CEUR Workshop Proceedings, Vol. 2043, CEUR-WS.org, 2017. http://ceur-ws.org/Vol-2043/paper-09.pdf.

[28] K. Hammar, Ontology Design Patterns in WebProtégé, in: *Proceedings of the ISWC 2015 Posters & Demonstrations Track co-located with the 14th International Semantic Web Conference (ISWC-2015), Bethlehem, PA, USA, October 11, 2015.*, S. Villata, J.Z. Pan and M. Dragoni, eds, CEUR Workshop Proceedings, Vol. 1486, CEUR-WS.org, 2015. http://ceur-ws.org/Vol-1486/paper_50.pdf.

[29] E. Blomqvist, A. Gangemi and V. Presutti, Experiments on pattern-based ontology design, in: *Proceedings of the fifth international conference on Knowledge capture*, 2009, pp. 41–48.

[30] E. Blomqvist, V. Presutti, E. Daga and A. Gangemi, Experimenting with eXtreme design, in: *International Conference on Knowledge Engineering and Knowledge Management*, Springer, 2010, pp. 120–134.

[31] K. Hammar, Ontology design patterns in use: lessons learnt from an ontology engineering case, in: *Workshop on Ontology Patterns in conjunction with the 11th International Semantic Web Conference 2012 (ISWC 2012)*, 2012.

[32] S. Peroni, A Simplified Agile Methodology for Ontology Development, in: *OWL: Experiences and Directions – Reasoner Evaluation*, M. Dragoni, M. Poveda-Villalón and E. Jimenez-Ruiz, eds, Lecture Notes in Computer Science, Vol. 10161, Springer, 2017, pp. 55–69.

[33] K. Hammar, *Content Ontology Design Patterns: Qualities, Methods, and Tools*, Vol. 1879, Linköping University Electronic Press, 2017.

[34] I. Hadar and P. Soffer, Variations in conceptual modeling: classification and ontological analysis, *Journal of the Association for Information Systems* **7**(8) (2006), 20.

[35] P. Shoval and I. Frumermann, OO and EER conceptual schemas: a comparison of user comprehension, *Journal of Database Management (JDM)* **5**(4) (1994), 28–38.

[36] S.S.-S. Cherfi, J. Akoka and I. Comyn-Wattiau, Conceptual modeling quality-from EER to UML schemas evaluation, in: *International Conference on Conceptual Modeling*, Springer, 2002, pp. 414–428.

[37] R. Agarwal and A.P. Sinha, Object-oriented modeling with UML: a study of developers' perceptions, *Communications of the ACM* **46**(9) (2003), 248–256.

[38] J. Krogstie, Evaluating UML using a generic quality framework, in: *UML and the Unified Process*, IGI Global, 2003, pp. 1–22.

[39] S. Lohmann, S. Negru, F. Haag and T. Ertl, Visualizing ontologies with VOWL, *Semantic Web* **7**(4) (2016), 399–419.

[40] S. Lohmann, V. Link, E. Marbach and S. Negru, WebVOWL: Web-based visualization of ontologies, in: *International Conference on Knowledge Engineering and Knowledge Management*, Springer, 2014, pp. 154–158.

[41] V. Wiens, S. Lohmann and S. Auer, WebVOWL Editor: Device-Independent Visual Ontology Modeling, in: *Proceedings of the ISWC 2018 Posters & Demonstrations, Industry*

and Blue Sky Ideas Tracks, CEUR Workshop Proceedings, Vol. 2180, 2018.

[42] M.K. Sarker, A.A. Krisnadhi and P. Hitzler, OWLAx: A Protégé Plugin to Support Ontology Axiomatization through Diagramming, in: *Proceedings of the ISWC 2016 Posters & Demonstrations Track co-located with 15th International Semantic Web Conference (ISWC 2016), Kobe, Japan, October 19, 2016.*, T. Kawamura and H. Paulheim, eds, CEUR Workshop Proceedings, Vol. 1690, CEUR-WS.org, 2016.

[43] C. Shimizu, P. Hitzler, Q. Hirt, D. Rehberger, S.G. Estrecha, C. Foley, A.M. Sheill, W. Hawthorne, J. Mixter, E. Watrall, R. Carty and D. Tarr, The Enslaved Ontology: Peoples of the historic slave trade, *Journal of Web Semantics* **63** (2020), 100567. doi:10.1016/j.websem.2020.100567.

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

[45] P. Hitzler and A. Krisnadhi, On the Roles of Logical Axiomatizations for Ontologies, in: *Ontology Engineering with Ontology Design Patterns – Foundations and Applications*, P. Hitzler, A. Gangemi, K. Janowicz, A. Krisnadhi and V. Presutti, eds, Studies on the Semantic Web, Vol. 25, IOS Press, 2016, pp. 73–80. doi:10.3233/978-1-61499-676-7-73.

[46] M.K. Sarker, D. Carral, A.A. Krisnadhi and P. Hitzler, Modeling OWL with Rules: The ROWL Protege Plugin, in: *Proceedings of the ISWC 2016 Posters & Demonstrations Track co-located with 15th International Semantic Web Conference (ISWC 2016), Kobe, Japan, October 19, 2016*, T. Kawamura and H. Paulheim, eds, CEUR Workshop Proceedings, Vol. 1690, CEUR-WS.org, 2016.

[47] M.K. Sarker, A. Krisnadhi, D. Carral and P. Hitzler, Rule-Based OWL Modeling with ROWLTab Protégé Plugin, in: *The Semantic Web – 14th International Conference, ESWC 2017, Portorož, Slovenia, May 28 – June 1, 2017, Proceedings, Part I*, E. Blomqvist, D. Maynard, A. Gangemi, R. Hoekstra, P. Hitzler and O. Hartig, eds, Lecture Notes in Computer Science, Vol. 10249, 2017, pp. 419–433. doi:10.1007/978-3-319-58068-5_26.

[48] C. Shimizu, A. Krisnadhi and P. Hitzler, On the Roles of Logical Axiomatizations for Ontologies, in: *Aplications and Practices in Ontology Design, Extration, and Reasoning*, G. Cota, M. Daquino and G.L. Pozzato, eds, Studies on the Semantic Web, IOS Press, to appear.

[49] C. Shimizu, P. Hitzler, Q. Hirt, A. Shiell, S. Gonzalez, C. Foley, D. Rehberger, E. Watrall, W. Hawthorne, D. Tarr, R. Carty and J. Mixter, The Enslaved Ontology 1.0: People of the Historic Slave Trade, Technical Report, Michigan State University, East Lansing, Michigan, 2019, available from https://daselab.cs.ksu.edu/projects/ontology-modeling-slave-trade.

[50] K. Hammar, Ontology design pattern property specialisation strategies, in: *International Conference on Knowledge Engineering and Knowledge Management*, Springer, 2014, pp. 165–180.

[51] P. Burnard, A method of analysing interview transcripts in qualitative research, *Nurse education today* **11**(6) (1991), 461–466.

[52] C.B. Seaman, Qualitative methods, in: *Guide to advanced empirical software engineering*, Springer, 2008, pp. 35–62.

[53] A. Dalal, C. Shimizu and P. Hitzler, Modular Ontology Modeling Meets Upper Ontologies: The Upper Ontology Alignment Tool, in: *Proceedings of the ISWC 2020 Posters & Demonstrations Track co-located with the 19th International Semantic Web Conference (ISWC 2020), Kobe, Japan, November 3, 2020*, 2020, In Press.

[54] S.J. Satpathy, Rules with Right hand Existential or Disjunction with ROWLTab, Master's thesis, Wright State University, 2019.

[55] C. Shimizu, L. McEwen and Q. Hirt, Towards a Pattern-Based Ontology for Chemical Laboratory Procedures, in: *Proceedings of the 9th Workshop on Ontology Design and Patterns (WOP 2018) co-located with 17th International Semantic Web Conference (ISWC 2018), Monterey, USA, October 9th, 2018*, M.G. Skjæveland, Y. Hu, K. Hammar, V. Svátek and A. Lawrynowicz, eds, CEUR Workshop Proceedings, Vol. 2195, CEUR-WS.org, 2018, pp. 40–51. http://ceur-ws.org/Vol-2195/research_paper_1.pdf.

[56] C. Shimizu, R. McGranaghan, A. Eberhart and A.C. Kellerman, Towards a Modular Ontology for Space Weather Research, *CoRR* **abs/2009.12285** (2020). https://arxiv.org/abs/2009.12285.

[57] A. Krisnadhi, Y. Hu, K. Janowicz, P. Hitzler, R.A. Arko, S. Carbotte, C. Chandler, M. Cheatham, D. Fils, T.W. Finin, P. Ji, M.B. Jones, N. Karima, K.A. Lehnert, A. Mickle, T.W. Narock, M. O'Brien, L. Raymond, A. Shepherd, M. Schildhauer and P. Wiebe, The GeoLink Modular Oceanography Ontology, in: *The Semantic Web - ISWC 2015 - 14th International Semantic Web Conference, Bethlehem, PA, USA, October 11-15, 2015, Proceedings, Part II*, M. Arenas, Ó. Corcho, E. Simperl, M. Strohmaier, M. d'Aquin, K. Srinivas, P.T. Groth, M. Dumontier, J. Heflin, K. Thirunarayan and S. Staab, eds, Lecture Notes in Computer Science, Vol. 9367, Springer, 2015, pp. 301–309. doi:10.1007/978-3-319-25010-6_19.

[58] A. Krisnadhi and P. Hitzler, Modeling With Ontology Design Patterns: Chess Games As a Worked Example, in: *Ontology Engineering with Ontology Design Patterns - Foundations and Applications*, P. Hitzler, A. Gangemi, K. Janowicz, A. Krisnadhi and V. Presutti, eds, Studies on the Semantic Web, Vol. 25, IOS Press, 2016, pp. 3–21. doi:10.3233/978-1-61499-676-7-3.

[59] L. Zhou, C. Shimizu, P. Hitzler, A.M. Sheill, S.G. Estrecha, C. Foley, D. Tarr and D. Rehberger, The Enslaved Dataset: A Real-world Complex Ontology Alignment Benchmark using Wikibase, in: *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*, M. d'Aquin, S. Dietze, C. Hauff, E. Curry and P. Cudré-Mauroux, eds, ACM, 2020, pp. 3197–3204. doi:10.1145/3340531.3412768.

[60] K. Hammar, E.O. Wallin, P. Karlberg and D. Hälleberg, The RealEstateCore Ontology, in: *The Semantic Web – ISWC 2019*, Lecture Notes in Computer Science, Vol. 11779, Springer, Cham, 2019, pp. 130–145. https://doi.org/10.1007/978-3-030-30796-7_9.

[61] K. Hammar, Ontology Design Principles for Model-Driven Applications – How Not to Shoot Yourself in the Foot with OWL, in: *Advances in Pattern-based Ontology Engineering (in print)*, E. Blomqvist, T. Hahmann, K. Hammar, P. Hitzler, R. Hoekstra, R. Mutharaju, M. Poveda, C. Shimizu, M. Skjaeveland, M. Solanki, V. Svátek and L. Zhouc, eds, Studies on the Semantic Web, IOS Press, 2021.

[62] T. Baker and E. Prud'hommeaux (eds), *Shape Expressions (ShEx) 2.1 Primer*, Final Community Group Report 09 October 2019, 2019, http://shex.io/shex-primer/index.html.

[63] H. Knublauch and D. Kontokostas (eds), *Shapes Constraint Language (SHACL)*, W3C Recommendation 20 July 2017, 2017, https://www.w3.org/TR/shacl/.