Towards an Ontology Network for the Reproducibility of Scientific Studies

Sheeba Samuel[1,2], Alsayed Algergawy[1] and Birgitta König-Ries[1,2]

[1]*Heinz Nixdorf Chair for Distributed Information Systems, Friedrich Schiller University Jena, Germany*
[2]*Michael Stifel Center Jena*

Abstract

Reproducibility is one of the fundamental characteristics of science. To reproduce scientific results, scientists need to manage and describe the provenance of end-to-end experimental pipelines. To understand, query, and reason how the results are derived, the provenance of the entire study needs to be described in an interoperable manner. Ontologies play an essential role in representing and interchanging provenance information generated in different systems, applications, and domains using a set of classes, properties, and restrictions. However, ontologies on describing provenance for scientific studies for different domains have been developed and used in isolation. They should be related to each other, aligned, and validated to form a network of interlinked ontologies, i.e., an ontology network. To this end, in this paper, we introduce *ReproduceMeON*, an ontology network for the reproducibility of scientific studies. The ontology network, which includes the foundational and core ontologies, attempts to bring together different aspects of the provenance of scientific studies from various applications to support their reproducibility. We present the development process of ReproduceMeON and the design methodology of developing core ontologies for the provenance of scientific experiments and machine learning using a semi-automated approach. We extend our scope to evolve ReproduceMeON to include ontologies for representing provenance for different subdomains like computational science, bioimaging, and microscopy.

Keywords

Ontology Network, Modeling, Core Ontology, Experiment, Machine Learning

1. Introduction

Reproducibility, the ability to get the same (or close-by) results when repeating an experiment under different conditions of measurement (e.g., experiment setup, method) [1], is essential for science as it helps scientists conduct better research in many ways: It allows researchers to check their results and verify the results of others, thus increasing trust in the scientific study. It also supports extending and building on top of others' works, thus promoting scientific progress. At the same time, achieving the reproducibility of scientific experiments is a complex real-world problem. Today, scientific studies in large collaborative research projects are often interdisciplinary and cover data and results from different disciplines. These scientific studies

involve experiments, their computational environment, wet lab experiments, workflows, computational experiments performed using data science or machine learning (ML) approaches, etc. Provenance, the source or origin of an object, plays a key role in the reproducibility of results. It helps understanding the data and sequence of steps performed by scientists, which led to creating the final result. Hence, researchers need to represent the provenance of results if they want to report the whole tale of the scientific study. Ontology-based descriptions of the provenance of data, steps, intermediary, and final results promise to enable reproducibility [2]. Since reproducibility is a complex domain and requirements to describe provenance and metadata for different research projects differ in specific aspects, it is not possible to build a large monolithic domain to cover the entire requirements. Instead, ontologies should be built in an integrated and modular way, forming a network. An ontology network or a network of ontologies (short: ON) is defined as a collection of single interconnected ontologies related to each other via various relationships such as alignment, modularization, and dependency relationships [3]. In this paper, we claim that ontologies for describing scientific studies for reproducibility should be organized as ON. We, therefore, introduce the methodology of developing the *ReproduceMe Ontology Network (ReproduceMeON)*, composed of ontologies that we have developed and others found in the state of art. For this, we first investigated the state of the art of ontologies in different relevant areas such as provenance, scientific experiments, ML, computational, microscopy, and scientific workflows through a systematic literature review. While the work involved in the development of *ReproduceMeON* touches upon many topics, the main focus of this paper is on the introduction of an abstract view of the ON architecture, where the proposed architecture is based on a three-layer view, including foundational, core, and domain ontologies. In particular, we propose a ontology matching-based approach to determine core concepts in each field (e.g., ML, provenance), which can be used later in the core ontology development. We focus here on how the ontologies have been automatically aligned and validated. The ontology alignment is done using three systems: OAPT [4, 5], AML [6] and LogMap [7].

2. Motivation and Use Case

We present a use case scenario showing the experimental workflow of the scientists we interviewed and collaborated with in projects like CRC ReceptorLight [2] and Werkstatt [8]. A Collaborative Research Center (CRC) is based on a number of interdisciplinary research projects consisting of several scientists possibly from different disciplines, who work together in teams towards a common goal. In our case, these include different subdisciplines of biology, medicine and computer science. The scientific studies conducted by researchers may consist of several computational and non-computational steps [2]. Figure 1 shows the need for developing an ON to describe the provenance of scientific studies in an interdisciplinary research project. Scientists perform wet lab activities like the preparation of samples and solutions, setting up the experiment's execution environment (room temperature, humidity), etc. These are non-computational steps which do not directly involve computational resources like computer, software, etc. In the next steps, several devices like a microscope to capture images of the receptor cell, an electrophysiological device to generate current, etc., are used in their experiments. The images
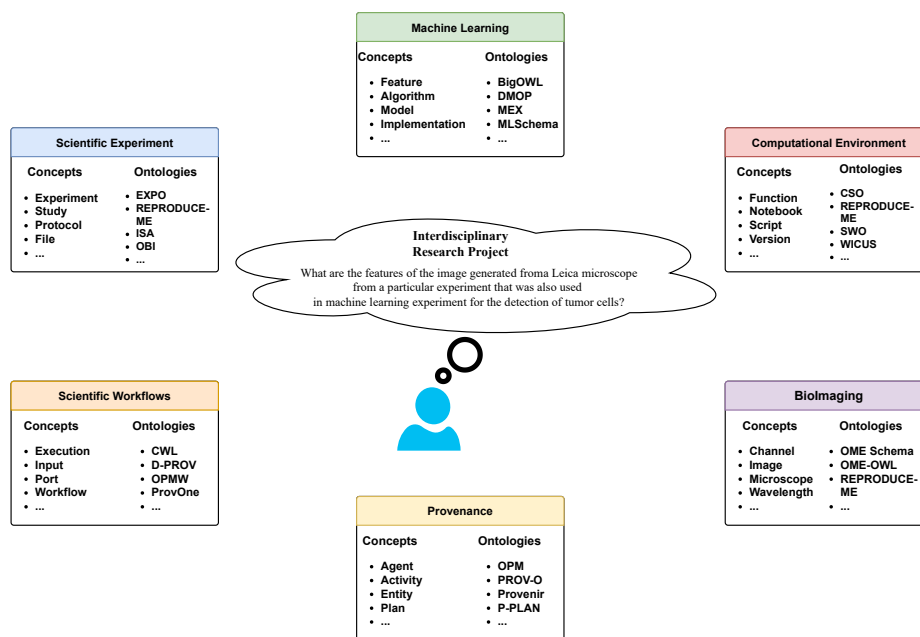
Machine Learning

Concepts	Ontologies
• Feature	• BigOWL
• Algorithm	• DMOP
• Model	• MEX
• Implementation	• MLSchema
• ...	• ...

Scientific Experiment

Concepts	Ontologies
• Experiment	• EXPO
• Study	• REPRODUCE-ME
• Protocol	• ISA
• File	• OBI
• ...	• ...

Computational Environment

Concepts	Ontologies
• Function	• CSO
• Notebook	• REPRODUCE-ME
• Script	• SWO
• Version	• WICUS
• ...	• ...

Interdisciplinary Research Project

What are the features of the image generated from a Leica microscope from a particular experiment that was also used in machine learning experiment for the detection of tumor cells?

Scientific Workflows

Concepts	Ontologies
• Execution	• CWL
• Input	• D-PROV
• Port	• OPMW
• Workflow	• ProvOne
• ...	• ...

BioImaging

Concepts	Ontologies
• Channel	• OME Schema
• Image	• OME-OWL
• Microscope	• REPRODUCE-ME
• Wavelength	
• ...	• ...

Provenance

Concepts	Ontologies
• Agent	• OPM
• Activity	• PROV-O
• Entity	• Provenir
• Plan	• P-PLAN
• ...	• ...

Figure 1: Motivation for developing an ON. The figure also shows our systematic literature review of existing ontologies in different areas related to reproducibility

acquired from microscopes are then analyzed by computational tools like proprietary software, scripts, or Jupyter notebooks based on the complexity of the problem and the skills of scientists. Reproducing a non-computational step is different from reproducing a computational step. The provenance of non-computational steps is usually neither machine-controlled nor automatic and often requires human involvement. Hence, the data, the steps, and the results from the computational and non-computational processes of a scientific experiment need to be interlinked and described in detail in an interoperable way [2]. In a complete workflow, the process starts with collecting data generated in the labs and moves on to analyzing and processing them using several computational techniques like ML.

The REPRODUCE-ME ontology [2] is our first attempt to developing an ontology to describe the complete path of a scientific experiment consisting of results from the computational and non-computational steps using semantic web technologies. The complete description, competency questions used, the development and evaluation of the REPRODUCE-ME ontology are explained in the paper [2]. It was developed by involving domain experts and computer scientists for the reproducibility of scientific experiments, initially focusing on the use case of biological imaging and microscopy [9]. It reuses existing ontologies, PROV-O [10] and P-Plan [11] and also models the provenance of the execution of scripts and computational notebooks like Jupyter notebooks. It was used and evaluated in the scientific data management platform CAESAR [2] in the CRC ReceptorLight project.

Though best practices were used in its development [12] and documentation[1] and it fulfills its

---

[1] https://w3id.org/reproduceme/

initial purpose, it was constructed in a monolithic way by providing all the terms based on the initial use case related to different areas and fields like biological imaging, microscopy, scripts, computational notebooks, etc. together in one OWL file. Over time, the need to modularize the REPRODUCE-ME ontology emerged with new requirements. New plans to reuse the ontology in other projects emerged as it provides the core concepts for describing scientific experiments' provenance. The reuse of computational provenance described by the ontology is used in projects like the FAIRification of the PREDICT workflow [13] and intended to be used in data science in ecological niche modeling [14]. The ontology is also used in computational tools like ProvBook [15] and ReproduceMeGit [16]. However, to use only its computational provenance part, currently, the whole ontology has to be imported into these tools and workflows, which in turn affects the reasoning and performance. The non-computational and computational aspects of the provenance of scientific studies were described in the ontology without identifying and separating the modules. Another requirement emerged from the Werkstatt project to describe the provenance of ML experiments [17]. However, it became challenging to extend the ontology in its current state. The lessons learned in the development of the REPRODUCE-ME ontology are used in the development of the ReproduceMeON[2]. An ON helps put together under one umbrella different modules required to describe the provenance of scientific studies.

3. Related Work

Several recent works have developed ON in different domains [18, 19, 20][3]. SEON [19] is a software engineering ON which is composed of a foundational ontology, two core ontologies, and several domain ontologies related to SE subdomains. Their alignment mechanism for integrating ontologies is by using the ontologies which are grounded in the foundational ontology and integrating two concepts if they have the same base type. Another recent approach presents developing an ON in human-computer interaction, HCI-ON [18] which is integrated to SEON [19]. New ontologies are added into the ON and aligned using their own annotation properties. The motivation behind building these ONs is to organize and structure knowledge in different domains. Many works have also pointed out the importance of modularizing ontologies [21]. Development of ON and the use of Ontology Design Patterns (ODP) are some of the available methods in the construction and management of modular and scalable ontologies [3, 21, 22]. Good modular ontologies should have good domain coverage, be formally rigorous, and reuse foundational ontologies according to [23].

Several ontologies have been developed covering different aspects of the reproducibility of scientific studies. In prior work [2], we have surveyed different provenance models and ontologies covering the computational and non-computational aspects of the reproducibility of scientific experiments. PROV-O, which provides fundamental concepts for the interoperable interchange of provenance information among heterogeneous applications and domains, is widely adopted by the scientific community and is reused and extended by different ontolo-

---

[2]To make it distinct, the REPRODUCE-ME ontology is a single ontology that was developed in [2] to describe the provenance of scientific experiments focusing on their computational and non-computational aspects. While the ReproduceMeON is a novel approach evolved from the REPRODUCE-ME ontology and contains a network of ontologies to describe the provenance of scientific studies.

[3]https://github.com/spice-h2020/SON, https://bimerr.iot.linkeddata.es/, https://github.com/rapw3k/glosis
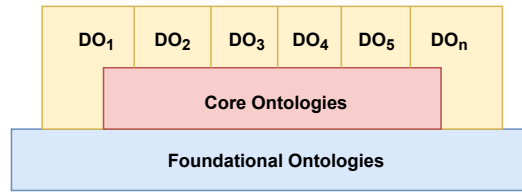
Figure 2: Three-layered ontology network architecture

gies [10]. Provenance ontologies like P-Plan, OPMW, D-PROV, DataONE, ProvONE have been mainly developed to represent computational processes in scientific workflows and to include specificities of particular Scientific Workflow Management Systems (SWfMS) [24]. In addition to the provenance and scientific workflow ontologies, various ontologies have been developed to capture the provenance of individual domains. The EXPO ontology [25] is developed to model scientific experiments by describing knowledge about experiment design, methodology, and results. The Ontology for Biomedical Investigations [26], developed as a community effort and widely adopted in the biomedical domain, describes experimental metadata in biomedical research, including planning, execution, and reporting. The recent work [27] presents the OWL representation of biological imaging data. Few ontologies have also been developed to describe computational provenance. Software Ontology (SWO) [28] models the data, the version, and the license used by the software. The REPRODUCE-ME ontology models the provenance of scientific experiments, bioimaging, scripts, and computational notebooks and their execution [2]. The ReproduceMeON is an initial novel approach to bring different ontologies together for representing the provenance of scientific studies for their reproducibility. With the development of the ReproduceMeON, the ability to align and import relevant modules from these ontologies becomes smooth. The design of ReproduceMeON considers important characteristics like being modular, considers international standards and reuses foundational ontologies. Our work aims to implement an ON by applying the characteristics and guidelines for developing modular, scalable, and reusable ontologies.

4. Development of an Ontology Network

In this section, we introduce the design and development scheme of *ReproduceMeON*. It is a novel approach that brings together knowledge from several domains, such as ML, provenance, and scientific computing, based on the three-layered architecture, as shown in Figure 2. Furthermore, the proposed approach builds on existing ontologies to enhance knowledge sharing and reuse. In general, Figure 2 shows that ON is organized into three layers: foundational, core, and domain-specific ontologies. According to this structure, we have to answer the following questions:

RQ1 Which are the foundational, core, and domain ontologies that compose the network?

RQ2 Which concepts and relations must be generalized to belong to a core ontology and specialized to belong to domain-specific ontologies?

RQ3 How should these ontologies in the ON be organized and relate to each other?

In the following, we describe how we can answer these questions.

4.1. Reproducibility Related Area Assimilation

To investigate existing ontologies in the area of reproducibility of scientific studies, we performed a systematic literature review [29]. The need for a systematic review arises from the requirement to develop an ON for the reproducibility of scientific studies by bringing together the existing ontologies that have been developed and used by researchers in different domains. The systematic review answers the research question **RQ1**. We used Google scholar to identify the existing ontologies in different areas related to reproducibility from 2006 to 2019. We limited the search to the following areas: *Provenance*, *Scientific Experiments*, *Scientific Workflows*, *Computational*, *Machine Learning*, and *Bio-imaging*. Information about the ontologies, including the developed year, imported ontologies, documentation, availability, content negotiation, formalization, and statistics, is available[4].

We found nine ontologies in Provenance, seven ontologies in Scientific Experiments, three in Bio-imaging, three ontologies in Computational, five ontologies in the ML domain. OPM, PROV-O, Provenir, P-Plan, OPMW, D-PROV, ProvOne, Research Object Ontology, Common Workflow Language are the ontologies we found in the area of provenance and scientific workflows. EXPO, SUMO, OBI, SMART Protocols, Investigation, Study Assay (ISA), The Minimum Information for Biological and Biomedical Investigation (MIBBI), Bioschemas, REPRODUCE-ME are the ontologies we found in the area of scientific experiments. MEX Ontology, ML Schema, Prov-ML, BigOWL, and DMOP are the ontologies we found in the area of ML. Software Ontology (SWO), WICUS ontology, Function Ontology, REPRODUCE-ME are the ontologies we found in the area of computational experiments and environment. We found OME Schema, REPRODUCE-ME, Ontology for an Integrated Image Analysis Platform, and Cellular Microscopy Phenotype Ontology (CMPO) in the area of bioimaging and microscopy.

We had to exclude some ontologies for the next phase of our study of identifying core ontologies because of the unavailability of the ontologies. Some are available in a different format other than the format used in ontologies (e.g., OME Schema is available in XML format). Table 1 shows a snapshot of the ontologies collected using the systematic literature review.

4.2. Core Ontologies Identification

The outcome of the first step is a set of reproducibility-related domains. In each domain, a number of existing ontologies have been identified and selected to construct the ON. To follow the three-layered architecture shown in Fig. 2, we need to opt which ontology in each domain can be used as a core ontology. After that, we build links between the core ontology and the other ontologies in the same domain (*intra-domain links*) and then build up links between ontologies from different domains (*inter-domain links*).

We started by investigating all collected ontologies, and if there was one that was well-defined and commonly used as a core ontology in the domain, we selected it as a representative core

---

[4]https://github.com/fusion-jena/ReproduceMeON

Ontology	Coverage	Serialization	Some Concepts
PROV-O	Provenance	TTL	Entity, Activity, Agent, ...
EXPO	Experiment	OWL	ScientificExperiment, ExperimentalTechnology, ...
ISA	Experiment	OWL	Investigation, Study, Assay, ...
SMART Protocol	Experiment	OWL	ExperimentalProtocol, LaboratoryProcedure, ...
REPRODUCE-ME	Experiment	OWL	Experiment, Dataset, Instrument, ...
MEX	ML	OWL	Algorithm, ClassificationProblem, Feature, ...
MLSchema	ML	TTL	Algorithm, Model, Run, ...
DMOP	Data Mining	OWL	ClassificationProblem, DataCharacteristic, ...
OME Schema	Microscopy	XML	Image, ImagingEnvironment, Instrument
Software Ontology	Computational	OWL	License, Software, SoftwareDevelopmentProcess, ...

Table 1
A snapshot of the ontologies collected using the systematic literature review
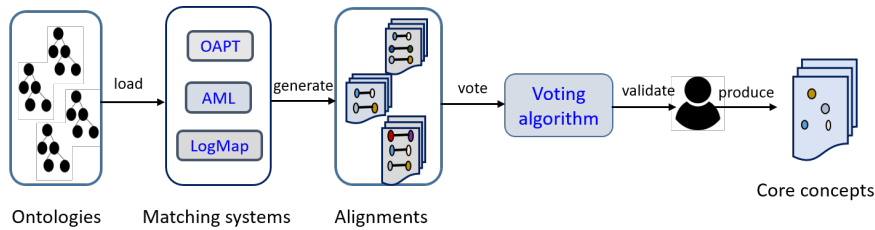

Figure 3: Core concepts determination

ontology. For example, the *PROV-O* ontology [10] is widely used in the provenance domain, providing the foundation to implement provenance applications in different domains, exchange, and integrate provenance information. Therefore, we selected it as the core ontology for the provenance part of the ON. For the other parts/domain composing the ON where it is hard to decide which ontology can be used as a core ontology, we propose an ontology matching-based approach. The general architecture of the proposed approach is shown in Fig. 3, where the set of ontologies belonging to each domain are loaded into three different matching tools. Each matching tool generates a matching result (an alignment) for every pair of ontologies.

- *Ontology loading:* For each domain, the collected set of ontologies are loaded and pre-processed (if needed) to be ready for matching, where each ontology pair constitutes a matching task. i.e., if we have n ontologies belonging to a domain, then $n \times \frac{(n-1)}{2}$ matching tasks are generated.

- *Alignment generation:* Ontology matching is a well-know solution to identify similar entities across a set of different ontologies [30]. We adopt the same idea to determine *intra-domain*, and *inter-domain* links during the development of the ON. Furthermore, together with a voting algorithm [31], ontology matching can help locate core concepts in each part/domain of the ON. To this end, we consider three well recognized matching systems, *LogMap*[7], *AML* [6], and *OAPT* [4, 5], as shown in Fig. 3. We implement a

```
<map>
    <Cell>
        <entity1 rdf:resource="http://www.opmw.org/ontology/WorkflowTemplate" />
        <entity2 rdf:resource="http://www.e-lico.eu/ontologies/dmo/DMOP/DMOP.owl#WorkflowTemplate" />
        <measure rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.9863</measure>
        <relation>=</relation>
    </Cell>
</map>
```
AML

```
<map>
    <Cell>
        <entity1 rdf:resource="http://www.opmw.org/ontology/WorkflowTemplate" />
        <entity2 rdf:resource="http://www.e-lico.eu/ontologies/dmo/DMOP/DMOP.owl#WorkflowTemplate" />
        <measure rdf:datatype="xsd:float">0.7</measure>
        <relation>=</relation>
    </Cell>
</map>
<map>
    <Cell>
        <entity1 rdf:resource="http://www.opmw.org/ontology/WorkflowTemplate" />
        <entity2 rdf:resource="http://www.e-lico.eu/ontologies/dmo/DMOP/DMOP.owl#WorkflowTemplate" />
        <measure rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.9976413490102886</measure>
        <relation>=</relation>
    </Cell>
</map>
```
LogMap

OAPT

Figure 4: A mapping example using three matching systems: AML, LogMap, and OAPT

pair-wise matching, where each pair of ontologies from the same domain is loaded into a matching system constituting a matching task. The corresponding matching result (*alignment*) is generated and saved into a local repository. An *alignment* is a set of mappings, usually expressed using the RDF alignment format defined by the ontology matching community. Each *mapping* (also called a correspondence), is a quintuple $< id, e, e', c, rel >$ where: id denotes a unique identifier of the mapping; e and e' are entities from two ontologies O and O' respectively; c denotes a measure of confidence, typically a value within the interval $[0, 1]$, and rel denotes the semantic relation between e and e' (equivalence ($\equiv$), more specific ($\sqsubseteq$), more general ($\sqsupseteq$), disjunction ($\perp$)). In the current implementation, we consider only equivalence ($\equiv$) relations.

The mapping example, shown in Fig. 4, illustrates that there is a corresponding between the *"WorkflowTemplate"* entity from the *BigOwl* ontology and the entity *"WorkflowTemplate"* from the *DMOP* ontology with different confidence values according to the used matching algorithm. This explains why we consider three different matching tools to achieve the task, as each tool measures the similarity between ontologies' entities based on different aspects.

- *Voting:* A vote corresponds to the number of times a mapping appeared in the sets generated by the matching systems. The consensus of vote 2, for instance, will contain mappings suggested by at least two systems. The more votes, the smaller is the size of the consensus alignment. We computed the number of mappings produced by applying Vote 2 and Vote 3 algorithms for scientific experiments and ML. Results are reported in Tables 2 and 3.

- *Alignment validating:* Generated alignments are in general not sufficient to be used to extract core concepts which can be used later for developing the core ontology for many reasons. First, they allow only a comparison of the systems to each other. Second, they may contain erroneous mappings, especially if the considered systems use the same background resources. And finally, valid alignments that have been found by only one system or none of them will be missing. For this reason, we first validate the generated

	Vote 2	Vote 3
No. of mapping	59	14
AML	83%	100%
LogMap	86%	100%
OAPT	54%	100%

Table 2

Voting for ML domain

	Vote 2	Vote 3
No. of mapping	82	22
AML	93%	100%
LogMap	37%	100%
OAPT	96%	100%

Table 3

Voting for Experiment domain

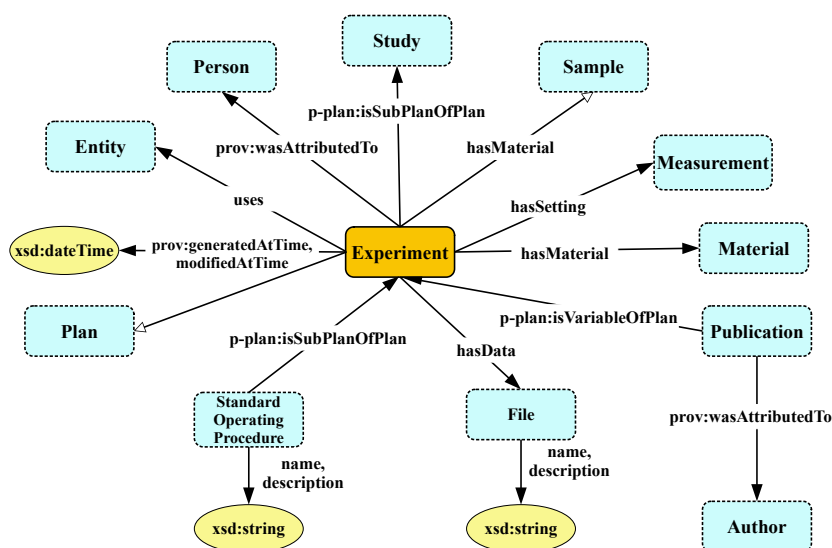alignments and add missing concepts based on our knowledge and experience.

Figure 5: A portion of schema diagram of the core ontology for scientific experiments. The orange-filled rectangular box represents the class that the diagram is depicting. The blue-filled rectangular boxes represent other classes in the ontology. The yellow-filled oval represents a data type. A subclass relationship is represented by an arrow with a white head and no label. An arrow with a solid tip represents the relationship mentioned in the label. The class at the solid tip of the arrow represents the range and the class at the other end of the arrow represents the domain of the relationship.

- *Core concept producing:* For each domain, the set of validated alignment is used to extract concepts that will be used later during the development of the core ontology. We consider both entities (concepts, relations) from each mapping. For example, the Vote 3 algorithm generate $< id, \text{`}http : //www.khaos.uma.es/perception/bigowl Algorithm', \text{`}http : //www.w3.org/ns/mlsAlgorithm', 1, \equiv >$, where the 'Algorithm' entities from *BigOwl* and *MLSchema* ontologies. As a result, we consider the term *Algorithm* as a core term for the ML domain. 'CrossValidation' is another core concept generated using the Vote 3 algorithm from the DMOP and Mex-Core ontologies. Figure 5 shows a portion of the conceptual model of the core ontology for scientific experiments. The classes are

generated from the validated alignments generated through this step. 'Study' is one of the core concepts validated using the Vote 3 algorithm from the ISA and REPRODUCE-ME ontologies. Currently, there are 37 core classes identified for scientific experiments and 35 classes for ML through this pipeline.

5. Discussion

We presented an abstract view of the ReproduceMeON and introduced the development of core ontologies focusing on the area of scientific experiment and ML. However, there are some open questions in the development of an ON. In the current state of the art, several approaches exist to align concepts from different ontologies. The most common approach is to import the entire ontology for the alignment between some concepts of the main ontology. Another approach is the usage of xref statements [32]. Using own annotation properties to align between classes from different ontologies is another approach. Each approach has its own pros and cons. However, importing entire ontology for alignment between concepts of different ontologies can affect performance and modularity. The ON design should be modular and should not affect the reasoning power and performance. In our approach of developing the proposed ON, we need to connect not only the ontologies from same domain but also from different domains. The question is whether we can use the same approach that we used in linking ontologies in the same domain to different domains. We also need to design ontologies so that we can easily plugin different ontologies into the network. The core relationships between the concepts also need to be identified and generated. Once the core ontologies are developed for the domains listed in Section 4, the linking of the core ontology to domain ontologies need to be determined. We plan to do a study on how new ontologies can be integrated into the ON, which are not aligned with the foundational and core ontologies. We plan to address these open questions in our future work. We also plan to do an extensive evaluation involving domain experts from each domain that we have selected for our current ON. The developed ontology will then be used to apply in the research projects as mentioned in Section 2.

6. Conclusion

This paper presented the need to build and organize ontologies for describing scientific studies for their reproducibility in an integrated and modular way, forming an ON. We introduced ReproduceMeON, an ontology network for the reproducibility of scientific studies. As the work involved in developing an ON is vast, in this paper, we focused on the abstract view of the ON architecture and the methodology for its development. We conducted a systematic literature review on the state of the art ontologies in different areas in provenance, scientific experiments, ML, computational, microscopy, and scientific workflows. We used the result from the review to develop the proposed ON, which includes foundational, core, and domain-specific ontologies for representing provenance for different areas like scientific experiments, computational science, biological imaging and microscopy, and ML. We use ontology matching techniques to select and develop core ontology for each sub-domain and link to other ontologies in the sub-domain. In the ON, we plan to build intra-domain and inter-domain links between the core ontologies and other ontologies from the same and different domains. In addition to the ontologies that are

already integrated or planned to be integrated into the network, we expect ReproduceMeON to continuously expand by incorporating other provenance ontologies related to the reproducibility of scientific studies.

Acknowledgments

References

[1] B. N. Taylor, C. E. Kuyatt, Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results, Technical Report, NIST Technical Note 1297, 1994.

[2] S. Samuel, A provenance-based semantic approach to support understandability, reproducibility, and reuse of scientific experiments, Ph.D. thesis, University of Jena, Germany, 2019.

[3] P. Haase, S. Rudolph, Y. Wang, S. Brockmans, D1. 1.1 networked ontology model (2006).

[4] A. Algergawy, S. Babalou, M. J. Kargar, S. H. Davarpanah, Seecont: A new seeding-based clustering approach for ontology matching, in: Advances in Databases and Information Systems - 19th East European Conference, ADBIS 2015, Poitiers, France, September 8-11, 2015, Proceedings, Springer, 2015, pp. 245–258.

[5] A. Algergawy, S. Babalou, F. Klan, B. König-Ries, Ontology modularization with OAPT, J. Data Semant. 9 (2020) 53–83.

[6] D. Faria, C. Pesquita, E. Santos, I. F. Cruz, F. M. Couto, Agreementmakerlight 2.0: Towards efficient large-scale ontology matching, in: Proceedings of the ISWC 2014 Posters & Demonstrations Track, ISWC 2014, Riva del Garda, Italy, October 21, 2014, volume 1272 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2014, pp. 457–460.

[7] E. Jiménez-Ruiz, B. C. Grau, Logmap: Logic-based and scalable ontology matching, in: The Semantic Web - ISWC 2011 - 10th International Semantic Web Conference, Bonn, Germany, October 23-27, 2011, Proceedings, Part I, 2011, pp. 273–288.

[8] S. Samuel, M. Shadaydeh, S. Böcker, B. Brügmann, S. F. Bucher, V. Deckert, J. Denzler, P. Dittrich, F. von Eggeling, D. Güllmar, et al., A virtual "werkstatt" for digitization in the sciences, Research Ideas and Outcomes 6 (2020).

[9] S. Samuel, B. König-Ries, REPRODUCE-ME: ontology-based data access for reproducibility of microscopy experiments, in: The Semantic Web: ESWC 2017 Satellite Events - ESWC 2017 Satellite Events, Portorož, 2017, pp. 17–20.

[10] T. Lebo, S. Sahoo, D. McGuinness, K. Belhajjame, J. Cheney, D. Corsar, D. Garijo, S. Soiland-Reyes, S. Zednik, J. Zhao, PROV-O: The PROV Ontology, W3C Recommendation 30 (2013).

[11] D. Garijo, Y. Gil, Augmenting PROV with plans in P-Plan: scientific processes as linked data, CEUR Workshop Proceedings, 2012.

[12] A. Gangemi, V. Presutti, Ontology design patterns, in: Handbook on ontologies, Springer, 2009, pp. 221–243.

[13] R. Celebi, J. R. Moreira, A. A. Hassan, S. Ayyar, L. Ridder, T. Kuhn, M. Dumontier, Towards fair protocols and workflows: the openpredict use case, PeerJ Computer Science 6 (2020) e281.

[14] M. L. Mondelli, A. Townsend Peterson, L. M. R. Gadelha, Exploring reproducibility and fair principles in data science using ecological niche modeling as a case study, in: Advances in Conceptual Modeling, Springer International Publishing, Cham, 2019, pp. 23–33.

[15] S. Samuel, B. König-Ries, Provbook: Provenance-based semantic enrichment of interactive notebooks for reproducibility, in: Proceedings of the ISWC 2018 Posters & Demonstrations, Industry and Blue Sky Ideas), USA, 2018, 2018.

[16] S. Samuel, B. König-Ries, Reproducemegit: A visualization tool for analyzing reproducibility of jupyter notebooks, in: B. Glavic, V. Braganholo, D. Koop (Eds.), Provenance and Annotation of Data and Processes, Springer International Publishing, Cham, 2021, pp. 201–206.

[17] S. Samuel, F. Löffler, B. König-Ries, Machine learning pipelines: Provenance, reproducibility and fair data principles, in: Provenance and Annotation of Data and Processes, Springer International Publishing, Cham, 2021, pp. 226–230.

[18] S. D. Costa, M. P. Barcellos, R. de Almeida Falbo, M. V. H. B. Castro, Towards an ontology network on human-computer interaction, in: Conceptual Modeling - 39th International Conference, ER 2020, Vienna, Austria, November 3-6, 2020, Proceedings, volume 12400 of *Lecture Notes in Computer Science*, Springer, 2020, pp. 331–341.

[19] F. Borges Ruy, R. de Almeida Falbo, M. Perini Barcellos, S. Dornelas Costa, G. Guizzardi, Seon: A software engineering ontology network, in: Knowledge Engineering and Knowledge Management, Springer International Publishing, Cham, 2016, pp. 527–542.

[20] C. Sila, O. Belo, V. Barros, Methodology for the development of an ontology network on the brazilian national system for the evaluation of higher education (ontosinaes), JISTEM-Journal of Information Systems and Technology Management 15 (2018).

[21] F. Ensan, W. Du, A modular approach to scalable ontology development, in: Canadian Semantic Web, Springer, 2010, pp. 79–103.

[22] M. C. Suárez-Figueroa, NeOn methodology for building ontology networks: specification, scheduling and reuse, Ph.D. thesis, Technical University of Madrid, 2012.

[23] M. d'Aquin, A. Gangemi, Is there beauty in ontologies?, Appl. Ontol. 6 (2011) 165–175.

[24] J. Liu, E. Pacitti, P. Valduriez, M. Mattoso, A survey of data-intensive scientific workflow management, J. Grid Comput. 13 (2015) 457–493.

[25] L. Soldatova, R. King, An ontology of scientific experiments, Journal of the Royal Society Interface 3 (2006) 795–803. URL: http://journals.royalsociety.org/content/u552845783800t73/fulltext.pdf. doi:10.1098/rsif.2006.0134.

[26] R. R. Brinkman, M. Courtot, D. Derom, et al., Modeling biomedical experimental processes with OBI, J. Biomedical Semantics 1 (2010) S7.

[27] J. Moore, et al., On bringing bioimaging data into the open(-world), 2019. URL: http://ceur-ws.org/Vol-2849/paper-06.pdf.

[28] J. Malone, A. Brown, A. L. Lister, J. Ison, D. Hull, H. Parkinson, R. Stevens, The Software Ontology (SWO): a resource for reproducibility in biomedical data analysis, curation and digital preservation, Journal of Biomedical Semantics 5 (2014) 25. doi:10.1186/2041-1480-5-25.

[29] B. Kitchenham, Guidelines for performing Systematic Literature Reviews in Software Engineering, Technical Report EBSE-2007-01, Keele University and University of Durham, 2007.

[30] P. Ochieng, S. Kyanda, Large-scale ontology matching: State-of-the-art analysis, ACM Comput. Surv. 51 (2018) 75:1–75:35.

[31] I. Harrow, E. Jiménez-Ruiz, A. Splendiani, M. Romacker, P. Woollard, S. Markel, Y. Alam-Faruque, M. Koch, J. Malone, A. Waaler, Matching disease and phenotype ontologies in the ontology alignment evaluation initiative, Journal of biomedical semantics 8 (2017) 1–13.

[32] A. Laadhar, E. Abrahão, C. Jonquet, Investigating one million xrefs in thirthy ontologies from the obo world, in: 11th International Conference on Biomedical Ontologies, 2020.