The Enslaved Ontology: Peoples of the Historic Slave Trade

Cogan Shimizu[a], Pascal Hitzler[a], Quinn Hirt[b], Dean Rehberger[c], Seila Gonzalez Estrecha[c], Catherine Foley[c], Alicia M. Sheill[c], Walter Hawthorne[c], Jeff Mixter[d], Ethan Watrall[c], Ryan Carty[c], Duncan Tarr[c]

[a]*Department of Computer Science, Kansas State University, USA*
[b]*Department of Computer Science and Engineering, Wright State University, USA*
[c]*Matrix, Michigan State University, USA*
[d]*OCLC Research, USA*

Abstract

We present the Enslaved Ontology (V1.0) which was developed for integrating data about the historic slave trade from diverse sources in a use case driven by historians. Ontology development followed modular ontology design principles as derived from ontology design pattern application best practices and the eXtreme Design Methodology. Ontology content focuses on data about historic persons and the event records from which this data can be taken. It also incorporates provenance modeling and some temporal and spatial aspects. The ontology is available as serialized in the Web Ontology Language OWL, and carries modularization annotations using the Ontology Pattern Language (OPLa). It is available under the Creative Commons CC BY 4.0 license.

Keywords: digital humanities, modular ontology, data integration, ontology design patterns, history of the slave trade

1. Introduction

The scourge of African enslavement was fundamental to the making of Europe, Africa, the Americas, and Middle East and parts of the Asian subcontinent. The enduring legacies of black bondage shape the moral questions of humanity in our times. We have seen in the past decade a growth in interest in the subject in film, on television, and in historical fiction. Historians have spilled much ink writing monographs aimed primarily at other scholars. At the same time, however, it is a worthy goal to expand the production of scholarly output and to bring what historians do to the general public. This aims to shed light on questions such as: How can we more effectively answer important moral questions? How can we make those questions part of a broader public discourse? What sources are available? How can we give broad access to them? And how in the decades to come will scholars answer questions about black bondage and its legacies when much valuable source material is deteriorating due to inattention, siloed scholarly activities, and underfunded archives?

During the past two decades there has been a significant shift in perceptions about what we can know about enslaved Africans, their descendants, and those who asserted ownership over them throughout the world. Those on the cutting edge of the digital humanities and social sciences have set about identifying, digitizing, analyzing, and making these resources available on innovative public history and cultural heritage websites. As a result, a growing number of collections of scanned original manuscript documents, digitized material culture, and databases, that organize and make sense of records of enslavement, are free and readily accessible for scholarly and public consumption.

Online databases about African slavery and the slave trade have a history that stretches back to the late 1990s and early 2000s, first provided on CD-ROMs, then on the World Wide Web as time progressed – and we provide key references in the related work section. Over time, a plethora of projects and teams generated a wide variety of databases with different foci and specializations. So, although this data is available through these data silos, this proliferation of different projects and databases presents scholars, students, and the interested public with a number of challenges:

- Most of these databases focus on the individuals of the slave trade, but data is often limited to the focus of the project. Further, the task of disambiguating (or merging) individuals across multiple datasets is nearly impossible given the current, siloed nature of all databases about slavery and the enslaved;
- There is no central, universally recognized clearinghouse for slave data. As such, it is difficult to find projects and databases;
- Individual projects and databases are isolated, preventing federated and cross project searching, browsing, and quan-

Email addresses: cogan.shimizu@coganshimizu.com (Cogan Shimizu), pascal@pascal-hitzler.de (Pascal Hitzler), hirtquinn@gmail.com (Quinn Hirt), rehberge@msu.edu (Dean Rehberger), seilage@msu.edu (Seila Gonzalez Estrecha), foleyc@msu.edu (Catherine Foley), asheill@msu.edu (Alicia M. Sheill), walterh@msu.edu (Walter Hawthorne), mixterj@oclc.org (Jeff Mixter), watrall@msu.edu (Ethan Watrall), cartyrya@msu.edu (Ryan Carty), tarrdunc@msu.edu (Duncan Tarr)

titative analysis;

- There are no best practices for digital data creation collectively agreed upon by the scholarly community;
- Important data is often lost or remain locked away in scholars' files, completely inaccessible to other scholars, students, descent communities, and the general public;
- Project participants rarely get scholarly credit for the work that goes into creating and releasing digital data;
- Humanists have little incentive to deposit datasets.

To address these challenges, the Enslaved project, funded by The Andrew W. Mellon Foundation and led by Michigan State University, is currently underway, and it is set to pioneer a new model for humanities scholarship. Enslaved brings together programmers, project managers, archivists, librarians, and historians in a collective endeavor and, over the years, with an expanding consortium of contributors. This collaborative approach, which is made possible by the World Wide Web, is set to challenge humanists to broaden their thinking about the production of knowledge; the sharing, as opposed to guarding, of research materials; and the benefits of collaboration. In sum, the model of Enslaved promises to disrupt conventions of humanities scholarship in much the way it would disrupt – for the better – historical perspectives on slavery and the individual lives of those enslaved.

The technical goal of Enslaved is thus to establish what we call the Enslaved Hub,[1] a website that provides one-stop querying and inspection capabilities for integrated historic data on the slave trade, originating from a diverse set of data sources and contributors, thereby allowing students, researchers and the general public to search over numerous databases to understand and reconstruct the lives of individuals who were part of the historical slave trade. To address the underlying data integration issues, Enslaved opted to follow the state of the art by establishing a knowledge graph, expressed in RDF, with an underlying schema in form of an OWL ontology. We call this the Enslaved Ontology; our modeling approach and its core concepts comprise the core of this paper.

The Enslaved Ontology expresses metadata record types and core fields that the Enslaved research team identified as frequently occuring in historic slave trade data projects. After months of evaluating datasets from the domain of slavery studies, Enslaved standardized metadata fields and developed controlled vocabularies for four types of records regularly found in these data collections: EVENT, PERSON, PLACE, and SOURCE. There are 43 Enslaved fields: 9 EVENT fields; 19 PERSON fields; 9 PLACE fields; and 6 SOURCE fields. Enslaved defined controlled vocabulary terms that standardize data in nine Enslaved fields. Documentation for Enslaved Metadata and Controlled Vocabularies is available at https://docs.enslaved.org/

[1] Eventually to be located at `http://enslaved.org/`

Like the metadata from which it originates, the Enslaved Ontology is the shared language that allows the Enslaved Hub to search over numerous disperate datasets to understand and reconstruct the lives of individuals who were part of the historical slave trade. Without a data model based on the records, fields, and terms included in the metadata and controlled vocabularies, it would be impossible for machine processing to make sense of all of the data available to the system.

The rest of this paper is structured as follows. In Section 2 we briefly describe our modeling approach. In Section 3 we provide a description of the ontology's key modules and their relationships. In Section 4 we briefly discuss the intended usage of the ontology. In Section 5 we discuss related work and in Section 6 we conclude.

The ontology is available at `https://docs.enslaved.org/`. Its detailed technical documentation is available from the same site and from [27].

2. Ontology Modeling Approach

We follow a modular ontology modeling approach based on ontology design patterns [3, 5, 6], as it will produce an ontology with desirable traits. Such a methodology is designed to ensure high quality and reuseability of the ontology, as well as cater to future expansions, both in terms of scope and in terms of granularity. These will allow the Enslaved Ontology to evolve as needs evolve and the number of collaborators increase. The modular ontology modeling approach and its rationale has been described in [16], and it is closely related to the eXtreme Design approach [2].

Following this methodology as laid out in [16] and further detailed in [9], we took the following subsequent steps.

Step 1: Define use case or scope of use cases.

The use case, including its anticipated future trajectory, was laid out in the Introduction, Section 1.

Step 2: Collect competency questions while looking at possible data sources and scoping the problem, i.e., decide on what should be modeled now, and what should be left for a possible later extension.

Competency questions were assembled from various sources. In January 2018, the Enslaved project team solicited search questions from the eight partner projects, seeking expert input from historians actively engaged in slavery studies and databasing. Over a four-week period, partners shared thoughts about potential audiences for the Enslaved project and how different audiences would have different search expectations and needs. Specific search suggestions focused on categories such as names, events, relationships, and place and time constraints. The Matrix team summarized this input.[2] Then the Enslaved

[2] The summary is available from `https://docs.enslaved.org/competencyquestions/v1/enslavedcompetencyquestions-v1.pdf`.

- Public 5: List the enslaved people in Reed County, NC, in the second half of the eighteenth century.
- Public 12: 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.
- K12 1: Who did Thomas Jefferson enslave at Monticello?
- K12 9: How many enslaved children lived in Boston when Phillis Wheatley lived there?
- Pro 4: What were the gender ratios of enslaved people identified as being of XXXX ethnicity?
- Pro 6: In what records does the enslaved person named XXXX appear? What were XXXX's professions? What places did he live? Who were his/her children and childrens children? Who did he marry?
- Pro 9: 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?
- Pro 20: What ever happened to Bernarda Angola, a Free African who ran away from her mistress Maria dos Santos Pereira, in June 1845?

Figure 1: Sample competency questions.

team invited four project partners to draft competency questions, queries about the data but formulated in natural language. These scholars created 17 questions from a Public user perspective, 14 questions from students and teachers in primary through secondary school (K12), and 24 questions from a professional scholarly user.[3] Example competency questions are listed in Figure 1.

The collected competency questions span the scope of an expanded Enslaved Hub, which goes significantly beyond the first stage of modeling and demonstration which we – and the ontology in the current version – needed to focus on. The historians in the team thus designed a suitable scope for a first stage of the effort, based on the competency questions and the availability of data sources. From the plethora of potential data sources, some of which are presented in the related work section, eight were selected, namely African Origins,[4] Voyages: The Trans-Atlantic Slave Trade Database,[5] Slave Societies Digital Archive,[6] Dictionary of Caribbean and Afro-Latin American Biography, Dictionary of African Biography and African American National Biography,[7] Freedom Narratives,[8] Legacies of British Slave-ownership,[9] The Liberated Africans Project[10] and Slave Biographies.[11] These sources were selected as they provided a range of different kinds of data and seemed representative as a starting point.

Enslaved used the eight original datasets as the foundation for the Ontology. Many of the underlying data points are the same across the data collections, for example Names of individuals, Gender, Race/Color, and freedom status (enslaved, owner, freed person). However, historians tend to be idiosyncratic in their data collection practices. Some copy verbatim from sources while others may have complex coding practices. Field names and values can vary in spelling, language and use antiquated and contemporary terms. Categorization also tends to be non-standardized. Therefore there are some challenges mapping disparate datasets to the data model. The original datasets were selected for the project because they included a wide range of challenges that needed to be addressed. For example, a dataset focused on biographies combines agent and event data in the single row while another dataset captured all agent data in one table and connected it to event data in another table using an Enslaved property. Despite these divergences, Enslaved has found that robust metadata documentation for legacy datasets and Enslaved has allowed coherent mapping the data model. To date Enslaved has mapped all eight original datasets to the Enslaved Ontology. From the outset, Enslaved intended to include only a subset of the data points in these datasets, those that commonly recur across historical slave trade data collections.

Step 3: Identify key notions from the data and the use case and identify which pattern should be used for each. Construct a set of modules from these.

The list of key notions was quickly finalized during the first in-person modeling meeting in summer 2018, where about a dozen researchers, including historians, data experts, and ontology engineers, met to draft the modules and the overall ontology. As an ontology for historic data, time, place, and provenance play a necessary role. The content focus of the ontology is on persons and key biographical or person data, more precisely name, age, sex, occupation, status (e.g. enslaved or freed), race, ethnolinguistic and/or geographical origin, participation in events, and relationships to other persons or organizations such as family relations or ownership relations. Events play a particular role in the data, as relevant historic records usually originate from specific events such as estate inventories. As further modules it seemed necessary to have a generic way of providing descriptions and external references, as well as a way to refer to specific research projects contributing data.

As for identifying ontology design patterns as a basis for corresponding modules, some were obvious choices, such as using the core of PROV-O [21] for provenance; using agent instead of person to include organizations or groups when needed; an agent roles pattern [15] for participations in events and for inter-agent relationships. Some person data, such as sex, occupation, enslaved/freed status, race, and origin seemed to be best captured by using controlled vocabularies, as were age categories, though numeric age also seemed desirable. It also seemed opportune to defer a complex modeling of names from different ethnic and linguistic origins and instead to go for a simple name

[3] The full list of competency questions is available from https://docs.enslaved.org/competencyquestions/v2/enslavedcompetencyquestions-v2.pdf.
[4] http://www.african-origins.org/
[5] http://www.slavevoyages.org/
[6] http://www.vanderbilt.edu/esss/
[7] https://hutchinscenter.fas.harvard.edu/AANB
[8] http://freedomnarratives.org/
[9] http://www.ucl.ac.uk/lbs/
[10] http://liberatedafricans.org
[11] http://slavebiographies.org/

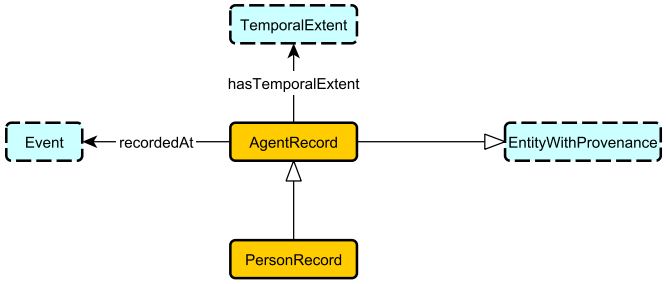

Figure 2: Schema Diagram for the AgentRecord module. Orange boxes indicate classes. Dashed blue boxes indicate classes which are part of another module. White-headed arrows indicate subclass relationships.

stub [18]. The usage of time was also rather restricted, so for the first stage a decision was made to develop a relatively simple placeholder module which could be refined later. A comprehensive treatment of historic places was clearly out of scope for the project, and since others pursue historic gazetteers in a principled way[12] we took a limited approach compatible with these efforts.

Since the focus is on historic data, and on historic persons, as reported by various (and possibly conflicting) sources, it was also obvious that the modeling, and the resulting knowledge graph, would contain possibly conflicting data as reported, rather than anything resembling a base truth. This is of course obvious from a historian's perspective, but is sometimes less prominently modeled in current approaches to knowledge graph schema design. In the Enslaved Ontology, we make this explicit by primarily speaking about the *records* of agents (so-called AgentRecords) as they pertain to different Agents.

Step 4: Put the modules together and add axioms which involve several modules.

We defer this discussion to the description of the modules and the ontology, in Section 3.

Step 5: Create OWL files.

OWL files were created using Protégé [23], and the OPLa Annotator, presented in [26], for adding additional metadata using the Ontology Design Pattern Representation Language (OPLa) [7].

3. Description of the Enslaved Ontology

As discussed previously, the focus of the Enslaved Ontology is on the records of historic agents. In this paper, we do not intend to replicate the comprehensive documentation, which can be found at [27]. Thus, we provide a broad overview, discuss key modeling choices, and give a few more detailed examples.

The key notion of our model is that of an agent record. Figure 2 shows the corresponding schema diagram. Note the use of temporal and provenance information, which are described in separate modules, and of the Event module: In this historic context, agent records were usually recorded at some historic events. We will discuss provenance in more detail below.

The schema diagram, of course, is just a simplified visualization of the module. In terms of formal model, it consists of the OWL axioms listed in Figure 3.[13] Here and elsewhere in the ontology, the primary purpose of the formal axiomatization is to disambiguate the model, i.e., we were striving for as complete an axiomatization as possible, while avoiding ontological overcommitments. Each axiom was discussed in detail between the ontology engineers and the historians on the team. The axiomatization is expresssed using the OWL 2 DL profile. Note that while it is not currently our primary goal to do formal reasoning over the ontology [8], we do not want to rule out such goals in the future (e.g. the use of reasoning for consistency checking). Furthermore, in order to fully encode the knowledge as determined to be important by stakeholders and domain experts, we make use of a number of features only expressible in OWL 2 DL, e.g. right-hand disjunctions appearing in Axiom 7 in the formalization of AgentRecord or Axiom 2 in the formalization of ExternalReference. These axioms, and others, may be found in more detail in the full documentation [27].

$$\text{AgentRecord} \sqsubseteq \text{EntityWithProvenance} \quad (1)$$

$$\text{PersonRecord} \sqsubseteq \text{AgentRecord} \quad (2)$$

$$\text{AgentRecord} \sqsubseteq =1\text{hasAgentRecord}^{-}.\text{Agent} \quad (3)$$

$$\text{PersonRecord} \sqsubseteq =1\text{hasPersonRecord}^{-}.\text{Person} \quad (4)$$

$$\text{AgentRecord} \sqsubseteq =1\text{hasTemporalExtent.TemporalExtent} \quad (5)$$

$$\text{AgentRecord} \sqsubseteq \leq 1\text{recordedAt.Event} \quad (6)$$

$$\text{AgentRecord} \sqsubseteq \leq 1\text{isDirectlyBasedOn.EntityWithProvenance} \quad (7)$$

$$\text{hasPersonRecord} \sqsubseteq \text{hasAgentRecord} \quad (8)$$

$$\text{Person} \sqsubseteq \geq 0\text{hasPersonRecord.PersonRecord} \quad (9)$$

Figure 3: OWL axioms for the AgentRecord module.

Systematically speaking, most of the axiomatization follows the template laid out for the OWLAx Protégé plug-in [25, 9], i.e., axioms were selected from those available in OWLAx, which – consistent with our previous modeling experiences – suffices for most modeling requirements. Indeed, all of the axioms in Figure 3, except for the last two, follow the OWLAx template. (1) and (2) are subclass relationships, (3) and (4) are combined existentials and inverse functionalities, (5) is a combined existential and functionality, (6) and (7) are functionalities, (8) is a subproperty relationship. (9) is an axiom which is actually a tautology, but is included for the benefit of humans trying to understand the ontology: it indicates that a person may have person records.

[12] `http://whgazetteer.org/`

[13] A primer on description logic and the notation can be found in [1].

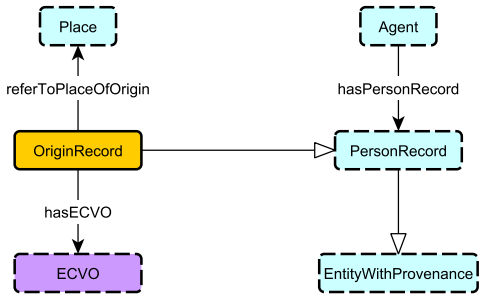

Figure 4: Schema Diagram for the OriginRecord module, color and shape usage is the same as in the previous diagram. Purple dashed boxes indicate a controlled vocabulary, i.e., all URIs of type ECVO (Ethnolinguistic Controlled Vocabulary of Origin) should be considered part of a controlled vocabulary.

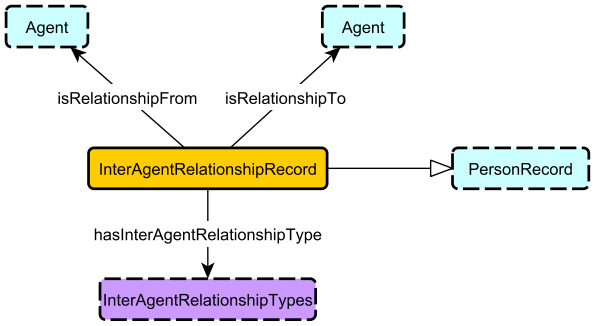

Figure 5: Schema Diagram for the InterAgentRelationshipRecord module, color and shape usage is the same as in the previous diagrams.

Several agent and person record modules are part of the Enslaved ontology; they are used for place of origin and ethnolinguistic origin of a person, for race, age, sex, occupation, name, and freedom status. They are also used for recording relationships between agents, and participations in events. We will now discuss a few of them.

The OriginRecord module is rather similar in structure to most of the other AgentRecord submodules, and so we use it to serve as a typical example. Its schema diagram can be found in Figure 4. We see that OriginRecord is a subclass of PersonRecord. A record of origin can be recorded in three ways, either by indicating a place (refering to the Place module), or by reference to a controlled vocabulary, the ECVO, which is being designed by Enslaved historians, or finally by referring to one ECVO and a place of origin. The OWL axiomatization which constitutes the module disambiguates the usage of this module: It prescribes that each OriginRecord can have at most one ECVO, and that each OriginRecord refers to at least one Place and/or at least one ECVO. Note that several places are allowed (as two places together again constitute a place, conceptually this may be considered the *union* of those places), and that there is no required relationship between place and ECVO, in case both are listed. The reasons for the latter are that, on the one hand, such relationships are rather controversial among historians, and that, on the other hand, such spatial information could be made part of the controlled vocabulary if desired. We anticipate that a refinement of this module (and of most others) may have to be done as usage of the ontology expands.

Records for inter-agent relationships, such as family or ownership relations, follow a standard relationship reification (or n-ary relation) pattern [11], with the addition of a controlled vocabulary for relationship types. The corresponding schema diagram can be found in Figure 5. The axiomatization, which we do not replicate here, follows the OWLAx templates. Note, though, that due to the reification it is not possible, in OWL DL, to specify that no agent can be in an InterAgentRelationship with itself, as non-simple OWL DL properties must not be irreflexive [22, 10].

Let us briefly pause to discuss the use of controlled vocabularies in our context. They are used in the Enslaved Ontology whenever a crisp categorization made sense from the domain experts' point of view, e.g., to enable efficient end-user querying of the data. The competency questions showed that queries for specific family relations or ethnolinguistic origins should occur frequently. In some cases, such as that for inter agent relationships, the controlled vocabulary entities act as types (in this case, for individuals belonging to InterAgentRelationshipRecord), and it could be argued that they should be modeled as subclasses of the InterAgentRelationshipRecord class, rather than as individuals. There are several reasons why we did not do that. (1) The controlled vocabulary is likely to change more rapidly than the rest of the ontology, and adding (or even removing) individuals appears to be less invasive rearding the ontology. That is, adding an individual of a class adds relatively little additional complexity as opposed to adding a new class, which could necessitate additional axiomatization. Indeed, we can now consider the controlled vocabularies to be seperate entities, with separate versioning, which can be updated without releasing a new version of the ontology. (2) In some cases, such as for the OriginRecord, it seems conceptually questionable to use the controlled vocabulary for classes. Thus, with all controlled vocabularies being individuals, the ontology has more coherence. (3) If needed, e.g., to establish formal relationships, expressed by OWL axioms, between controlled vocabulary items, then it is possible, in OWL DL, to map between these individual and corresponding class identifiers – this is a form of typecasting, laid out in detail in [20]. We acknowledge, though, that it could have been done differently; since typecasting is possible, though, the choice does not seem to make a huge difference.

Records for event participation, such as a baptism or a slave rebellion, follow a standard agent role pattern [15], which is a reification pattern. We again make use of controlled vocabularies for role types. The corresponding schema diagram can be found in Figure 6. Axiomatization is relatively straightforward following the OWLAx templates.

Event is a key module in the Enslaved ontology, as the creation of historical records usually happens at certain events, such as baptisms. While there are existing ontologies to capture event, such as the Simple Event Model [31], they did not fit our pur-

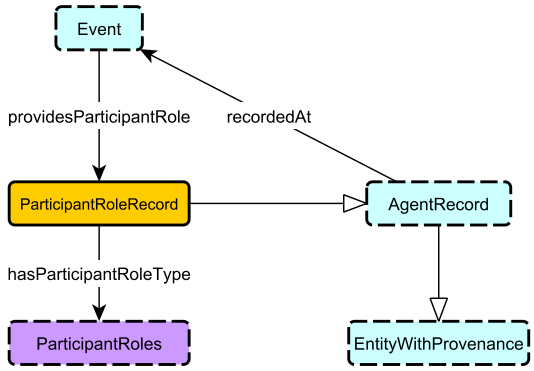

Figure 6: Schema Diagram for the ParticipantRoleRecord module, color and shape usage is the same as in the previous diagrams.

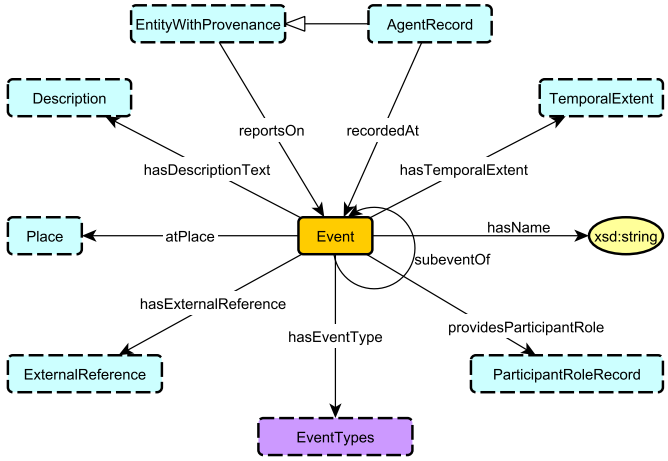

Figure 7: Schema Diagram for the Event module, color and shape usage is the same as in the previous diagrams. Yellow oval nodes indicate data types.

poses, mostly because they were much more detailed than required at this time. We thus derived our event module from a very simple event pattern [17] with appropriate modifications. E.g., at this stage there was no need for full-fledged spatiotemporal extents, and we added a controlled vocabulary to record event types, always keeping in mind that the module may have to be replaced by a module with finer granularity in the future. A schema diagram can be found in Figure 7.

For the axiomatization, which mainly follows the OWLAx templates, we required to express that, if an agent record is recorded at an event, then that agent record also reports on this event. This can be expressed as a (first order predicate logic) rule as

$$\text{AgentRecord}(x) \wedge \text{recordedAt}(x, y) \wedge \text{Event}(y) \rightarrow \text{reportsOn}(x,y).$$

Using a technique known as *rolification* [19, 24], this rule can

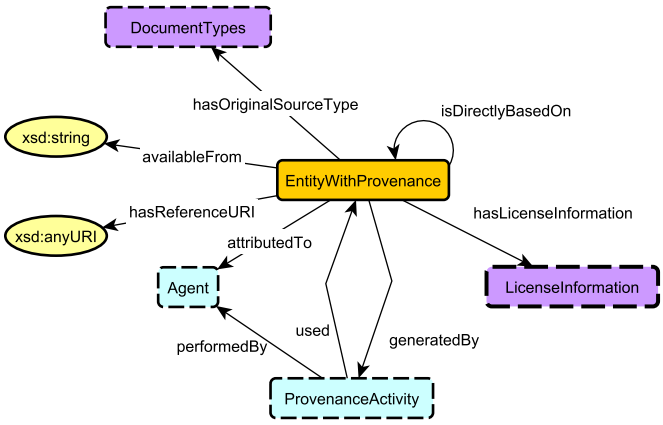

Figure 8: Schema Diagram for the Provenance module, color and shape usage is the same as in the previous diagrams.

be converted to OWL DL as the following three axioms.

$$\text{AgentRecord} \sqsubseteq \exists \text{agentRecordSelfProperty.Self}$$
$$\text{Event} \sqsubseteq \exists \text{eventSelfProperty.Self}$$
$$\text{agentRecordSelfProperty} \circ \text{recordedAt} \circ \text{eventSelfProperty} \sqsubseteq \text{reportsOn}$$

For provenance information, we borrowed the core pattern from PROV-O [21, 28] as a template, and added controlled vocabularies for license information and document types of the original source, which is important for Enslaved use case scenarios. We also added the option to record from where an entity is available (e.g., which database) together with a reference URI pointing to the exact database entry. This module's schema diagram can be found in Figure 8. The axiomatization follows mainly OWLAx templates. We record, in addition, that the original source type of an entity with provenance is the same as that of the entity with provenance it is directly based on, and vice-versa, which can be expressed as property chain axioms, as follows.

$$\text{isDirectlyBasedOn} \circ \text{hasOriginalSourceType} \sqsubseteq \text{hasOriginalSourceType}$$
$$\text{isDirectlyBasedOn}^{-} \circ \text{hasOriginalSourceType} \sqsubseteq \text{hasOriginalSourceType}$$

The ontology has additional modules, in particular preliminary ones for place and time, which we do not discuss in more detail. However, we provide an overview schema diagram for the whole ontology in Figure 9. Further details can of course be found in the referenced technical report.

Modular ontologies are one way of addressing the maintenance, interoperability, extensibility, and reusability of ontologies. To do so, each module is annotated using the Ontology Design Pattern Representation Language (OPLa) [7] using the OPLa Annotator plugin [26] for Protégé [23]. These annotations allow us to fully describe the structure of a modular ontology

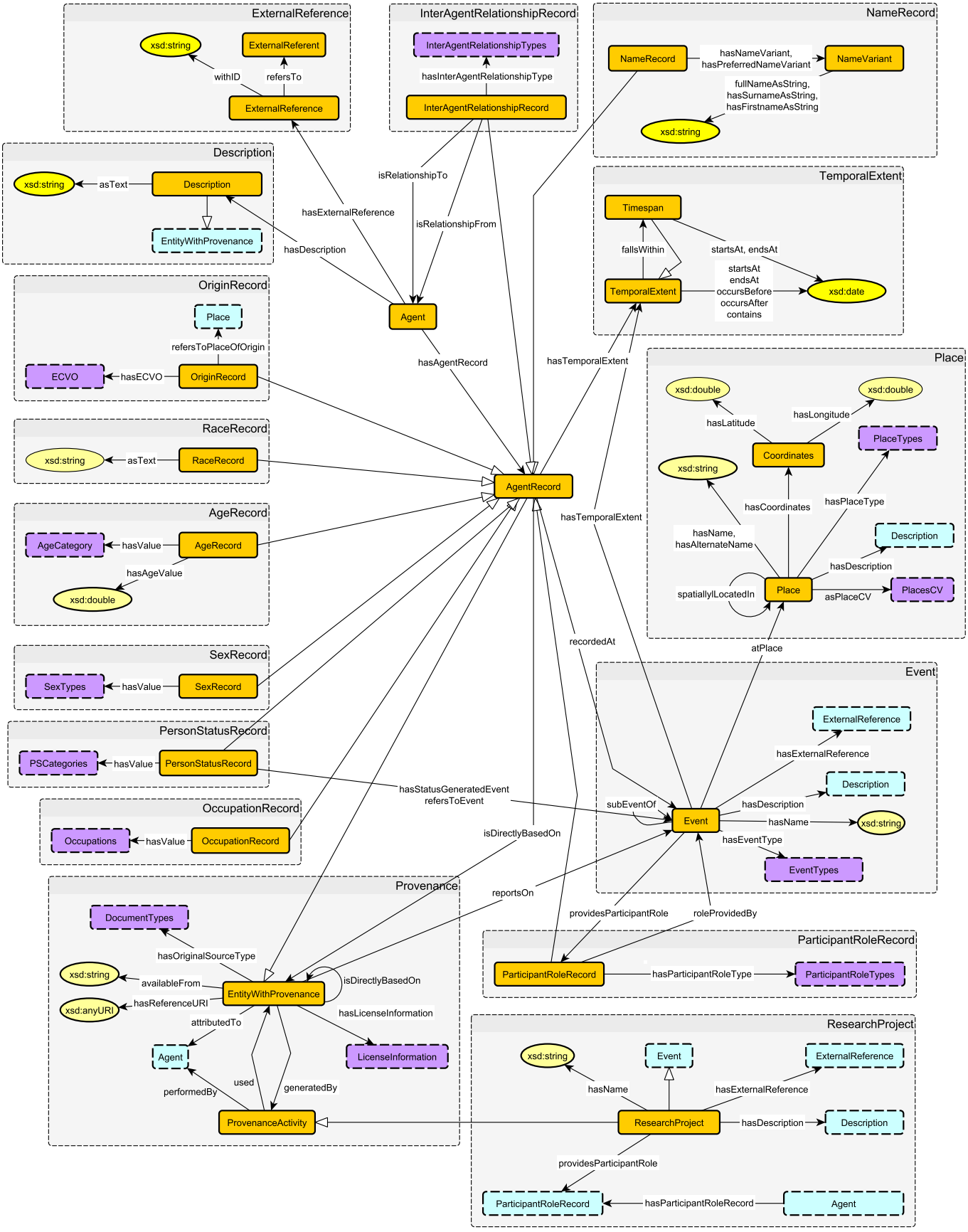

Figure 9: Overview schema Diagram for the Enslaved ontology, color and shape usage is the same as in the previous diagrams. The gray frames indicate (some of the) modules.

in a machine-readable way. Briefly, OPLa shows how the modules that comprise a modular ontology are interrelated and from which patterns those modules were derived. This, in turn, makes it easier to replace or modify modules as the ontology evolves.

The Enslaved Ontology is available from the Enslaved Hub[14] and is available under the Creative Commons Attribution License CC BY 4.0,[15] which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

4. Intended Usage

The Enslaved project aims to connect the different communities pertaining to the historic slave trade. The nature of this endeavor calls for a bootstrapping approach, where a demonstrator hub would first be limited in scope, diversity, and number of datasets integrated, however with the expectation that it would grow significantly once under way, as additional collaborators and contributors start sharing their data for the integration. The Enslaved Ontology, due to its modular nature, has the capacity to expand and evolve alongside the Enslaved Hub.

Additionally, we wish to enable community driven population of the knowledge graph. Currently, the Enslaved Ontology serves as the underlying schema for the Enslaved knowledge graph. The axiomatizations provided in earlier sections and in [27] are a way for disambiguating the model—at this time, there is no intent to do formal, deductive reasoning over the ontology. Instead, the ontology is meant for human consumption, and for informing the knowledge graph structure.

The Enslaved Ontology is also used to inform the population interface; it has been completely mapped into a Wikibase installation, the same technology that which underlies Wikidata [32]. This enables compatibility with the greater data and software ecosystem, for example the Wikimedia Foundation's[16] Wikipedia,[17] while maintaining a different path for the curation of the data. Blazegraph[18] is used as the underlying triplestore. Classes, properties, and individuals (instances) may be queried through a SPARQL query endpoint.

5. Related Work

5.1. Databases About the Slave Trade

Online databases about African slavery and the slave trade have a history that stretches back to the late 1990s and early 2000s. Then three projects were trailblazing. First, in 1993 Edward L. Ayres and William Thomas launched Valley of the Shadows at the Virginia Center for Digital History.[19] The site is a digital archive of primary sources centered on the lives of people, both black and white, in Virginia and Pennsylvania during the era of the American Civil War. It allows readers to take multiple pathways through documentation from the period and, in so doing, to build a variety of narratives that tell history in a way that a book cannot. Second, in March 2000, Gwendolyn Midlo Hall published a CD-ROM with the Louisiana State University Press. The CD had information about over 100,000 slaves who labored in colonial Louisiana. In 2001, Hall launched much of the same information on a website sponsored by University of North Carolina and I-Biblio, Afro-Louisiana History and Genealogy.[20] An update and revision of the dataset would later be incorporated into Slave Biographies.[21] Third, in 1999 David Eltis, Stephen D. Brehrendt, David Richardson, and Herbert S. Klein published The Trans-Atlantic Slave Trade: A Database on CD-ROM with Cambridge University Press. The CD had information about 27,233 Atlantic slave ship voyages from 1595 to 1866. In 2006, the team launched Voyages: The Trans-Atlantic Slave Trade Database.[22] The Voyages site has grown, currently holding information about almost 36,000 slave voyages. For both Afro-Louisiana and Voyages, the advantages of the World Wide Web over a CD-ROM were obvious: audiences grew, information was available at no cost to anyone with Internet access, and data could be updated.

Other projects followed and were housed on the Internet. Slave Biographies: The Atlantic Database Network, which is home to updated data from Hall's I-Biblio site and to other datasets, is a digital project hosted by Michigan State University and Matrix: The Center for Digital Humanities and Social Sciences. It provides open-source demographic information, including names, ethnicities, skills, occupations, and illnesses, about individual slaves in colonial Louisiana and Maranhão, Brazil. In addition, there is Cornell University's Freedom on the Move,[23] Emory University and Harvard University's W.E.B. Du Bois Institute's African Origins.[24] and Vanderbilt University's Slave Societies Digital Archive.[25] Still other projects are supported at universities and heritage institutions both within and outside the United States. Among them are the Liberated Africans Project,[26] Digital Archaeological Archive of Comparative Slavery,[27] Legacies of British Slave-ownership,[28] Marronnage in Saint-Domingue, Haiti,[29] Baptismal Records Database for Slave Societies,[30] and Studies in the History of the African Diaspora.[31]

[14] https://docs.enslaved.org/
[15] http://creativecommons.org/licenses/by/4.0/
[16] https://en.wikipedia.org/wiki/Wikimedia_Foundation
[17] https://en.wikipedia.org/wiki/Wikipedia:About
[18] https://blazegraph.com/
[19] http://valley.lib.virginia.edu/
[20] http://www.ibiblio.org/laslave/
[21] http://www.slavebiographies.org/
[22] http://www.slavevoyages.org/
[23] http://freedomonthemove.org/
[24] http://www.african-origins.org/
[25] http://www.vanderbilt.edu/esss/
[26] http://www.liberatedafricans.org/
[27] http://www.daacs.org/
[28] http://www.ucl.ac.uk/lbs/
[29] http://www.marronnage.info/en/index.html
[30] http://bardss.matrix.msu.edu/
[31] http://tubman.info.yorku.ca/publications/shadd/

5.2. Related Digital Humanities Ontologies

To the best of our knowledge, there is no other ontology that aims to model the lives, times, and movements of peoples in the historic slave trade. However, there is an active community of historians and digital humanitarians utilizing Semantic Web technologies. We discuss a few such efforts in the digital humanities to provide context for our work. These chosen works are intended to be representative of a fairly extensive body of literature. We selected them based on relevance, shared vision, and recency, as well as attempting to choose representative types of topics.

dataLegend [12] is a platform for linking historical data, first published in 2015. Essentially, it allows for a researcher to upload their data, link it to other's data, and view the results of their mapping. This project is similar in vision to the Enslaved Hub, but with broader scope. In particular, the Enslaved project focuses on those people involved in the historic slave trade, where as dataLegend allows for the mapping of arbitrary historical data. While this is not undesirable, it is outside the scope of our work to be similarly comprehensive. However, dataLegend does remain a source of inspiration as Enslaved grows.

Europeana[32] is a massive in scope, digital archive with over 58 million metadata records of various media, first launched in 2008. Europeana acts as a discovery portal; it redirects users to different institutions' datasets. This work is similar in vision to the Enslaved Hub, but has a much more relaxed semantics for interrelating datasets, using Schema.org annotations.

Semantic National Biography of Finland[33] [14, 13] (BiographySampo) is a collection of short biographies of historic peoples that have been extracted from large biographical collections published in 2017. BiographySampo uses a combination of different vocabularies and models to represent historic peoples and the events and roles that connected them. This has similar goals as the Enslaved Ontology (e.g. the tracking of historical peoples); we see multiple points for alignment and inspiration as the ontology evolves.

Historical Ecology and Recipes from Newspapers [29, 30] are two projects, first published in 2018, that attempt to extract ecological and recipe information from historical newspapers. This approach is somehow similar to ours, in the sense that it is about extracting records of items of interest from historical documents and tracking how they change over time.

Additionally, we identify two upper ontologies that have been used in the digital humanities.

The **Simple Event Model** [34] (SEM) is an upper ontology for describing events, in particular those relating to history, cultural heritage, geography and multimedia. It captures some spatial and temporal data and different ways of describing how people, places, and objects interact. The SEM is a rather finely granular, complex, and robust model, but exceeds our current modeling needs. As the Enslaved Ontology is a modular ontology, it would be possible to use SEM instead of our Event module. In the future, as the Enslaved Ontology evolves, we may consider replacing the current Event module with an adaptation of the SEM.

The **International Committee for Documentation's Conceptual Reference Model** (CIDOC-CRM) is an upper ontology for "describing the implicit and explicit concepts and relationships used in cultural heritage documentation."[35] In the same manner that SEM provides a more heavyweight and more finely granular model, so does CIDOC-CRM. As such, we developed the Enslaved Ontology independently, but due to its modular nature, as the ontology evolves and needs change, an adaptation of the relevant portions of CIDOC-CRM will be considered.

5.3. Considerations on External Ontologies and Modules

During the development of the Enslaved Ontology, we also considered a number of external ontologies for use. We discuss a representative selection of them here. Furthermore, we may include alignments from our developed modules to these external ontologies in subsequent versions of the Enslaved Ontology.

OWL-Time is an ontology of temporal concepts encoded in OWL 2 DL [4]. The Time ontology is as robust as it is complex. Our modeling needs were met by simply using some basic temporal relationships and the XML schema datatypes, so we felt it unnecessary to import OWL-Time.

Geonames Ontology[36] is a geographical database containing the names and locations of over ten million places. However, it is not sufficient for our purposes as we are dealing, primarily, with historical places.

6. Conclusions

In this paper, we have presented the Enslaved Ontology (V1.0) which has been developed for use in integrating a wide variety of heterogeneous data sources found in the historian research communities. The ontology was developed in a modular fashion, thus facilitating future maintenance and extensibility—critical and beneficial design aspects, as the expected growth of the Enslaved Hub must be matched by the Enslaved Ontology. The ontology incorporates modules for capturing the spatial, temporal, and provenance aspects of historic events and agents (i.e., organizations or persons). The Enslaved Ontology is serialized in OWL and is equipped with annotations describing the modular structure, further improving its reusability in the future. It may be found online at `https://docs.enslaved.org/`.

[32] `https://www.europeana.eu/portal/en`
[33] `https://seco.cs.aalto.fi/projects/biografiasampo/en/`
[34] `https://semanticweb.cs.vu.nl/2009/11/sem/`
[35] `http://www.cidoc-crm.org/`
[36] `https://www.geonames.org/`

Acknowledgement. This work was supported by The Andrew W. Mellon Foundation through the Enslaved project. Cogan Shimizu acknowledges partial support from the Dayton Area Graduate Studies Institute.

References

[1] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003.

[2] E. Blomqvist, K. Hammar, and V. Presutti. Engineering Ontologies with Patterns – The eXtreme Design Methodology. In P. Hitzler, A. Gangemi, K. Janowicz, A. Krisnadhi, and V. Presutti, editors, *Ontology Engineering with Ontology Design Patterns – Foundations and Applications*, volume 25 of *Studies on the Semantic Web*, pages 23–50. IOS Press, 2016.

[3] E. Blomqvist and K. Sandkuhl. Patterns in ontology engineering: Classification of ontology patterns. In C. Chen, J. Filipe, I. Seruca, and J. Cordeiro, editors, *ICEIS 2005, Proceedings of the Seventh International Conference on Enterprise Information Systems, Miami, USA, May 25-28, 2005*, pages 413–416, 2005.

[4] S. Cox and C. Little, editors. *OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax (Second Edition)*. W3C Recommendation, 11 December 2012. Available at http://www.w3.org/TR/owl2-syntax/.

[5] A. Gangemi. Ontology design patterns for semantic web content. In Y. Gil, E. Motta, V. R. Benjamins, and M. A. Musen, editors, *The Semantic Web – ISWC 2005, 4th International Semantic Web Conference, ISWC 2005, Galway, Ireland, November 6-10, 2005, Proceedings*, volume 3729 of *Lecture Notes in Computer Science*, pages 262–276. Springer, 2005.

[6] P. Hitzler, A. Gangemi, K. Janowicz, A. Krisnadhi, and V. Presutti, editors. *Ontology Engineering with Ontology Design Patterns – Foundations and Applications*, volume 25 of *Studies on the Semantic Web*. IOS Press, 2016.

[7] P. Hitzler, A. Gangemi, K. Janowicz, A. A. Krisnadhi, and V. Presutti. Towards a simple but useful ontology design pattern representation language. In E. Blomqvist, Ó. Corcho, M. Horridge, D. Carral, and R. Hoekstra, editors, *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.*, volume 2043 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2017.

[8] P. Hitzler and A. Krisnadhi. On the roles of logical axiomatizations for ontologies. In P. Hitzler, A. Gangemi, K. Janowicz, A. Krisnadhi, and V. Presutti, editors, *Ontology Engineering with Ontology Design Patterns – Foundations and Applications*, volume 25 of *Studies on the Semantic Web*, pages 73–80. IOS Press, 2016.

[9] P. Hitzler and A. Krisnadhi. A tutorial on modular ontology modeling with ontology design patterns: The Cooking Recipes Ontology. Technical report, Data Semantics (DaSe) Lab, Wright State University, Dayton, OH, USA, 2018. Available from `http://pascal-hitzler.de/topics/publications.html`.

[10] P. Hitzler, M. Krötzsch, and S. Rudolph. *Foundations of Semantic Web Technologies*. Chapman & Hall/CRC, 2010.

[11] R. Hoekstra. *Ontology Representation – Design Patterns and Ontologies that Make Sense*, volume 197 of *Frontiers in Artificial Intelligence and Applications*. IOS Press, 2009.

[12] R. Hoekstra, A. Meroño-Peñuela, A. Rijpma, R. Zijdeman, A. Ashkpour, K. Dentler, I. Zandhuis, and L. Rietveld. The dataLegend ecosystem for historical statistics. *J. Web Semant.*, 50:49–61, 2018.

[13] E. Hyvönen, P. Leskinen, M. Tamper, H. Rantala, E. Ikkala, J. Tuominen, and K. Keravuori. BiographySampo—publishing and enriching biographies on the semantic web for digital humanities research. *The Semantic Web: ESWC*, 2019. To appear.

[14] E. Hyvönen, P. Leskinen, M. Tamper, J. Tuominen, and K. Keravuori. Semantic National Biography of Finland. In E. Mäkelä, M. Tolonen, and J. Tuominen, editors, *Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference, DHN 2018, Helsinki, Finland, March 7-9, 2018.*, volume 2084 of *CEUR Workshop Proceedings*, pages 372–385. CEUR-WS.org, 2018.

[15] A. Krisnadhi. The role patterns. In P. Hitzler, A. Gangemi, K. Janowicz, A. Krisnadhi, and V. Presutti, editors, *Ontology Engineering with Ontology Design Patterns – Foundations and Applications*, volume 25 of *Studies on the Semantic Web*, pages 313–319. IOS Press, 2016.

[16] A. Krisnadhi and P. Hitzler. Modeling with ontology design patterns: Chess games as a worked example. In P. Hitzler, A. Gangemi, K. Janowicz, A. Krisnadhi, and V. Presutti, editors, *Ontology Engineering with Ontology Design Patterns – Foundations and Applications*, volume 25 of *Studies on the Semantic Web*, pages 3–21. IOS Press, 2016.

[17] A. Krisnadhi and P. Hitzler. A core pattern for events. In K. Hammar, P. Hitlzer, A. Lawrynowicz, A. Krisnadhi, A. Nuzzolese, and M. Solanki, editors, *Advances in Ontology Design and Patterns*, volume 32 of *Studies on the Semantic Web*, pages 29–38. IOS Press / AKA Verlag, 2017.

[18] A. Krisnadhi and P. Hitzler. The stub metapattern. In K. Hammar, P. Hitlzer, A. Lawrynowicz, A. Krisnadhi, A. Nuzzolese, and M. Solanki, editors, *Advances in Ontology Design and Patterns*, volume 32 of *Studies on the Semantic Web*, pages 39–64. IOS Press / AKA Verlag, 2017.

[19] A. Krisnadhi, F. Maier, and P. Hitzler. OWL and Rules. In A. Polleres, C. d'Amato, M. Arenas, S. Handschuh, P. Kroner, S. Ossowski, and P. F. Patel-Schneider, editors, *Reasoning Web. Semantic Technologies for the Web of Data – 7th International Summer School 2011, Galway, Ireland, August 23-27, 2011, Tutorial Lectures*, volume 6848 of *Lecture Notes in Computer Science*, pages 382–415. Springer, Heidelberg, 2011.

[20] A. A. Krisnadhi, P. Hitzler, and K. Janowicz. On the capabilities and limitations of OWL regarding typecasting and ontology design pattern views. In V. A. M. Tamma, M. Dragoni, R. S. Gonçalves, and A. Lawrynowicz, editors, *Ontology Engineering – 12th International Experiences and Directions Workshop on OWL, OWLED 2015, co-located with ISWC 2015, Bethlehem, PA, USA, October 9-10, 2015, Revised Selected Papers*, volume 9557 of *Lecture Notes in Computer Science*, pages 105–116. Springer, 2015.

[21] T. Lebo, S. Sahoo, and D. McGuinness, editors. *PROV-O: The PROV Ontology*. W3C Recommendation, 30 April 2013. Available at http://www.w3.org/TR/prov-o/.

[22] B. Motik, P. Patel-Schneider, and B. Parsia, editors. *Time Ontology in OWL*. W3C Recommendation, 19 October 2017. Available at https://www.w3.org/TR/owl-time/.

[23] M. A. Musen. The Protégé project: a look back and a look forward. *AI Matters*, 1(4):4–12, 2015.

[24] M. K. Sarker, A. Krisnadhi, D. Carral, and P. Hitzler. Rule-based OWL modeling with ROWLTab Protégé plugin. In E. Blomqvist, D. Maynard, A. Gangemi, R. Hoekstra, P. Hitzler, and O. Hartig, editors, *The Semantic Web – 14th International Conference, ESWC 2017, Portorož, Slovenia, May 28 - June 1, 2017, Proceedings, Part I*, volume 10249 of *Lecture Notes in Computer Science*, pages 419–433, 2017.

[25] M. K. Sarker, A. A. Krisnadhi, and P. Hitzler. OWLAx: A Protégé plugin to support ontology axiomatization through diagramming. In T. Kawamura and H. Paulheim, editors, *Proceedings of the ISWC 2016 Posters & Demonstrations Track co-located with 15th International Semantic Web Conference (ISWC 2016), Kobe, Japan, October 19, 2016.*, volume 1690 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2016.

[26] C. Shimizu, Q. Hirt, and P. Hitzler. A Protégé plug-in for annotating OWL ontologies with OPLa. In A. Gangemi, A. L. Gentile, A. G. Nuzzolese, S. Rudolph, M. Maleshkova, H. Paulheim, J. Z. Pan, and M. Alam, editors, *The Semantic Web: ESWC 2018 Satellite Events – ESWC 2018 Satellite Events, Heraklion, Crete, Greece, June 3-7, 2018, Revised Selected Papers*, volume 11155 of *Lecture Notes in Computer Science*, pages 23–27. Springer, 2018.

[27] C. Shimizu, P. Hitzler, Q. Hirt, A. Sheill, S. Gonzalez, C. Foley, D. Rehberger, E. Watrall, W. Hawthorne, D. Tarr, R. Carty, and J. Mixter. The Enslaved Ontology 1.0: Peoples of the historic slave trade. Technical report, Enslaved: Peoples of the Historic Slave Trade, March 2019. Available from `http://docs.enslaved.org`.

[28] C. Shimizu, P. Hitzler, and C. Paul. Ontology design patterns for Winston's taxonomy of part-whole relations. In M. G. Skjæveland, Y. Hu, K. Hammar, V. Svátek, and A. Lawrynowicz, editors, *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.*, volume 2195 of *CEUR Workshop Proceedings*, pages 2–13. CEUR-WS.org, 2018.

[29] M. van Erp, J. de Does, K. Depuydt, R. Lenders, and T. van Goethem.

Slicing and dicing a newspaper corpus for historical ecology research. In C. Faron-Zucker, C. Ghidini, A. Napoli, and Y. Toussaint, editors, *Knowledge Engineering and Knowledge Management - 21st International Conference, EKAW 2018, Nancy, France, November 12-16, 2018, Proceedings*, volume 11313 of *Lecture Notes in Computer Science*, pages 470–484. Springer, 2018.

[30] M. van Erp, M. Wevers, and H. C. Huurdeman. Constructing a recipe web from historical newspapers. In D. Vrandecic, K. Bontcheva, M. C. Suárez-Figueroa, V. Presutti, I. Celino, M. Sabou, L. Kaffee, and E. Simperl, editors, *The Semantic Web - ISWC 2018 - 17th International Semantic Web Conference, Monterey, CA, USA, October 8-12, 2018, Proceedings, Part I*, volume 11136 of *Lecture Notes in Computer Science*, pages 217–232. Springer, 2018.

[31] W. R. van Hage, V. Malaisé, R. Segers, L. Hollink, and G. Schreiber. Design and use of the Simple Event Model (SEM). *J. Web Semantics*, 9(2):128–136, 2011.

[32] D. Vrandecic and M. Krötzsch. Wikidata: a free collaborative knowledgebase. *Commun. ACM*, 57(10):78–85, 2014.