Exploring the Use of Ontologies and Reasonable Ontology Templates (OTTR) within Health Registry Systems

Victoria Varzhel

Informatics: Design, Use, Interaction
60 ECTS

University of Oslo
Department of Informatics

15.11.19

ABSTRACT

In this thesis we explored the use of ontologies and Reasonable Ontology Templates (OTTR) within the context of health registries in Norway. We contributed to further knowledge about this topic by reusing the existing ontologies and applying OTTR templates to a real-life case - health registries. The project starts with presenting the theoretical background about ontologies, their components and languages, ontology reuse, design patterns and OTTR. The presented use case is the use of health registries in Norway. Responsibility for the national registries is shared among different institutions, which results in lack of standardisation and data quality.

To answer the research questions, we reused the PICO and CSEO ontologies and designed the OTTR templates for the health registry variables. Results demonstrated that (1) reuse of biomedical ontologies occurs through merging ontologies and their modules, but it entails manual effort from both an ontology engineer, and a domain expert; (2) despite some issues, the OTTR templates can be applied to the real life case of health registry variables. The created templates can potentially be used in a registry builder application where templates will be used to construct an OWL file with the necessary variables for the registry.

SUMMARY

My contributions to the research in this thesis are:

- Formulating the research questions with the help of my supervisor;
- Studying the theoretical background and the use case;
- Writing the competency questions with the help of my supervisor and a medical professional;
- Manual analysis and evaluation of the existing biomedical ontologies to be reused;
- Designing the Health Registry Variables ontology in Protege, reusing PICO and CSEO ontologies;
- Designing the OTTR templates with the help of the OTTR expert.

ACKNOWLEDGMENTS

I would like to take this opportunity to offer my deepest thanks to my supervisor, Laura Slaughter for her guidance and support. My supervisor consistently allowed this paper to be my own work, but steered me in the right direction whenever she thought I needed it. Thank you for all the invaluable comments, insights and discussions.

Furthermore I would like to thank Daniel Lupp for his help with designing of OTTR templates and Linn Brandt for advising me on the subjects of competency questions and the use case.

Last but not least, I would like to extend my gratitude to my family and friends for their continued encouragement and support, especially to Michal Przybylski, who has supported me all along the way.

Victoria Varzhel
Oslo, November 2019

TABLE OF CONTENT

LIST OF ABBREVIATIONS

ADL	Activity of Daily Living
BFO	Basic Formal Ontology
CP	Content Ontology Design Patterns
CQs	Competency Questions
CSEO	Cigarette Smoking Exposure Ontology
DAG	Directed Acyclic Graph
DL	Description Logic
DNT	Door to Needle Time
EHR	Electronic Health Records
FHHO	Family Health History Ontology
FMA	Foundational Model of Anatomy
FOL	First Order Language
HRV	Health Registry Variable
MED	Medical Entities Dictionary
NCI	National Cancer Institute
NLP	Natural Language Processing
ODP	Ontology Design Patterns
ORSD	Ontology Requirements Specification Document
OTT	Onset to Treatment Time
OTTR	Reasonable Ontology Templates
OWL	Web Ontology Language
PCAO	Prostate Cancer Ontology
PICO	Population/ Intervention/ Comparison/ Outcome
RCD	Read Codes
RDF(S)	Resource Description Framework (Schema)
XML	Extensible Markup Language

LIST OF IMAGES AND TABLES

1. INTRODUCTION

Digitalisation and standardization is one of the goals of e-health. There are numerous ontologies that are available in the biomedical domain. Despite all the standardization work there are still some areas in health and medical administration that lack the necessary standard level. In this master thesis I discussed the use case of registry variables in norwegian medical registries and how the reuse of ontologies and Reasonable Ontology Templates can be applied in creating better systems to manage registries and registry data.

1.1. BACKGROUND

Nowadays it is impossible to separate biomedical informatics from the development of computer technologies (Masic, 2014). Medical informatics is a constantly evolving field, that has progressed significantly during the past years. The volume of patient data is increasing constantly, including laboratory and diagnostic tests, social information, CAVE etc.(Dugas *et al.*, 2016). Digitalisation and data integration in medicine are important fields of research (Prasser *et al.*, 2018) and (Smith *et al.*, 2007). Solving the problem of data integration will improve the quality of medical services. There are 820 biomedical ontologies that are currently represented on Bioportal (*NCBO BioPortal*, no date). They represent different domains in the health sector. It is often the case that domains in these ontologies overlap. That raises the question of how can the process of creating the new ontologies be improved.

One of the solutions for efficient ontology creation is ontology reuse methods and ontology matching. There are studies about ontology reuse via modularisation (Doran, 2006), matching and integration techniques (Caldarola and Rinaldi, 2018), linguistic matching (Caldarola and Rinaldi, 2016), etc. Despite the amount of studies on ontology reuse, there is no particular method which is agreed to be used when reusing ontologies (Zulkarnain, Meziane and Crofts, 2016). Also, there are some advantages and disadvantages of applying the ontology reuse methods as presented in this thesis.

Another way to address the challenge of ontology creation is through applying the Ontology Design Patterns (ODPs). Studying the recurring patterns in ontologies and its reuse is one of the focuses of ontology researchers. There are several types of ODPs, that are used for different purposes (*OPTypes - Odp*, no date). ODPs can simplify the creation of new ontologies and ensure uniform modeling. Ontology templates were introduced as a way to "define and instantiate recurring patterns within an ontology" (Lupp, 2019).

When developing different ontologies, we need to be able to evaluate them. There are various evaluation approaches that are discussed by ontology researchers, e.g. gold standard, corpus-based, task-based and criteria-based (Raad and Cruz, 2015). An ontology can be evaluated from different perspectives too, for instance quality and correctness, which implies different methods (Hlomani, 2014).

1.2. OBJECTIVES

The main focus of this thesis is to explore how can we reuse existing ontologies and apply OTTR templates within a real life case of Norwegian health registries. We studied existing biomedical ontologies, modeling issues, evaluation and OTTR templates. We also applied the templates to an application which can be used for creation of new registry ontologies.

1.3. RESEARCH QUESTIONS

The main goal of this project is to facilitate the creation of registry variables that are consistent and that convey clear machine-readable definitions of the registry data to be collected. Towards these goals, this thesis explores the following research questions:

1. How can we reuse existing biomedical ontologies towards the goal of facilitating data integration of registries?
2. What are the design issues when applying OTTR templates to health registries variables modelling and implementing a ontology-driven system?

1.4. METHOD SUMMARY

Here we summarize methods that were used during the research. Refer to Chapter 4 for more details. The following methods were used to answer the research questions:

- designing of competency questions, that were used for further evaluation of the created ontology;
- extending the PICO ontology to model the Health Registry Variables ontology in Protege;
- evaluation and analysis of the ontologies to be reused;
- designing the OTTR templates.

1.5. THESIS OUTLINE

This master thesis has the following outline:

- Chapter 1 is an introduction to this master thesis. We present the background for the thesis, review the objectives, list the research questions and summarize methods.
- Chapter 2 focuses on the theoretical background behind ontologies. The definitions, main concepts, languages and description logistics are mentioned. Moreover, we discuss biomedical ontologies which are the type of ontologies that are relevant for this thesis. In this chapter we also present the theory behind the two focus areas of this thesis: Ontology Reuse and ODPs.
- Chapter 3, Use Case, describes the environment of our work. We present the current status of Health Data Program and medical registries in Norway. We also outline the standardization issues between registries and discuss how to improve the interoperability between registries and EHR.
- In Chapter 4, Methods, we describe in detail “how” we answer the research questions asked in the Introduction (Chapter 1.3). This includes the methods of reuse of ontologies, their evaluation and creating the OTTR templates.
- Chapter 5, Results, refers directly to the research questions outlined in Chapter 1.3. In this chapter we present an example of reusing the biomedical ontologies, results of Health Registry Variables ontology evaluation and OTTR templates.
- Last Chapter of this thesis, Chapter 6 - Discussion of Results, summarizes all the findings and experiences acquired during our work. In this chapter we focus on showing the “pros and cons” of both ontology reuse and OTTR templates as well as proposing a future work based on this thesis.

2. THEORETICAL BACKGROUND

In this chapter we are discussing what an ontology is and its main components. The problem of medical ontologies is approached from different angles. We looked at works focusing on interoperability between registries and EHR, methods for ontology merging and finally practice of modelling the biomedical ontologies.

2.1. ONTOLOGY

Data processing has become an essential part of solving any task or problem in multiple fields of human knowledge , for instance engineering, IT, medicine, social sciences etc. The volume and variety of data is increasing and can be gathered from various sources, both structured and unstructured. Data storage is often up to IT experts who are familiar with the tools to extract the data. Thus domain experts often have to learn the ways to work with data or rely on IT specialists. Another issue is maintenance. In most cases data is not something static, it changes and needs to be constantly updated.

Ontology, as a philosophical term, is “the discipline investigating the structure of reality, which can be structured into a series of integrative levels, such as the physical, the biological, the mental and the cultural, and each level acts as a base for the emergence of more complex levels.”(Chatterjee, 2017). Ontology, as used in AI, refers to an engineering artifact, which consist of a vocabulary (some domain of knowledge that describes a certain nature) and relations and hierarchy between the vocabulary concepts.

Ontology is an abstract model that represents the domain of knowledge and is independent of how data is stored. In an ontology-based data access, end-users do not need to know the technical part of an ontology to extract data. This can be done with the help of queries which are close to a human language, thus quite simple to use. “The ontology is connected to data sources with the help of mappings which translate between queries over the ontology to queries over the data sources.” (Lupp, 2019).

Designing a high-quality ontology is still a debatable topic among researchers. Nevertheless, it should be a combination of an accurate representation of the end-user’s domain understanding and correct information requirements kept. This requires a close collaboration between ontology engineers and domain experts. Also ontologies should be easily maintainable and allow automation whenever possible, which “eliminates potential sources of human error.” (Lupp, 2019).

2.1.1. MAIN COMPONENTS OF ONTOLOGY

There are different kinds of knowledge representation formalisms for ontologies. Despite the fact that all of them may be very different, they have a common minimal set of components, which are listed further (Ruiz and Hilera, 2006).

Class is something that represents a concept, which is a core of most ontologies. For example a travelling domain where locations (cities, towns, etc.), housing (hotels, camping sites, apartments, etc.) and transportation (cars, trains, planes, ships, etc.) would symbolize concepts.

Classes can be organised in **taxonomies** which are also called **hierarchies**. A class can have a subclass that represents a more specific context than its parent. For example, cars can be sedan, SUV, pickup. etc.

There are different approaches for developing of class hierarchy (Noy and McGuinness 2001):

- top- down approach means defining the most general classes and continuing to specialization to subclasses;
- bottom-up approach starts with defining the most specific concepts and further moving on to their generalization;
- combination approach is a combination of both top-down and bottom-up development processes.

In some ontologies a class is allowed to have more than one parent, which is called multiple inheritance. That means that all the subclasses inherit information from multiple parents. If a class is allowed to have only one parent, it is called single inheritance. Some consider multiple inheritance to be a bad practice and may cause problems. According to (Takai-Igarashi *et al., no date)* “the problem will also become an obstacle for keeping consistency in the ontology.”

The ontology as seen by Thomas Gruber is a graph of nodes (concepts) connected by different types of links. These nodes and is-a relations form a Rooted Directed Acyclic Graph (DAG). Rooted means that there is a one “highest node” which is a concept. Directed means that is-a relation has a direction, which can point upward or downwards. Acyclic means that if you start from a node and continue to other nodes using is-a relation, you can never come back to where you started. Most nodes in the graph have other types of relationships than is-a, which are called semantic relations. Higher nodes in a graph mean more general concepts, lower represent specific concepts. There is also an unrooted DAG, which allows several roots. However most often an artificial root is introduced, for example “Thing”.

Attributes are usually aspects or parts to which objects in an ontology are related. It is a way to describe an object. Attributes can be classes or individuals and its value can be a complex data type. For example, a car can have such attributes as name and different parts, such as a special type of engine, transmission, etc.

Relations specify how classes are related to other classes, how they interact. There are two types of relations:

- taxonomies that organise classes into subclasses. It is "a kind of" or "is-a" relationship. For example, Enzyme is *a kind of* Protein which is a kind of Macromolecule. *hasComponent* (reverse of *isPartOf*) relation also falls under that category, for example a car *hasComponent engine;*
- associative relations that relate classes though the tree structure. They can be nominative (*hasName*), locative (hasLocation), the ones that represent functions or property (*hasFunction*) and many others (*What is an Ontology?*, no date)

Instances represent objects or individuals. For example Hemoglobin is an instance of class Proteins. Ontologies are not supposed to contain instances, because ontology should conceptualize the domain. Though sometimes it is challenging to tell the difference between an instance and a concept when modeling the ontologies. That might be very relative, depending on the ontology itself (*What is an Ontology?*, no date). For example *CRP* can be an instance of a *Blood Test* concept. At the same time CRP can be argued to be a concept that is represented by other instances.

Axioms is an assertion in logical form. The simple example of the subclass axiom would be Child *is a subclass of* a Person. Another example is that if Person A *isRelated* to Person B, that means that Person B *isRelated* to Person A.(Kaljurand, no date) (*What is an Ontology?*, no date).

2.1.2. WEB ONTOLOGY LANGUAGE

The logical languages can be classified into two groups: traditional and web-based languages. The traditional languages were developed in the field of AI by the 1990s (Rebstock, Janina and Paulheim, 2008). Most of the web-based ontology languages were based on XML syntax, but they have different terminologies and a level of expressiveness.

Web Ontology Language (OWL) became a W3C recommendation in 2004 and is currently the standard ontology language. It was built on Description Logics (ref. Chapters 2.1.3 and 2.1.4) and combines its expressiveness with Resource Description Framework (RDF) technology , e.g. URIs and namespaces.

OWL represents complex knowledge about concepts and relations between them. It is built on the RDF triple model, but has more possibilities to build expressions and more vocabulary than RDF(S). The expressions are formulated similar to the first-order logic (Rebstock, Janina and Paulheim, 2008). W3C classified OWL into three sub languages which are OWL Lite, OWL DL and OWL Full (Taye 2011).

- OWL Lite is the simplest and less expressive version of OWL, that allows to express relationships with maximum cardinality equal to 0 or 1;
- OWL DL uses Description Logic to express relations between things and their properties. It is built on OWL Lite;
- OWL Full is based on OWL Lite and DL and thus possesses the highest expressiveness of all three (Taye 2011).

There are three kinds of mutually disjoint properties in OWL: (1) Object property, (2) Datatype property and (3) Annotation property.

Object properties link individuals to individuals, for example personA *hasFriend* personB. Datatype properties link individuals to data types, for example Victoria *hasAge* 28. So they cannot be reflexive, symmetric, transitive, etc. The built-in datatypes can be: integer, long, double, string, boolean and others. Annotation properties are ignored by reasoners, it has no logical implication, for example *rdfs:label, dc:creator (OWL Web Ontology Language Overview, no date)*.

The second edition of OWL (OWL 2) was developed in 2011. It has a similar structure to OWL 1, but there are some significant differences between them. OWL 2 includes new functionalities and features, for instance keys, property chains, richer data types etc. Moreover, it includes new profiles (*OWL 2 Web Ontology Language Profiles (Second Edition)*, no date) and new syntax - Manchester Syntax (*OWL 2 Web Ontology Language Manchester Syntax (Second Edition)*, no date). Manchester Syntax is used in Protege as concept descriptions. It is close to DL constructs, but in the form of text, because DL symbols are hard to find on the keyboard.

OWL 2 is based on the DL SROIQ(D), which stands for:

- S for ALC + role transitivity;
- R for (complex) roles inclusions,
- O for closed classes;
- I for inverse roles;
- Q for qualified cardinality restrictions;
- D for data types.

OWL separates the notions of classes, properties and individuals, but OWL 2 introduces punning, which means that one URI is used both as a class and an individual. Semantically these are still two different things, which are treated differently (*Punning - OWL*, no date).

OWL 2 has different profiles according to different DLs (*OWL 2 Web Ontology Language Profiles (Second Edition)*, no date), for example:

- OWL 2 QL, designed for database integration;
- OWL 2 EL, language with polynomial time reasoning, suitable for large ontologies;
- OWL 2 RL, compatible with rule-based inference tools.

2.1.3. DESCRIPTION LOGICS

Description Logics (DL) is a "popular knowledge representation formalism in the context of the Semantic Web."(Hotz *et al.*, 2014). It is used for "formal description of concepts and roles (relations)."(Obitko, no date)

There are various DL languages (*AL, EL, DL-Lite, RL* etc.) which allow different logics. *ALC* (Attributive concept Language with Complements) allows just a certain set of concepts , for example atomic, universal, bottom, negation, union etc. *EL* on the other hand includes only the following concepts: atomic, universal, bottom, singular enumeration, intersection and existential restriction. It doesn't support negation, cardinalities, inverse roles etc. So every DL language allows a different set of concepts (Baader and Nutt, 2002).

Knowledge in DL is represented by two components, as presented below:

$$K = (T, A)$$

where:

- K is knowledge;
- T is a set of terminology definitions, "a vocabulary of an application domain" (Baader and Nutt, 2002) which is called TBox. For example $Woman \equiv Person \sqcap Female$;
- A ia a set of "assertions about named individuals in terms of this vocabulary" (Baader and Nutt, 2002) which is called ABox. For example *isBrotherOf (john, sam), Person(amy)*.

2.1.4. DESCRIPTION LOGIC IN BIOMEDICAL INFORMATICS

Biomedicine is a complicated domain of knowledge. It creates some issues to be faced by DL community:

- developing of scaled implementations;
- developing an architecture that satisfies the users' need for simplicity and formal constraints for managing the ontology;
- developing an ontology which is rich enough and can be reused;
- developing expressive formalisms, which are good enough to cope with medical constructs.

Description Logics were applied in at least five applications in medical informatics (Baader *et al.*, 2003):

- representation of information in health records or terminology development;
- intelligent UI;
- decision support and semantic indexing;
- semantics oriented NLP;
- semantic integration of IS.

However the biggest challenge is to use DL notations "in a form that users can use to meet real problems - whether in representation of medical records, indexing of information for decision support, or supporting user interfaces and natural language processing." (Baader *et al.*, 2003).

2.1.5. ONTOLOGY ENGINEERING

Ontology engineering has been a subject of interest for researchers for some time already. The ontology development process consists of different stages and processes. Also there are numerous methods to construct ontologies and each has pros and cons.

To build a good ontology it is important to identify its requirements. The Ontology Resource Specification Document (ORSD) is used to "to state why the ontology is being built, what its intended uses are, who the end-users are, and which requirements the ontology should fulfill"(Suárez-Figueroa, Gómez-Pérez and Villazón-Terrazas, 2009). There are several activities that ORSD includes:

- finding and reusing the existing knowledge resources so that they can be reengineered into ontologies;
- finding and reusing of the existing ontology resources , e.g. modules, design patterns etc.;

- verifying the ontology during its development.

The way an ontology is constructed depends on many factors. There multiple methods for ontology engineering. For instance, Alexander C. Yu in his work (Yu, 2006) pointed out several methods for constructing ontologies:

- Representation formalisms , e.g. KIF, LOOM, Ontolingua etc. For example KIF is based on declarative semantics. It focuses on describing facts than processes or procedures. “Knowledge can be described as objects, functions, relations, and rules...Its semantics are based on a subset of description logics (DLs).”(Yu, 2006)
- Fundamental ontological theories are based on principles of classification and subsumption. These are “fundamental theories that deal with, first, the relationships between classes and their instances and, second, the taxonomic relationships between classes.”(Yu, 2006)
- General ontology development methodologies are based on starting with manual collecting and coding of common sense knowledge. After extracting enough knowledge and entering it into the system, machine learning tools and tools to analyze natural language can be used to help with the process of collecting and adding more knowledge. These methodologies can be used for: (1) identifying the purpose of an ontology, (2) ontology developing, (3) ontology evaluation and (4) documentation.
- Top-level ontologies “describe the most general concepts or categories that are presumed to be common across domains.”(Yu, 2006) The examples are DOLCE, BFO etc. Top-level ontologies are used as a “formal foundation for building domain ontologies”. This can improve the integration between current and future ontologies. Domain ontology can be built first, then it can be integrated with the top-level ontology. The author gives an example - you, a surgeon’s scalpel, your arm, and your wristwatch. Your life, the movement of your blood through your blood vessels, and the execution of a surgical procedure would be the occurrence. There are many biomedical ontologies which are based on BFO (Basic Formal Ontology) which is an upper ontology (*Basic Formal Ontology (BFO) | Home*, no date). The advantages and disadvantages of top-level ontologies are discussed in detail in Chapter 2.6, Ontology Reuse .

(Bontas, Schlangen and Niepage, 2005) present a method to construct a semantic web based medical ontology “which is used for the semantic annotation of medical reports, and evaluate the engineering process against a reuse-oriented approach” (Bontas, Schlangen and Niepage, 2005). In their paper they discuss using NLP (Natural Language Processing) to analyze texts from target domain and its use in creating the ontology. They divided the developing of the ontology into several stages:

- Analysis of the domain. They were working with lung pathology, which requires a number of domain experts (pathologists) to identify the key terms used in the ontology.

- Conceptualization of the domain knowledge. At that stage they used an NLP tool (The OntoSeed Suite) to determine the list of occuring word sequences by analyzing the pathology reports. These groupings are clustered again by “hierarchical clustering algorithm” (Bontas, Schlangen and Niepage, 2005). That method helps relations between terms in the future ontology, for example sub-classes etc.
- Implementation, refinement and evaluation of a prototype ontology.

For example, such concepts as “tumor cell” and “lung tissue” (or Tumorzelle and Lungengewebe in german) were used to test the NLP method. These examples show that suffixes and prefixes can be useful when modeling the concepts. Tumor *cell* can be modeled as a subclass of cell concept or by means of property *locationOf*, meaning that cell is located in the tumor. Lung tissue is a similar example, as it can be a subclass of tissue concept or modeled by means of *locationOf* property. (Bontas, Schlangen and Niepage, 2005)

As we see there are multiple different ways to design an ontology. There is no one single perfect solution. A design method should be chosen according to the type and the purpose of the ontology to be created.

2.2. INTRODUCTION TO BIOMEDICAL ONTOLOGIES

Research on ontology has become a significant part of biomedical informatics. “Ontology design is becoming increasingly recognized as central to medical informatics and even more so to bioinformatics.”(Rosse and Mejino, 2003). New ontologies are developed all the time, in addition to numerous existing ones (*AberOWL*, no date). The more data there is the more there is a need for its structuring, and ontology is the best way so far to process data. “...the exponential increases in biological data and knowledge have also led to an awareness of the usefulness of ontological methods in biology and, hence, to subsequent efforts to exploit these techniques.”(Yu, 2006) Therefore designing ontologies that can be used by both IT- and domain experts is one of the key solutions for improving medical services.

Need for biomedical ontologies arises with time. More and more ontologies are created constantly. There are many examples of ontologies that are actively used nowadays, but there are some that are no longer in use (for example GALEN project).

- Foundational Model of Anatomy (FMA) is an example of one of the most structured biomedical ontologies. “It relied on a set of declared principles, high-level schemes, Aristotelian definitions, and a frame-based formalism.”(Yu, 2006). It was developed on the basis of the anatomical part of UMLS ontology and represents the human body from the macromolecular and microscopic levels.

- The Medical Entities Dictionary (MED) is based on concepts used in laboratory, pharmacy, billing systems and radiology. MED is used in real-world applications. “MED has been used to support various applications such as data retrieval from medical records, “just in time” medical education, expert systems, data mining, and knowledge-based terminology maintenance.”(Yu, 2006)

- National Cancer Institute (NCI) Thesaurus is a DL based terminology. It is used for data annotation, as well as its search and retrieval from NCI repositories. It was also intended that it will be used for “automated indexing, bibliographic retrieval, and linkage of heterogeneous resources” but several issues were discovered related to concept definitions, formation of terms, properties and DL representation (Yu, 2006).

- Gene ontology (*Gene Ontology Resource*, no date) is an example of a complex structure that combines three ontologies: Cellular component, Molecular function, and Biological process. Their combination allows to see the whole picture of the newly found genes, as three questions need to be answered when it happens:

  1. Where is it in the cell?
  2. Functions on the molecular level?
  3. Which biological processes are influenced?

  By using that model, GO can look at the same gene from different angles.

- (‘Unified Medical Language System (UMLS)’, 2009) is quite a big ontology. It includes various types of data from different knowledge fields. “Major groupings of semantic types include those for organisms, anatomical structures, biological function, chemicals, events, physical objects, and concepts or ideas.”(Yu, 2006) Not only biomedical data is introduced in that ontology.

- Finally, SNOMED CT (*SNOMED Home page*, no date) is the most comprehensive ontology that exists nowadays. It has the DL based structure which allows to represent various concepts and interrelations between them. That leads to minimizing the “concept redundancy and ambiguity.”(Yu, 2006) Despite that, SNOMED CT has various flaws in its design , for instance incomplete concept descriptions (Bodenreider *et al.,* 2004). Also (Ceusters et al., 2003) classified SNOMED problems into three categories:
  - not correct assignment of is-a relationship caused by human error;
  - shifts in meaning while migrating from SNOMED RT to CT and redundant concepts caused by technology flaw;
  - lack of ontological theory.

There are many existing medical ontologies that were designed in different ways and have their advantages and disadvantages. It is important while using an ontology in an application to understand its purpose, strengths and weaknesses and what it was designed for.

2.3. ONTOLOGY DESIGN PATTERN

“An ontology pattern is a set of ontological elements, structures or construction principles that intend to solve a specific engineering problem and that recurs, either exactly replicated or in an adapted form, within some set of ontologies, or is envisioned to recur within some future set of ontologies.”(Blomqvist, 2009)

(Blomqvist *et al.*, 2010) see ODPs as a “modeling solution to solve a recurrent ontology design problem.” According to them ODPs consist of the set of “prototypical ontology entities” and “metadata about its use cases, motivations, …, links to the other patterns etc”.

The purpose of ODPs is to solve the designing problems in ontologies. An ODP can be represented by a single element, a module of an ontology or by the logical structure of the whole ontology. ODPs reduce the need for experience engineers while creating new ontologies. Those with less experience “can apply the well-defined solutions provided in the patterns.”

Ontology design patterns can be also used to automatically construct the ontologies. To do so the ODPs need to be specific enough to be used in automatic ontology construction and generic enough to be used in the construction of several ontologies with the same domain of knowledge (Blomqvist and Sandkuhl, 2005).

There are various types of ODPs that can be divided into six groups: Reasoning, Structural, Content, Lexico-Syntactic, Presentation and Correspondence ODPs. See the graphical representation of the ODP types in Figure 1 (Zagorulko, Borovikova and Zagorulko, 2018).

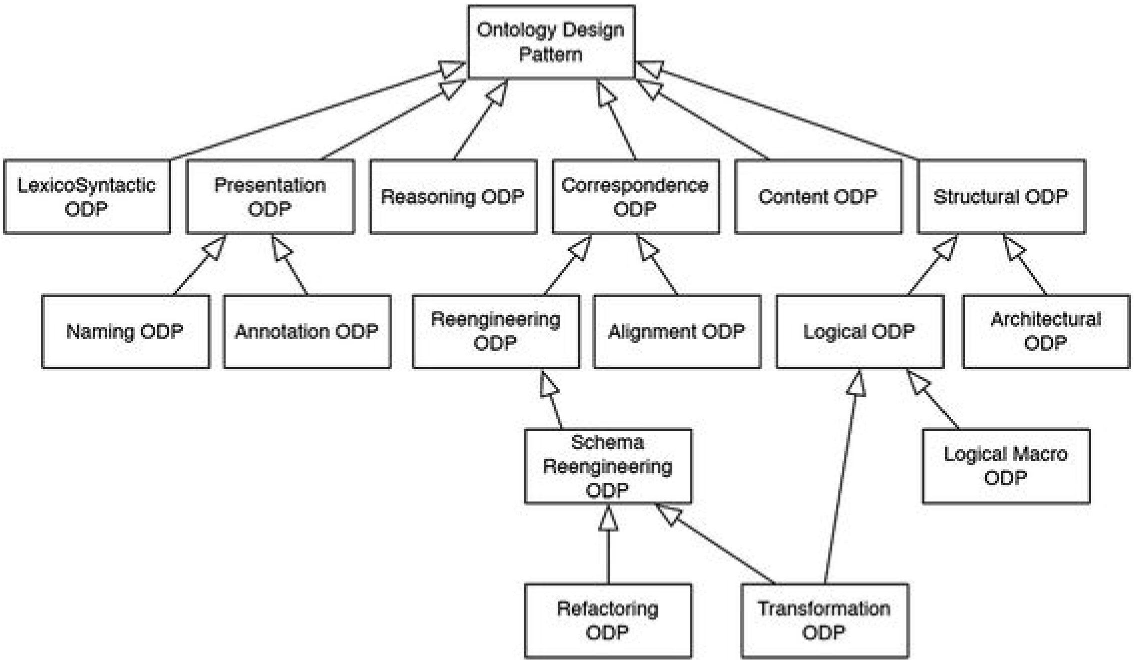


Figure 1. The graphical representation of the ODP types

Lexico-Syntactic ODPs are linguistic structures that contain certain types of words in a defined order and allow deriving conclusions about the meaning they express.

Presentation ODPs are defined by usability and readability of ontologies from a user perspective. It is considered a good practice that is focused on ontology reuse. They consist of Naming and Annotation ODPs. Naming one is focused on the correct way to name files, elements of ontology , for instance classes, properties, relations etc. It is a good practice that standardise an ontology and improves its readability and thus makes it more understandable by humans. Annotation ODPs are focused on annotation properties that improves the readability of an ontology.

Reasoning ODPs are the applications of the Logical ODPs. They are focused on getting certain results depending on what kind of reasoning engine is used. The examples of Reasoning ODPs are classification, subsumption, inheritance, materialization, de-anonymizing, etc. When Reasoning OPs are used in ontology it lets the system know what should be done to that ontology so that evaluations, queries etc can be carried out.

Correspondence ODPs include two types: Reengineering and Alignment (also called Mapping) ODPs.

- **Reengineering ODP** is a solution that allows designers to transform a conceptual model, which is not necessarily of an ontology type, into a new ontology. It takes elements of a source model (for example UML model, thesaurus concept etc) and transforms it into a target model, which is an ontology. Alignment ODPs are used to create semantic associations between two different ontologies without changing the logical types (e.g. owl:Class) of its elements.

- **Content ODPs (CPs)** represent the conceptual design patterns, not the logical ones. That means that if Logical ODPs focus on the ontology design independently from the concept, CPs on the opposite focus on the concept. That is why it is mostly used to represent the specific domain knowledge.

Structural ODPs are divided into Architectural and Logical ODPs.

- **Architectural ODPs** influence the way an ontology is shaped in general. Its main focus is the overall structure of the ontology. It can be internal and external types. Internal is defined by a combination of Logical OPs and used in the design phase of creating an ontology. External is defined by meta-level constructs, for example the modular architecture pattern which consists of a network of ontologies, which are similar to modules that build the ontology.
- **Logical ODPs** consist of logical constructs and solve the problem of expressivity. The signature of Logical OPs is empty, meaning that it does not include any set of classes and the relations between them except for the predefined owl:Thing in OWL. That makes Logical OPs content independent that is it does not depend on the domain context. At the same time they are dependent on the expressivity of logical formalism used for representation.

All in all the two types that are most discussed are Logical and Content ODPs. The logical one focuses on the logical structure, it is focused on expressivity of a language, as well as its common issues and problems.

Content ODPs are often seen as instatiations of logical ODPs. They contain actual classes, properties etc. That is why content ODPs are usually created for the special domain of knowledge.

2.4. ODPS AND BIOMEDICAL ONTOLOGIES

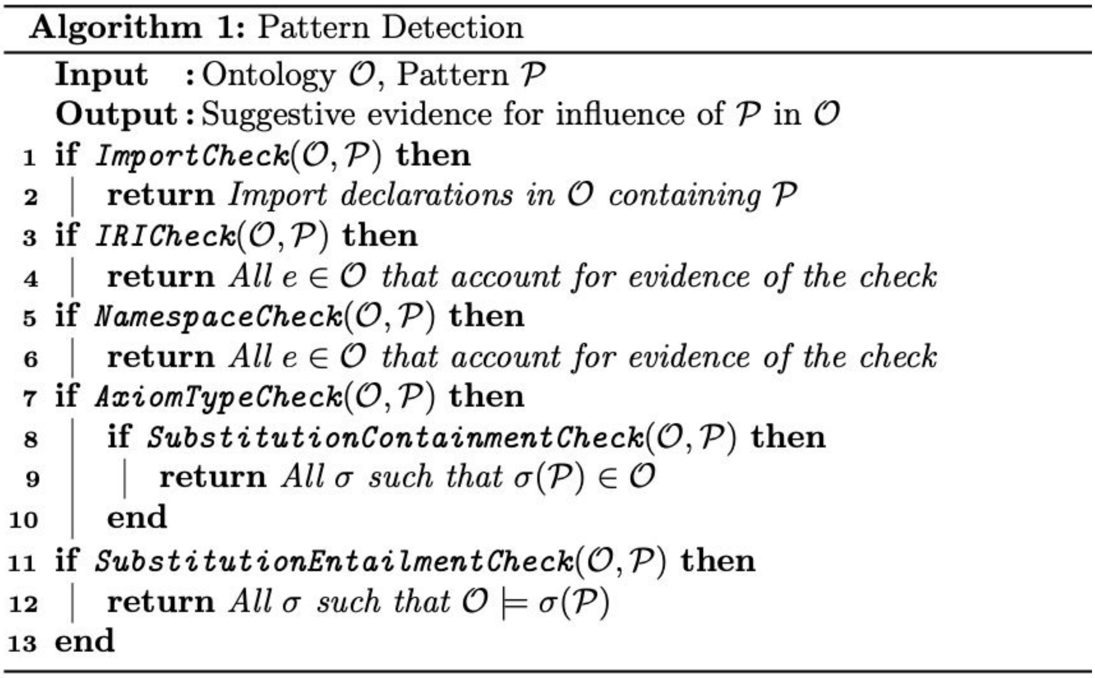
(Gangemi and Presutti, 2009) claim that the most neglected area of ontology design is reusability. This is mostly due to a large set of ontologies and their complexity. They believe that portable and sustainable ontologies (e.g. FOAF and SKOS) are a good choice for a new approach to ontology design. They suggest to create and use small ontologies as building blocks in ontology design. It is called *Content Ontology Design Patterns (CP)* which can be used as a modelling component. An ontology built with CPs has “appropriate dependencies between them, plus the necessary design expansion based on specific needs” in accordance to the authors.

CPs are seen as "special networked ontologies, which cover a set of competency questions, which represent a problem they provide a solution for." (Blomqvist *et al.*, 2010)

Use of ODP in biomedical ontologies was discussed by Christian Kindermann, Bijan Parsia, and Uli Sattler (Kindermann, Parsia and Sattler, 2019). According to them, "there is little empirical work to support the often claimed benefits provided by ODPs." but they try to develop a technique that would allow to "develop algorithmic techniques to automate the identification of a given ODP's influence."

There are many various frameworks for working with patterns in ontologies, but no standard is used while working with them in practise. (Kindermann, Parsia and Sattler, 2019) focus on ODPs that are "captured by a set of axioms or an OWL ontology." They have developed several techniques for reuse. Import containment method is based on ontology import declaration. It can be imported by its name or location. Signature overlap is another technique that is based on copying a given component into an ontology. There are other techniques , for example logical and lexical variation, logical axiom and logical expression agreement.

Kindermann et al. also developed an algorithm based on string comparison of an ontology and a pattern, see Figure 2.



```
Algorithm 1: Pattern Detection
Input  : Ontology $\mathcal{O}$, Pattern $\mathcal{P}$
Output : Suggestive evidence for influence of $\mathcal{P}$ in $\mathcal{O}$
if ImportCheck($\mathcal{O}, \mathcal{P}$) then
    return Import declarations in $\mathcal{O}$ containing $\mathcal{P}$
if IRICheck($\mathcal{O}, \mathcal{P}$) then
    return All $e \in \mathcal{O}$ that account for evidence of the check
if NamespaceCheck($\mathcal{O}, \mathcal{P}$) then
    return All $e \in \mathcal{O}$ that account for evidence of the check
if AxiomTypeCheck($\mathcal{O}, \mathcal{P}$) then
    if SubstitutionContainmentCheck($\mathcal{O}, \mathcal{P}$) then
        return All $\sigma$ such that $\sigma(\mathcal{P}) \in \mathcal{O}$
    end
if SubstitutionEntailmentCheck($\mathcal{O}, \mathcal{P}$) then
    return All $\sigma$ such that $\mathcal{O} \models \sigma(\mathcal{P})$
end
```

Figure 2. Pattern detection algorithm

The result of their research did not give enough evidence of ODPs' influence in biomedical ontologies.

M. Egana developed an Ontology PreProcessor Language - “an axiom-based language for selecting and transforming portions of OWL ontologies, offering a means for applying ODPs.” (Egaña *et al.*, 2008) It was developed to address the issue of manual construction of ODPs in complex ontologies, which is prone for mistakes.

ODPs might not always be practical enough, especially when constructing big and complex ontologies. This is because constructing ODPs for particular modelling tasks require a considerable amount of manual work (Skjæveland *et al.*, 2017). The Reasonable Ontology Templates (OTTR) is a solution for that issue.

2.5. OTTR

As mentioned in Chapter 2.4, ODPs have their limitations, particularly practical implementation of ODPs in OWL ontologies. “...the natural possibilities are either to import its OWL implementation, which includes the whole pattern as-is, or by cloning (parts of) it.”(Lupp, 2019) This solution cannot be called sustainable since it still requires a significant amount of manual work.

“OTTR (Reasonable Ontology Templates) is a language with supporting tools for representing and instantiating RDF graph and OWL ontology modelling patterns. It is designed to improve the efficiency and quality of building, using, and maintaining knowledge bases.” (*Reasonable Ontology Templates (OTTR)*, no date). Because the template in itself is an OWL ontology or RDF graph, it can be debugged, published, instantiated and maintained, all according to W3C standards (Skjæveland *et al.*, 2017). OTTRs allow modeling patterns to be applied to OWL ontology engineering in an efficient way, without unnecessary manual work.

With OTTR, it is possible to:

- build a knowledge base by adding instances to templates;
- get information from the knowledge base by using queries for templates instances;
- present and visualize the knowledge base as a set of instances at different levels of abstraction;
- improve the quality and maintainability of the knowledge base through structural and semantic analysis of the templates which construct the knowledge base;
- specify a template based on other templates, because template definitions can be nested;
- set different types of parameters for the template, , e.g. typed, optional, nin-blank and default;
- to use any RDF resource as an argument, including lists.

The OTTR templates were applied to an existing well known Pizza ontology. The template called NamedPizza based on a pattern, takes 3 arguments: name, optional country of origin and toppings list. These arguments are called instances of the pattern. A template consists of the head or a signature and the body. For example a pizza ontology:

1. NamedPizza(?Pizza : 1 class,?Country : ? individual,?Toppings : + class) ::
2. SubClassOf(?Pizza,:NamedPizza),
3. SubObjectHasValue(?Pizza,:hasCountryOfOrigin,?Country),
4. SubObjectAllValuesFrom(?Pizza,:hasTopping,_:blank),
5. ObjectUnionOf(_:blank,?Toppings),
6. x | SubObjectSomeValuesFrom(?Pizza,:hasTopping,?Toppings)

Line 1 is a head of an ontology, lines 2-6 would be its body.
The design of the pattern can be easily updated without changing the instances.

Using of OTTR have the following benefits:

- better abstractions, which makes the knowledge base easier to understand;
- uniform modelling, which improves the use and maintenance of ontologies;
- separation of design and content;
- open standards support etc.

The goal of OTTR is to enable the use of ODP. OTTR separates between "modelling and populating of an ontology."(Lupp, 2019). That means that ontology engineers do not need to understand every detail of a knowledge domain, they can focus on patterns and relationships instead. At the same time the domain experts do not have to understand the technical details of how the ontology works, but rather focus on the correct concepts and their relations.

As we mentioned before, OTTR solves the problem of extensive manual work while applying ODPs on complex ontologies. The regular OWL ontology is created by "recursively expanding an OTTR by replacing any containing OTTR with the pattern it represents."(Lupp, 2019)

2.6. ONTOLOGY REUSE

(Doran, Tamma and Iannone, 2007) suggest that one of the possible ways to achieve efficient ontology reuse is through a process called ontology module extraction. They built a model and an algorithm of how to extract the necessary modules. The biggest advantage of this method is that it allows module extraction independent of the ontology language
(Bontas E.P. *et al.,* 2005) examined cases concerning ontology reuse in medicine and e-recruitment domains. They distinguish between ontology merging and integration. They also point out that ontology reuse has both costs and benefits. Depending on multiple factors

the process of reusing (merging) the existing ontology can cost a significant amount of time and effort. That is mostly because the existing modules are used in a new context.

They took as an example a case study - the project “A Semantic Web for Pathology” and “usage of ontologies in a retrieval system for image and text data in the medical domain”(Bontas E.P. *et al.,* 2005). More than one hundred knowledge sources were analyzed and selected for further reuse. Among those sources are - SNOMED, Digital Anatomist, UMLS Semantic Network and XML-HL7 “a standard XML-based format for the representation of patient data and patient records” (Bontas E.P. *et al.,* 2005) as well as Immunohistology Guidelines which are used in diagnostic procedures. The outcome of reusing these existing knowledge sources is that the significant part of the pathology domain was actually not covered by the existing ontologies despite their size, for example SNOMED. The authors had to manually merge data from the ontologies as well as build the “pathology” part. The analysis of their work showed that the process of merging the data was extremely costly. Despite the increased interoperability the benefits of reuse were “outweighed by their costs, because of the difficulties related to the evaluation and (technical) management of large scale ontologies and because of the costs of the subsequent refinement phase.”(Bontas E.P. *et al.,* 2005)

Another issue is how to choose ontologies that are to be reused. (Katsumi and Grüninger, 2017) argue that it is a challenge to choose between the different candidate ontologies. They introduce the notion of preference between ontologies which is “based on an assessment of relative accuracy and precision.” The limitations of ontology reuse were also discussed in the paper.

According to (Talebpour, Sykora and Jackson, 2018) the selection for reuse “depends on different social and community related metrics and metadata.” Based on a survey, they determined a set of metrics that ontology engineers pay attention to when choosing an ontology for reuse. The respondents were mainly experienced ontology researchers, professors and engineers who worked within various domains, e.g. biomedical, business, industry, etc. The most important factors to choose the ontologies for reuse were:

- content and coverage (classes, relationships, properties, etc);
- availability of documentation;
- structure (hierarchy);
- domain coverage;
- ontology availability.

(Simperl, 2009) defined the four stage workflow of ontology reuse based on the analysis of several use cases:

- Discovering the ontologies, that is search for the relevant ontological resources;

- Selecting an ontology to be reused. This entails understanding the ontology, checking whether it is relevant for the application and proofreading;
- Customization of relevant ontologies. This process means reengineering and adjusting the ontology to be reused;
- Integration into the application ontology.

Reuse of existing ontologies can be a beneficial project, however due to many factors the process of merging existing ontologies might cost a significant amount of time and effort. The process of reuse consists of multiple steps, which may vary depending on the use case.

2.6.1. THE OBO FOUNDRY AND BFO

The Open Biological and Biomedical Ontology (OBO) Foundry is a collaboration of ontology developers that are devoted to the shared principles. The goal is "to develop a family of interoperable ontologies that are both logically well-formed and scientifically accurate." (*The OBO Foundry*, no date) The principles of the OBO foundry include open use, common syntax and relations, non-overlapping and scoped content.

The Basic Formal Ontology is based on OBO foundry principles. It does not contain any physical, biomedical or other science terms, and thus does not cover any science domains. Other OBO Foundry ontologies, , for example Gene Ontology, Human Disease Ontology, Protein Ontology and others, are built on the basis of BFO ontology.

Several different groups have approached the issue of designing the biomedical ontologies. M. Brochhausen (Brochhausen *et al.*, 2013) constructed an ontology for the biobank-administration domain according to the principles of the OBO Foundry. Their goal was to construct an ontology where it would be possible to query biobank data from different perspectives: administrative, population and specimen. They developed an ontology and called it Ontologized MIABIS (Minimum Information About BIobank Data Sharing). They used Basic Formal Ontology (BFO) to "facilitate reuse and harmonisation across ontologies"(Brochhausen *et al.*, 2013). They believe that is the best way to construct that type of ontology due to the fact that it is easier to reuse, integrate and build further ontologies using the same approach, which is a key point of OBO Foundry principles. Due to that they were able to import several ontologies while creating OMIABIS, e.g. PNO (Proper Name Ontology), OBI (Ontology of Biomedical Investigations) and OMRSE (Ontology of Medically Relevant Social Entities). All of these ontologies are modelled using BFO.

That means, though, that if an ontology is not built according to the principles of the OBO Foundry it makes the integration process more complicated.

Another example of the use of the upper level ontology is CSEO - the Cigarette Smoke Exposure Ontology. According to research “The core of any ontology is a controlled vocabulary that attempts to describe a unified definition for all terms and concepts in a particular subject area.”(Younesi *et al.*, 2014). CSEO is based on the axioms provided by BFO and ExO, which is an exposure ontology. ExO is used in ontologies related to the exposure domain.

CSEO has two different versions. The main one, which is based on BFO, and the second one, the controlled vocabulary version. The main version consists of BFO hierarchy with the inserted ExO hierarchy. They called it a “computer readable” format of CSEO. The second version is called “expert readable” and is meant to be “intuitive and easy to navigate by medical and biological experts”.

2.6.2. OBO FOUNDRY CRITICISM

According to Barry Smith, the goal “is to develop a suite of coherently defined bio-ontological relations that is sufficiently compact to be easily learned and applied, yet sufficiently broad in scope to capture a wide range of the relations currently coded in standard biomedical ontologies.”(Smith *et al.,* 2005). A In his article “Relations in biomedical ontologies” he stated that the problem with OBO ontologies is with relations and distinctions between them. He gives multiple examples of ontologies that are based on the OBO Foundry principles, e.g. Mouse Anatomy ontology, Gene ontology, The Drosophila Anatomy ontology and others. They have similar types of relations , for example is_a, part_of, develops_from etc.

He states “While considerable effort has been invested in the formulation and definition of terms in biomedical ontologies, too little attention has been paid in the ontological literature to the associated relations.”(Smith *et al.,* 2005). He gives OBO’s Cell Ontology as an example. There we can see both *derives_from* and *develops_from* as relations.

There are similar issues in the Gene Ontology which is based on the controlled vocabulary. GO is considered to be a standard in the field of genetics and biology in general. However, GO Consortium admitted that its controlled vocabulary has some “problematic features…, which are destined to raise increasingly serious obstacles to the automatic integration of biomedical information in the future.”(Smith and Kumar, 2004). By those features B. Smith meant mostly the issues of compositionality and syntactic regimentation.

Certain steps have been done to improve GO, which are based on “re-expressing the existing GO schema in a description logic (DL) framework.” That helped to detect some types of errors and problems, which was later implemented in later versions of GO. The problem is that DL is mostly understood only by DL specialists, which means that they would have to take part is creation of ontologies. To solve that issues the more “radical approach” was taken, which went to the level of “re-examining the basic definitions of the relations used in GO and in related ontologies…”(Smith *et al.,* 2005). That is supposed to benefit the biologists by avoiding the issues that are usually dealt with by DL professionals.

2.7. ONTOLOGY MAPPING

Development of ontologies for the same or similar domain raises the need to unify those ontologies to avoid scattering of data.

There are many different existing tools to match ontologies. It is common that various ontologies, even the ones that are based on different knowledge domains, are similar. Currently there are dozens of various tools for ontology matching developed by different universities. For example KitAMO (*KitAMO*, no date) which was created in Linköpings University in Sweden. It is a “tool for evaluating ontology alignment strategies and their combinations”. It is based on “comparative evaluation of the non-interactive alignment components, including alignment algorithms, combination algorithms and filters”.

Another project is from 2014 and is called SOMER (*SOMER: Semantic Ontology Matching using External Resources*, no date) from the University of Lisboa, Portugal. They developed a method that analyzes background knowledge “through evidence and information content provided by unstructured text and annotation corpora”(*SOMER: Semantic Ontology Matching using External Resources*, no date). SOMER was used on real ontologies in the biomedical and geospasial domains, but can as well be used in other areas.

C-OWL (Stuckenschmidt *et al.*, 2004) was used to merge especially medical ontologies. C-OWL is an extension of OWL that allows alignment mapping between different complex ontologies that might be incompatible on the semantic level. The so called bridge rules are used to define relations between different ontologies. These rules state that a concept from an ontology Oi is:

- more general;
- more specific;
- equivalent;
- disjoint;
- overlapping

compared to another ontology Oj.

These rules are represented in Figure 3.

$$i:x \xrightarrow{\sqsubseteq} j:y, \;\; i:x \xrightarrow{\sqsupseteq} j:y, \;\; i:x \xrightarrow{\equiv} j:y,$$

$$i:x \xrightarrow{\perp} j:y, \;\; i:x \xrightarrow{*} j:y,$$

Figure 3. Bridge rules between ontology Oi and Oj

They used GALEN (*The Galen Project | Universiteit Utrecht*, no date), Tambis (*The TAMBIS Project*, no date) and Unified Medical Language System (UMLS) ontologies for mapping using C-OWL. The goal was to establish a connection between the ontologies, which supplement each other, as visualized in Figure 4.

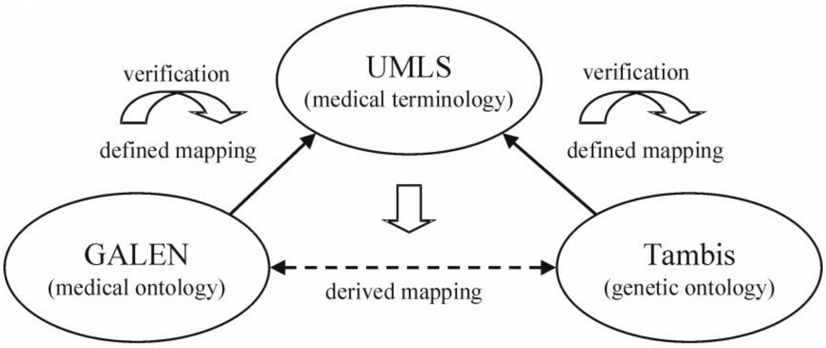


Figure 4. C-OWL ontologies mapping

They conclude that C-OWL is a suitable tool for merging ontologies. it “provides powerful reasoning support for the verification and derivation of mappings and even supports the process of merging terminologies based on existing mappings.”(Stuckenschmidt *et al.*, 2004.)

3. USE CASE

To understand the scope of this master thesis it is important to present the use case, which is the use of health registries in Norway. We discuss the issue of standardizing and how the registries are used in real life.

In Norway there are 11 central medical registries regulated by Folkeinstitutet (*Helseregistre og andre registre*, no date):

- Pregnancy Termination;
- Biobank;
- Cause of Death;
- Cardiovascular Disease;
- Medical Birth;
- Surveillance System for Communicable Diseases (MSIS);
- Immunisation (SYSVAK);
- Surveillance System for Use of Antibiotics and Healthcare Associated Infections (NOIS);
- Surveillance System for Antimicrobial Drug Resistance (NORM);
- Surveillance System for Virus Resistance (RAVN);
- Prescription Database (NORPD).

Central or statutory registries are used for statistics, health analysis, research and planning in health sector all over Norway.

In addition there are 53 national medical registries, also referred to as the medical quality registries. A registry contains information about the whole treatment period, for example diagnostics, treatment itself and its follow-up, as well as a result of a treatment. The main purpose of the medical registries is to improve the quality of medical services for a patient and to avoid unnecessary variation in healthcare and treatment quality. Some of the national registries were created on the basis of the central registries.. The registries are managed by various units or organizations in the hospitals. The official status of a national medical registry is approved by the Department of Health and are created on the basis of the Health Registry and Regulations. National health and quality registries are important tools in the work of public health monitoring, statistics, health analysis, quality assurance, quality improvement and research. In addition to national registries there are numerous local and regional registries which are not included in this research. (*Medisinske Kvalitetsregistre*, no date)

The examples of national medical quality registries are:

- Gastronet (*Gastronet*, no date);
- Child Cancer (*Nasjonalt kvalitetsregister for barnekreft*, no date);
- Stroke (*Norsk hjerneslagregister*, no date) see Figure 5

Tilstand før det aktuelle hjerneslaget

Boligforhold
☐ 1 Egen bolig uten hjemmesykepleie/hjemmehjelp
☐ 2 Egen bolig med hjemmesykepleie/hjemmehjelp
☐ 3 Omsorgsbolig med døgnkontinuerlige tjenester
☐ 4 Sykehjem, både korttids- og langtidsopphold
☐ 9 Ukjent

Sivilstatus
☐ 1 Gift/samboende
☐ 2 Enke/enkemann
☐ 3 Enslig
☐ 9 Ukjent

Bosituasjon
☐ 1 Alene
☐ 2 Sammen med noen
☐ 9 Ukjent

Forflytning
☐ 1 Alene – inne og ute
☐ 2 Alene inne
☐ 3 Med hjelp (av andre)
☐ 9 Ukjent

Toalettbesøk
☐ 1 Alene
☐ 2 Med hjelp
☐ 9 Ukjent

Påkledning
☐ 1 Alene
☐ 2 Med hjelp
☐ 9 Ukjent

Modified Rankin Scale
☐ 0 Ingen symptomer
☐ 1 Ingen betydelig funksjonssvikt
☐ 2 Lett funksjonssvikt
☐ 3 Moderat funksjonssvikt
☐ 4 Alvorlig funksjonssvikt
☐ 5 Svært alvorlig funksjonssvikt
☐ Ikke utført

Figure 5. Emergency form for the Stroke Registry

Figure 6 shows how the variables from the Stroke registry are represented on the health registries variables website (*Nasjonale registervariabler - metadata*, no date). The variables presented in the report are based on data from 2017, while registries are usually updated every year. That means that some of the variables might be outdated, which might cause some difficulties while searching for variables during the research.

Er du yrkesaktiv nå?	Bo, arbeid, funksjon	Inn	Valg	1 / 1	0	1 Ja 2 Nei 9 Ukjent
Var du yrkesaktiv da du fikk hjerneslaget?	Bo, arbeid, funksjon	Inn	Valg	1 / 1	0	1 Ja 2 Nei 9 Ukjent
Sivilstatus	Bo, arbeid, funksjon	Inn	Valg	2 / 3	1	1 Gift/samboende 2 Enke/ enkemann 3 Enslig 9 Ukjent
Boligforhold	Bo, arbeid, funksjon	Inn	Valg	2 / 3	1	1 Egen bolig uten hjemmesykepleie/hjemmehjelp 2 Egen bolig med hjemmesykepleie/hjemmehjelp 3 Omsorgsbolig med døgnkontinuerlige tjenester 4 Sykehjem 9 Ukjent
Bosituasjon	Bo, arbeid, funksjon	Inn	Valg	1 / 2	1	1 Pasienten bodde alene 2 Pasienten bodde sammen med noen 3 Pasienten bodde i institusjon/sykehjem 9 Ukjent

Figure 6. Representation of variables from the Stroke Registry

Both central and national registries were developed over time, the first ones are dated back to the 1950s. By that time there were no national or international standards, terminology or coding system, as well as no requirements for the use of variables. As mentioned above the national quality registries are managed by various organisations which means that the variables of the registries do not have a common standard format. Data from registries is collected through different systems and often via manual filling out paper forms. It is common that data from the registries is overlapping with information registered in Electronic Health Records and has to be registered multiple times in various local systems. The issue of improving interoperability between registries and EHR is addressed in chapter 3.3.1.

Different variables can have the same meaning in different registries which makes it particularly difficult to search through them and analyze registry data afterwards. It is common that a patient has different conditions, registered by different doctors and hospitals. Thus information about that patient is stored in several registries. Another issue is that it is hard to know which data can be merged or which variables are related. It is very common that data needs to be transferred from one hospital system to another, but due to non centralised registry system, that process takes more effort and time than needed. All that makes it particularly difficult to maintain the whole system. Standardization might be one of the solutions for the problem.

The challenge of standardizing registries data is important on the national level. One of the main current goals in e-Health department is to connect data from registries with data from Electronic Health Records. The implementation of these systems would be a huge improvement in patient data accuracy which will result in better quality of health services. However, to start implementing those systems the registries and their variables need to be standardized.

The Ministry of Health and Care Services has initiated work to develop a common regulation for medical quality registries. The objective is that a regulation will be able to replace the licenses these registrations have today, which can provide a more predictable framework.

3.1. HEALTH DATA PROGRAM

The quality of health registries is essential for health services and its improvement. The medical quality registries are designed to document treatment results from the whole country and give basis for quality improvement and research. The variety of variables makes it harder for researchers to find the information they require. A great deal of work has been done already in mapping of variables in national health registries (*KARTLEGGING AV VARIABLER I NASJONALE HELSEREGISTRE*, no date). There is much interest in creating an ontology with variables from medical registries.

Health Data Program (*Helsedataprogrammet*, no date) was established in march 2017 by Department of E-health. The project is part of the strategy for modernization and coordination of national health registry data. The Health Data Program is designed for better use, quality, easier reporting and safer handling of health data. The program consists of a “Harmonising project” which goal is to give a better access to health data to those who are interested, for instance researchers and those who work in the health sector (*Prosjekt Harmonisering i Helsedataprogrammet*, no date). Another goal is to improve the quality of registries data and its management. The way to do it would be to standardize the creation and use of registries. The project is divided into three sub projects: Integration and interaction; Metadata, coding and terminology and Architecture management (*KARTLEGGING AV VARIABLER I NASJONALE HELSEREGISTRE*, no date)

- Integration and interaction is about sharing data between registries and from registries to the health platform and researchers. It is focused on establishing the common terminology for registries.
- Metadata, coding and terminology subproject involves improving and designing common terminology. This will improve the analysis process and its quality since it would be easier for different actors to get access to data.
- Architecture management is focused on establishing various architecture mechanisms for health registries.

3.2. STANDARDIZATION ISSUE

The standardization process might be time consuming. Certain types of variables e.g., demographics are often collected in all or most registries. Other data, for example data related to cardiology can be collected not only in cardiology registries, but also in other registries connected to cardiac diseases. This might be an issue for registries focused on surgery, endocrinology and other clinical domains, and despite the fact that data is pretty common, most registries are not using a common format. For example different formats for one of the most common variable - date of birth, which can be year, month and year, the full date etc. There is currently no standard that defines how common data elements should be defined and captured in registries.

One of the examples is a variable “smoking status” (‘RØYKESTATUS’)) which is used in 28 registry forms in its 9 different variations. In addition to that there are variables , for instance “amount of smoking years” (in 2 registries), “amount of cigarettes per day” (in 5 registries), “amount of months without smoking” (in 4 registries). One of the variations of the “smoking status” is a “Do you smoke daily?” variable.

Other examples of the variables with many variations are: “civil status”, “living conditions”, “occupation”, “education”, “gender” etc.

There is a significant amount of work planned for registries in the future (*Nasjonal strategi: Slik skal vi jobbe med kvalitetsregistrene fremover*, no date). The national strategy includes 11 goals to be achieved by the end of 2020. One of the goals is that 85% of national medical quality registries represent the results reported by the patients themselves. Standardization process is crucial for the data to be presented accurately.

3.3. THE USE OF MEDICAL REGISTRIES

Medical quality records have been created to document treatment results, provide a basis for quality improvement and research. Information from the registries is used mainly for statististics. The example of the main findings from the Child Cancer register (*Nasjonalt kvalitetsregister for barnekreft*, no date) for the past years are:

- The amount of cancer cases among children is stable, not increasing;
- Children who get cancer in Norway get it regardless of where they live;
- The amount of children who get radiation therapy during the period of 2010-2017 is 2 times smaller than during 2002-2009.

When analyzing registry data different variables can be taken into consideration depending on the purpose of research. For example, (Javor *et al.*, 2019) analyzed data from the Stroke Unit Registry in Austria to find out the differences between stroke risk factors and treatment variables between rural and urban regions in Austria. Using the postal code of each patient they collected geographical information. Other than geographical data they used such variables as gender, risk factors, several pre-hospital variables, DNT (door to needle time), OTT (onset to treatment time) and finally outcome rates, for example mortality and rehabilitation variables.

In another example, the study of (Torgersen, 2003) used the data from Stroke registries in Norway to determine the validity of stroke identification by dispatch guidelines. In this research they compared patients who were suspected to have a stroke vs patients who were diagnosed with stroke. After analyzing variables related to symptoms and diagnosis they concluded that there was a quite high percentage of stroke prediction. Research like this could show us how precise the diagnosis are and what kind of symptoms are most common for strokes.

3.3.1. IMPROVING INTEROPERABILITY BETWEEN REGISTRIES AND EHR

Improving interoperability between health registries and Electronic Health Records (EHR) is one of the goals in medical informatics. EHR contains data about each and every patient in the country. By implementing data from registries this system could be constantly updated.

Seth Blumenthal, based in Chicago, works on the same problem the Norwegian system has - lack of standardization. In the beginning of his paper (Blumenthal, 2018) he explains why the standardization of the whole system is it a challenge. He complains about the lack of standards, while this is necessary for a well functioning healthcare system. “Standards that allow meaningful semantic and process interoperability either do not yet exist or are not yet feasible to apply and thus have not been widely adopted in health IT.”(Blumenthal, 2018). At the same time Blumenthal agrees that implementation of both systems is a complicated task mostly due to variety of data. He states that “...it is difficult to aggregate data across multiple registries to come up with uniform data, …, and health outcomes across a wide variety of patient populations, diseases and conditions and the clinical specialties that provide care for those patients.”(Blumenthal, 2018).

(Blumenthal, 2018) points out the following priority areas:

- Develop common data element standards for data that are used in many registries e.g., demographics, conditions, etc;
- Develop standardized information models that support needs common across many registries e.g., QCDR measures, measure sets needing harmonization, cross-cutting measures;
- Develop nationally-accepted specialty-specific data standards e.g., for pathology specimen collection, and make them generally available;
- Improve patient matching; particularly useful in specialties that have a significant overlap i.e., anesthesia;
- Identify specific instances of concepts in registries e.g., a system to standardize the way anything is tagged or identified in registries;
- Develop technical interoperability standards or guidance for registries.

After conducting interviews with informaticists, they identified a number of steps that should be done. “The first step is to identify the clinical or operational scenario that may benefit from improved interoperability. Next step would be to identify an information model “that documents and abstracts the data and information needs at a level that allows the needed information to be stored and communicated.”(Blumenthal, 2018). Once these models are created an implementation guidance can be created.

3.3.2. PICO MODEL

PICO (*PICO | helsebiblioteket*, no date) is often used in health care “as a strategy for formulating questions and search strategies and for characterizing clinical studies or meta-analyses”(*PICO ontology*, no date), because it is a complicated and time-consuming process to collect and analyze all the data from the various studies about a particular clinical problem. (Mavergames, Oliver and Becker 2013)

PICO is an ontology that structures information and allows to ask the correct and focused questions. PICO stands for:

- **Population/patient or problem** - Which population group is it about?;
- **Intervention** - What is it you are interested in with that population group? Are there any measures taken , e.g. intervention or exposure?
- **Comparison** - Which measures would you like to compare? Does intervention has an alternative?
- **Outcome** - What kind of outcome are you interested in? Which ones are relevant?

PICO works like a filter which allows to classify and organise big amount of data. The PICO Ecosystem was created by Cochrane (*Cochrane*, no date) to describe the process of “filtering”. It is an iterative process with the main objectives being quality and standardisation of data. Figure 7 gives a graphical representation of the PICO ecosystem.

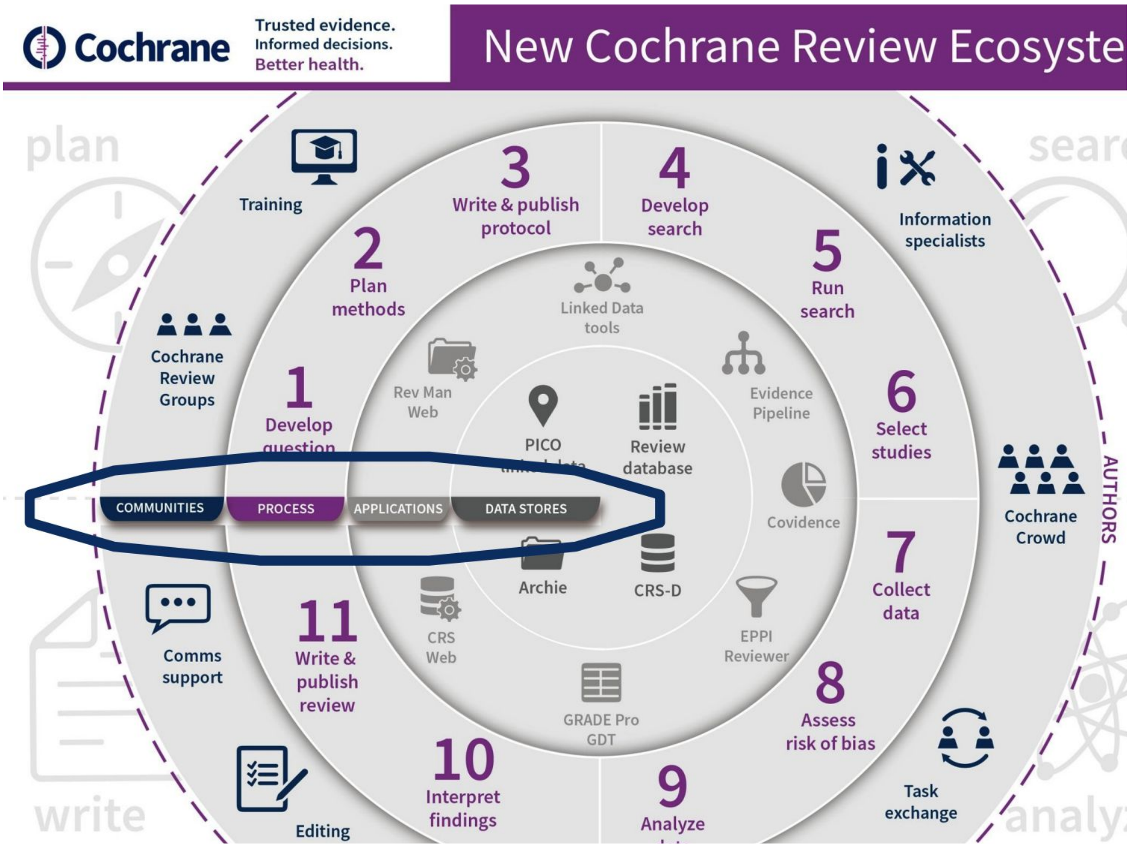


Figure 7. The graphical representation of the PICO Ecosystem

Analyzing registry data with the help of PICO would structure the data accordingly. This would simplify the process of analyzing the data for future analysis and statistics by tagging the data correctly.

4. METHODS

We started with collecting competency questions for the future ontology. That ontology was evaluated with the queries based on the collected competency questions. Afterwards, we modelled a Population ontology, based on PICO, with concepts that are based on demographic and addiction registry variables. Next, we manually analyzed ontologies that could be reused. They were evaluated by how often they had been reused before, amount of classes and properties and how those classes can be related to the registry variables. Then we explored how the OTTR templates for registry variables could be created. We analyzed what characterizes the variables and what were their components, that could be used for templates. At last, we explored how the templates could be used in an application.

4.1. COMPETENCY QUESTIONS

Competency questions (CQs) is an important part of an ontology development. They are crucial for understanding how the future ontology will be used and thus how it should be modelled. These are the questions that help to define the scope and the content of an ontology. CQs are supposed to represent all types of relations in an ontology. For example, in our case, if we take a group of demographic variables, we need to be able to know the age, family status, employment status etc. Thus, due to the abundance of variables, the amount of competency questions can be unlimited. We have invited a healthcare professional, who has experience with the registry variables, to help us with writing competency questions. We took into consideration how the registries are used in research (see Chapter 3.3 The Use of Medical Registries), and designed the questions accordingly. The CQs were grouped into questions about individuals (population), the variables themselves, the combination of individuals and variables, free text information and questions related to outside modeled information. They were used to evaluate the future ontology. Some of the CQs were translated to DL queries for the evaluation.

4.2. REUSE OF PICO ONTOLOGY

We modeled the existing registry variables to build the evidence base, using the PICO ontology. The Population part of the ontology was chosen to be reused in our ontology. The structure of the Population part of the PICO is presented in Figure 8 (*PICO ontology*, no date).

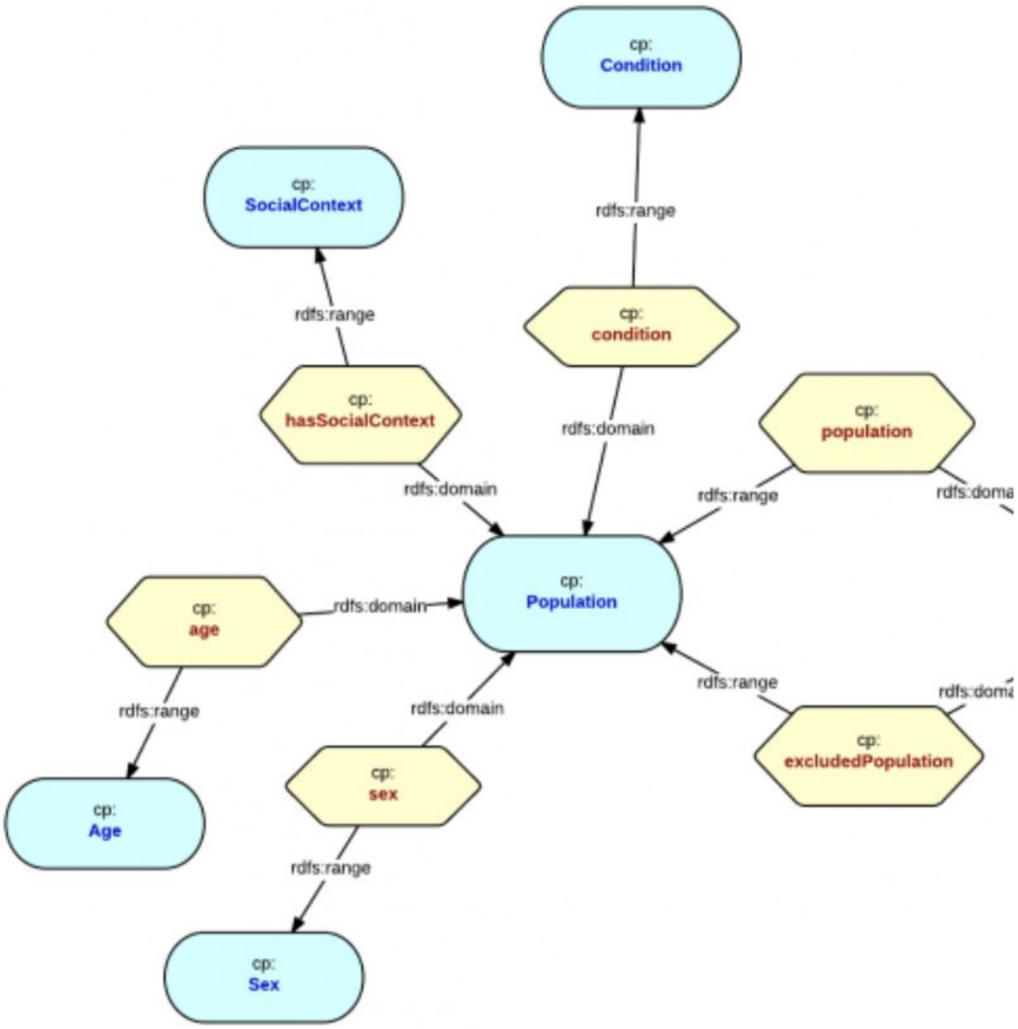


Figure 8. The Population part of the PICO Ontology

All the classes that are in the Population module if the PICO ontology were reused in our ontology, that is Condition, SocialContext, Age and Sex.
The registry variables were chosen to be included in our ontology on the basis of the competency questions and relation to a population group. The following groups of variables were chosen to be modeled in the health registry ontology:

- Condition of a patient (Cancer, Stroke, Osteoporosis)
- Social context and addiction (Alcohol, Drugs, Smoking)
- Sex (Male, Female)
- Age (Age when hospitalised, age in years and days)

Several other classes were added to the Population ontology, e.g. Medication Exposure, Language and Employment.

4.3. REUSE OF EXISTING BIOMEDICAL ONTOLOGIES

In addition to reusing the PICO design pattern, we analyzed several biomedical ontologies to be reused in ours. We used the theoretical background described in Chapter 2.6, especially the works of Talebpour et al., Doran et al. and Simperl. Their research results were used to create the steps for the ontology reuse process.
The process of reuse of other biomedical ontologies consisted of the following steps:

1. Finding ontologies

We searched for ontological resources that were relevant for the health registry variables ontology using Bioportal. It is a known repository containing biomedical ontologies. The ontologies to be reused were searched by reviewing the most visited ontologies and by inputting the concept that needs to be reused in our ontology. The choice of classes depended on the concepts that were needed for the Population ontology.

2. Selecting ontologies

This step included activities such as understanding the ontology and checking whether it was relevant for our purpose. We applied evaluation methods suggested by Doran et al. and Talebpour et al. We manually analyzed ontologies applying the following criteria:

- popularity, if the ontology is popular and has been used, it is likely to be correct (Doran, Tamma and Iannone, 2007);
- content and structure;
- scope, that is the coverage of the ontology;
- purpose of the ontology itself, whether is fits our purpose;
- relevant classes and their characteristics, , e.g. amount of children;
- class mappings;
- how the ontology was created;
- how the ontology was evaluated.

3. Integration of the chosen ontology

After the ontology was chosen, we manually extracted and reused the relevant module in our ontology.

4. Evaluation of the result ontology

Our ontology was evaluated with competency questions, which were collected before. We created several individuals and converted questions into DL queries to evaluate the ontology in Protege.

4.4. DESIGNING THE OTTR TEMPLATES

As discussed in Chapter 2.4 and 2.5, the OTTR templates are suited best for practical application in a real ontology and for specific problem. In this thesis we discuss the topic of registry variables standardization, thus the templates should be chosen accordingly. To design the OTTR templates the following steps have been made:

1. Analysis of the health registry variables

To create the templates that would reflect the variables accurately, they needed to be analyzed. The hreg.no website contains an overview of all the variables grouped by category. The following information about the variables were used in the templates:

- variable name;
- variable form, for example text;
- values type;
- and SNOMED CT mapping.

2. Designing the OTTR templates

The OTTR templates were designed in such a way that they are general enough to be an actual template, at the same time they have to be precise enough to be used in the registry variables domain. We consulted with the library of the OTTR templates, which contains common patterns. First, we created the general template for all the variables. Then we continued with creating template according to the variables type.

3. Potential use of OTTR templates

This step included designing the registry builder tool that would use OTTR templates to create the registry ontologies based on input variables.

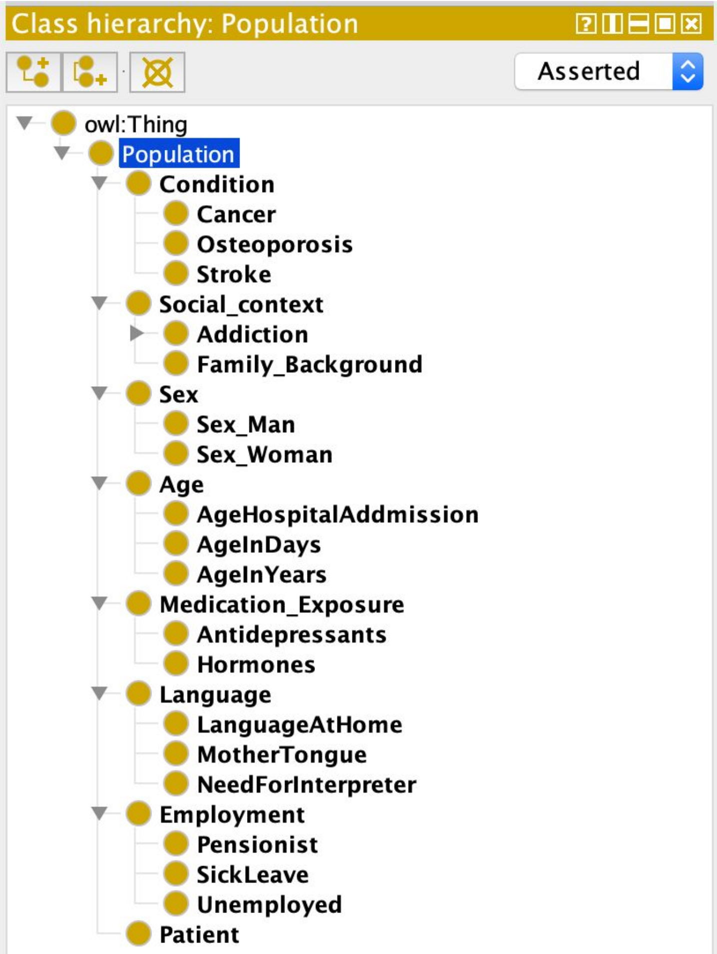
5. RESULTS

In this chapter we present the following results:

- reuse of the Population part of the PICO ontology;
- reuse of the CSEO smoking module;
- evaluation of the created ontology, as a combination of the Population pattern and the Smoking module of the CSEO ontology;
- designing the OTTR templates.

5.1. REUSE OF PICO DESIGN PATTERN

We built a population ontology based on PICO ontology. The Population structure was reused in our ontology. We focused on demographic, addiction and variables related to diseases, which fits into the Population design. We also added variables related to employment, language and medical exposure. See the outline of the created ontology in Figure 9.



- owl:Thing
  - Population
    - Condition
      - Cancer
      - Osteoporosis
      - Stroke
    - Social_context
      - Addiction
      - Family_Background
    - Sex
      - Sex_Man
      - Sex_Woman
    - Age
      - AgeHospitalAddmission
      - AgeInDays
      - AgeInYears
    - Medication_Exposure
      - Antidepressants
      - Hormones
    - Language
      - LanguageAtHome
      - MotherTongue
      - NeedForInterpreter
    - Employment
      - Pensionist
      - SickLeave
      - Unemployed
    - Patient

Figure 9. The outline of the created ontology below

5.2. REUSE OF EXISTING BIOMEDICAL ONTOLOGIES

We have used BioPortal (*Welcome to the NCBO BioPortal | NCBO BioPortal*, no date) to manually examine various ontologies and evaluate reuse of specific classes in our ontology.

The following ontologies were analyzed:
- FHHO (*Family Health History Ontology | NCBO BioPortal*, no date);
- RCD. (*Read Codes, Clinical Terms Version 3 (CTV3) | NCBO BioPortal*, no date)
- PCAO (*Prostate Cancer Life Style Ontology | NCBO BioPortal*, no date);
- CSEO (*Cigarette Smoke Exposure Ontology | NCBO BioPortal*, no date).

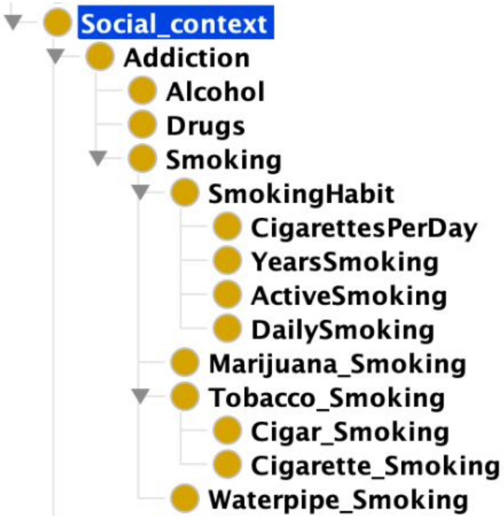
This choice of ontologies was based on different types of registry variables, that are used in our Population based ontology. FHHO contains numerous concepts of family health history. RCD includes concepts connected to clinical findings. PCAO and CSEO contain classes related to lifestyle variables,. See Table 1 for a summary of the comparison.

Ontology name	**FHHO**	**RCD**	**PCAO**	**CSEO**
No. of classes	238	140,065	636	20,085
Amount of properties	433	3	0	91
Class mapping	2	9	1	31
Potential reusable module	Family relations	Clinical findings	Habits and Behaviors	Smoking
Class children	11 children: Adopted_Kin_Relation, Blood_Relation, Female_Relation, Fictive_Kin_Relation, Foster_Kin_Relation, Male_Relation, Organ_Donor, Organ_Recipient, Parent_Relation, Partner_Relation, Step_Kin_Relation.	3 children: Disorders, History and Observations, Morphology findings	8 children: Alcohol, Coffee, Diet style, Leisure time, Sex Behavior, Sleep, Tea or polyphenols, Tobacco smoke	3 children: Marijuana Smoking, Tobacco Smoking, Waterpipe Smoking

Reference	(Saranto, 2009)	(Sidhu, 2009)	(Min *et al.*, 2009)	(Younesi *et al.*, 2014)
Project name	Formalizing Nursing Knowledge: From Theories and Models to Ontologies	Clinical Terms Version 3 (The Read Codes)	Integration of Prostate Cancer Clinical Data Using an Ontology	CSEO – the Cigarette Smoke Exposure Ontology
Purpose	Represent family history in an inclusive manner consistent with nursing conceptualisation	Provide terminology for describing relevant clinical summaries and administrative data for general practice	Provides a description of the concepts and relationships for prostate cancer, Development of the Prostate Cancer Information System (PCIS)	Support annotation of experimental data sets and exposure path to outcomes
Re-use Method	OWL using Protege, iterative process of creating drafts and their evaluation, SNOMED CT as a reference terminology	N/A	OWL-DL using Protege, merging concepts from NCI thesaurus and FMA, Concepts connected by is-a relationship that form DAG structure.	Built on the basis of BFO and ExO superclasses are used as root concepts for CSEO, Protege 4.2, combination of bottom-up and top-down methods.
Evaluation	Completeness of the domain of family health history and ontology's utility for nursing	N/A	N/A	Three-dimensional - structural measure, functional measure and usability profile

Table 1. Comparison of different ontologies

The Smoking module from the CSEO was chosen to be reused in our ontology. We also added the following registry variables related to smoking habits: *aktiv røyker, røyker du daglig?, antall år røykt and antall sigaretter per dag* (active smoking, do you smoke daily?, amount of smoking years, amount of cigarettes per day), as shown in Figure 10.

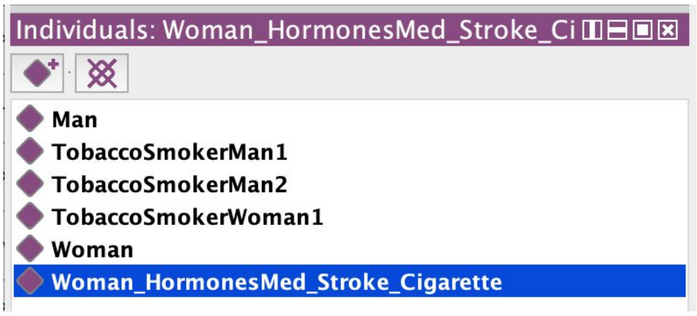


- Social_context
  - Addiction
    - Alcohol
    - Drugs
    - Smoking
      - SmokingHabit
        - CigarettesPerDay
        - YearsSmoking
        - ActiveSmoking
        - DailySmoking
      - Marijuana_Smoking
      - Tobacco_Smoking
        - Cigar_Smoking
        - Cigarette_Smoking
      - Waterpipe_Smoking

Figure 10. An example of reuse of CSEO ontology

5.2.1. EVALUATION USING COMPETENCY QUESTIONS

Several individuals were created to have the data for evaluation, as presented in Figure 11.



Individuals: Woman_HormonesMed_Stroke_Ci

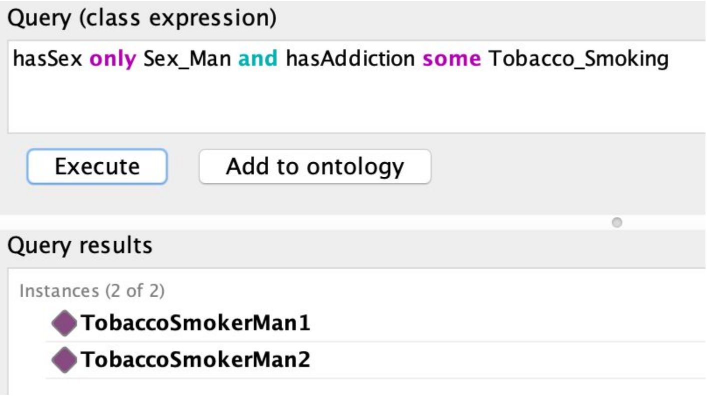
- Man
- TobaccoSmokerMan1
- TobaccoSmokerMan2
- TobaccoSmokerWoman1
- Woman
- Woman_HormonesMed_Stroke_Cigarette

Figure 11. Individuals created in the evaluation

We have divided the questions into categories: Individual (the actual data about the patient), queries about the variables themselves, queries that are both variables and individual data, the outside queries that ask the question about the one who fills out the form. These questions were transformed into queries to extract data.

To evaluate the reuse of smoking module, we used several competency questions that are focused on smoking. See a full list of competency questions in Appendix A.

1. *All individuals who are men and have a tobacco smoking addiction (see Figure 12).*

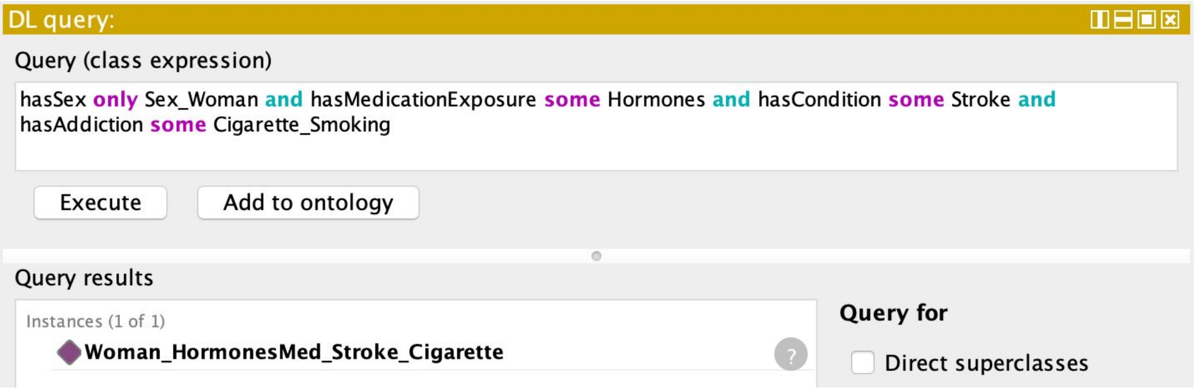


```
hasSex only Sex_Man and hasAddiction some Tobacco_Smoking
```

Instances (2 of 2)
- **TobaccoSmokerMan1**
- **TobaccoSmokerMan2**

Figure 12. Query 1 with results

2. *All individuals who are women on hormonal medication, who had a stroke and smoke cigarettes (see Figure 13).*

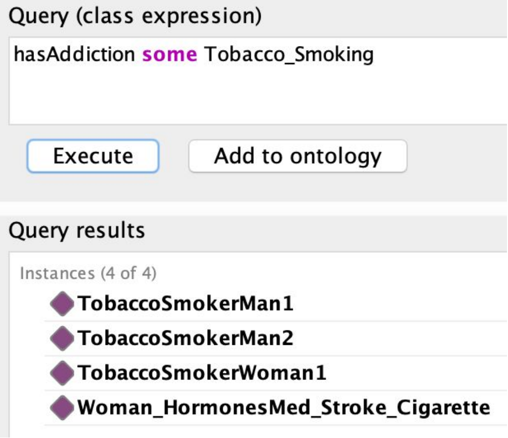


```
hasSex only Sex_Woman and hasMedicationExposure some Hormones and hasCondition some Stroke and
hasAddiction some Cigarette_Smoking
```

Instances (1 of 1)
- **Woman_HormonesMed_Stroke_Cigarette**

Figure 13. Query 2 with results

3. *All individuals who are smoking tobacco (see Figure 14).*



```
hasAddiction some Tobacco_Smoking
```

Instances (4 of 4)
- **TobaccoSmokerMan1**
- **TobaccoSmokerMan2**
- **TobaccoSmokerWoman1**
- **Woman_HormonesMed_Stroke_Cigarette**

Figure 14. Query 3 with results

The evaluation showed that the ontology correctly depicts the information from the registries and the reused smoking module is successfully merged into our ontology and can interact with other classes.

5.3. OTTR TEMPLATES

We started with creating the general template for Health Registry Variables (HRVs).

> *hrv-template:HealthRegistryVariable[owl:NamedIndividual ?name, ? xsd:string ?formText, ? List<owl:Class> ?Parents, ? List<owl:Class> ?Children, ? List<owl:Class> ?RelatedClasses, owl:Class ?HRVType]::*

Where ?name is the name of the variable, ?formText is a type of question, ?Parents is a list a classes that represents the parent classes of the variable. ?Children is a list of classes that on the opposite represents the children classes of the variable. ?RelatedClasses is a list of related classes in other ontologies. ?HRVType is a type of a variable. It can be Categorical, Continuous and Free Text, depending on the values that the variable accepts.
For example, the smoking variable “røykestatus” is a categorical type. Depending on the registry røykestatus has various categorical values, for instance:

- aldri røykt daglig (never smoked daily);
- dagligrøyker (daily smoker);
- eks-dagligrøyker (ex daily smoker);
- røykestatus ukjent (smoking status unknown).

Mutually exclusive values mean that it is possible to choose only one option. A categorical type means that it would be possible to choose several.
Below is a body of the template.

> We say that the variable has type of HRV class
> *ottr:Triple(?name, rdf:type, hrv-class:HRV),*
>
> Name has form text ?formText.
> *ottr:Triple(?name, hrv-prop:hasFormText, ?formText),*

Name is a subclass of all elements of ?Parents.

> *cross | o-owl-ax:SubClassOf(?name, ++?Parents),*

Name is a superclass of all elements in ?Children.

> cross | o-owl-ax:SubClassOf(++?Children, ?name),

Name is in some way related to all elements of ?RelatedClasses. We have defined the property and called it hrv-prop:relatedTo.

cross | ottr:Triple(?name, hrv-prop:relatedTo, ++?RelatedClasses),

Name is a registry variable of type ?HRVType (categorical, continuous, freetext)

ottr:Triple(?name, rdf:type, ?HRVType).}.

After designing the most general template, we started working on the more specific ones. In categorical templates we needed to show that the variable accepts the list of values, which are not mutually exclusive.

hrv-template:CategoricalHRV[owl:NamedIndividual ?name, List<xsd:string> ?vals, List<xsd:string> ?SNOMEDCodes, ? xsd:string ?formText, ? List<owl:Class> ?Parents, ? List<owl:Class> ?Children, ? List<owl:Class> ?RelatedClasses] :: {

Here we call the previously created template and add the information about the type of the variable.

hrv-template:HealthRegistryVariable(?name, ?formText, ?Parents, ?Children, ?RelatedClasses, hrv-class:CategoricalHRV),

We also need to state that name may only have values from ?vals, which is a list of strings. Since ?name and ?vals are of different types (class and list), we cannot relate those two. To solve this problem we used the blank node _:u. After that we use the zipMin function to match the data from values and the SNOMED codes.

o-owl-ax:SubObjectAllValuesFrom(?name, hrv-prop:hasValue, _:u),
cross | ottr:Triple(++?vals, rdf:type, _:u),
zipMin | ottr:Triple(++?vals, hrv-prop:hasCode, ++?SNOMEDCodes).}.

After that we created the templates for Mutually Exclusive, Continuous and Free Text HRV templates. See Appendix B for the rest of created OTTR templates.

6. DISCUSSION

The objective of the research was to explore the possibilities of the use of existing ontologies and the OTTR templates within Health Registry variables. In order to do that, we created an ontology based on reused PICO ontology and CSEO ontology smoking module, and created the OTTR templates for the registry variables. In this chapter we discuss the results and present the way the OTTR templates can be used in a registry builder application.

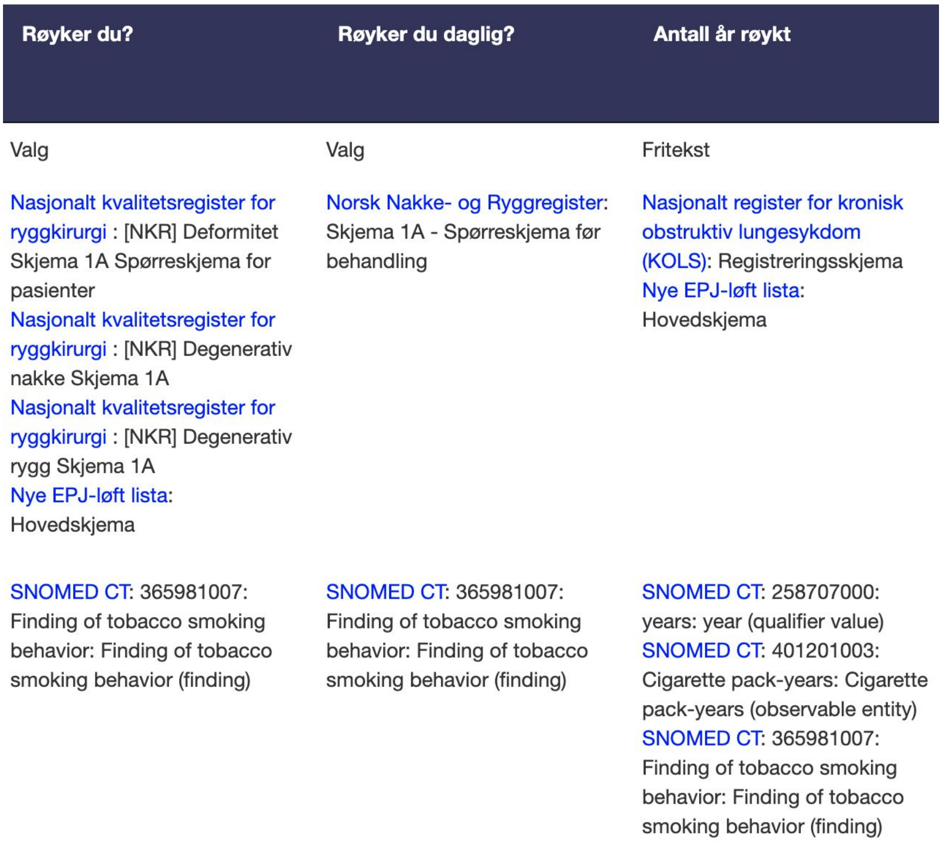
6.1. REUSE OF PICO ONTOLOGY

The Population ontology was created based on the PICO ontology. In chapter 5.1 we presented the results of reusing the PICO-Population design pattern. The Population part was chosen because we wanted to design variables that do not require a medical background. That is why we chose the variables that are related to demographic data, for instance age, sex, employment etc., lifestyle and addiction, and a few condition variables.

The results indicate that reusing PICO ontology pattern suits our purpose and the registry variables, that we used can be modeled according to the PICO design. At the same time, there are certain limitations. The ontology that was created is based on Population part of the PICO ontology, not including Intervention, Comparison or Outcome. Designing four complete ontologies would go beyond the scope of this master thesis. Nevertheless, even reusing the Population part of PICO was important for our research as we found out that PICO's design can be used to model all the registry variables. Applying the whole PICO to all the variables would tag the registry data, which would help filtering the data for statistics.

6.2. REUSE OF CSEO ONTOLOGY

Most of the registry variables are mapped to SNOMED CT or Volven codes, which is very valuable, but that mapping is not perfect. For example, the smoking status in the registries is expressed by different variables, e.g. *røykestatus*, *røyk*, *aktiv røyker*, *snus* and others. All of these variables are mapped to the same SNOMED CT code 365981007 - Finding of tobacco smoking behavior. Figure 15 shows variable names on the top, which are mapped to SNOMED CT codes below.



Røyker du?	Røyker du daglig?	Antall år røykt
Valg	Valg	Fritekst
Nasjonalt kvalitetsregister for ryggkirurgi : [NKR] Deformitet Skjema 1A Spørreskjema for pasienter Nasjonalt kvalitetsregister for ryggkirurgi : [NKR] Degenerativ nakke Skjema 1A Nasjonalt kvalitetsregister for ryggkirurgi : [NKR] Degenerativ rygg Skjema 1A Nye EPJ-løft lista: Hovedskjema	Norsk Nakke- og Ryggregister: Skjema 1A - Spørreskjema før behandling	Nasjonalt register for kronisk obstruktiv lungesykdom (KOLS): Registreringsskjema Nye EPJ-løft lista: Hovedskjema
SNOMED CT: 365981007: Finding of tobacco smoking behavior: Finding of tobacco smoking behavior (finding)	SNOMED CT: 365981007: Finding of tobacco smoking behavior: Finding of tobacco smoking behavior (finding)	SNOMED CT: 258707000: years: year (qualifier value) SNOMED CT: 401201003: Cigarette pack-years: Cigarette pack-years (observable entity) SNOMED CT: 365981007: Finding of tobacco smoking behavior: Finding of tobacco smoking behavior (finding)

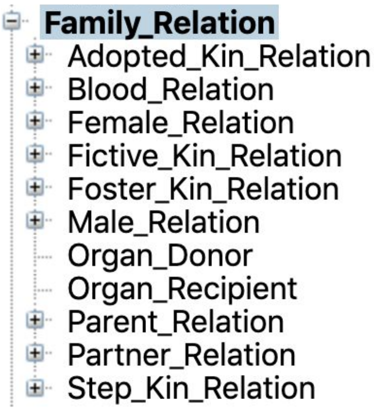
Figure 15. SNOMED CT variables and codes

SNOMED CT is one of the richest existing biomedical ontologies, and still it is not enough for all the registry variables. The variables *røyker du?* and *røyker du daglig?* are different, but are still mapped to the same code. There are multiple examples of such mapping of the variables, which can be misleading during further research. That is why it is important to reuse several ontologies, which requires a significant amount of analysis, good overview of all the relevant sources and a share of manual work.

We tried to find out whether reusing of existing ontologies improve the process of creation of new ontologies. The results were presented in detail in chapter 5.2. It shows that the reuse methods can be useful in our case, but it requires a share of manual work and a knowledge of biomedical ontologies database. The process of searching and selecting the ontology to be reused required certain background knowledge about existing biomedical ontologies. The results of the search showed that popularity of the ontology does not mean that it automatically fits our purpose. Also, the concept search outputs the list of all the ontologies with the same class, but often only a few of those classes could be considered to be reused.

After analyzing various ontologies we found out that despite their richness not all variables can be covered by existing ontologies. For example, the registry variables *fødselsmåned, alder i dager, D-nummer* are not covered in other biomedical ontologies. There are plenty of other examples of variables that do not match.

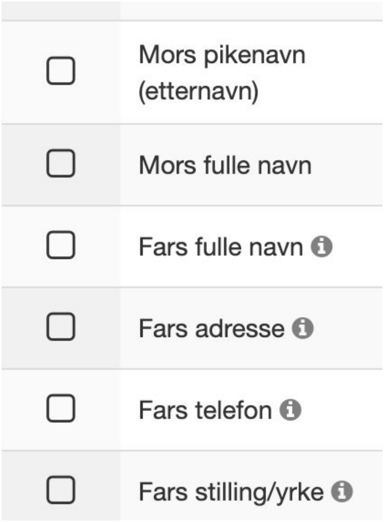
Another issue is that some ontologies are too broad for our purpose, for example the Family Health History Ontology, presented in Figure 16.



- Family_Relation
  - Adopted_Kin_Relation
  - Blood_Relation
  - Female_Relation
  - Fictive_Kin_Relation
  - Foster_Kin_Relation
  - Male_Relation
  - Organ_Donor
  - Organ_Recipient
  - Parent_Relation
  - Partner_Relation
  - Step_Kin_Relation

Figure 16. Example of Family Relation Ontology

That ontology focus is family health history, which is too detailed and mostly does not match with our purpose. The list of registry variables includes such detailed information as mother's maiden name, parent's address, profession etc. See Figure 17 for visualization.



☐ Mors pikenavn (etternavn)
☐ Mors fulle navn
☐ Fars fulle navn
☐ Fars adresse
☐ Fars telefon
☐ Fars stilling/yrke

Figure 17. Example of FHHO's registry variables

If the module from FHHO was reused in our ontology it would have required a significant amount of manual work, including editing the merged module and adding the new classes.

We know that by reusing other ontologies we improve the quality of newly created registries. If the *purpose* of the ontology fits our purpose, reusing of ontology benefits our goal. We reused a smoking module of the CSEO ontology. It focuses on cigarette exposure, which fits our purpose since there are multiple registries that have smoking related variables. We learned that before reusing any module, an ontology has to be evaluated. Also, the reused module can be more detailed than the registry variables, as long as it has the same purpose.

At the same time, our results have some limitations. We reused one module of one ontology - CSEO to create a test ontology with a certain set of variables. The results of module reuse do not cover the whole ontology and all the registry variables. Another issue that we faced is that sometimes it is quite difficult to reuse different parts of ontologies since they have different purpose than our ontology as stated in chapter 2.6 Ontology Reuse. There are might be the classes with the same name, but their meaning and the depth of its surroundings might not fit our purpose. Even after reuse, a certain amount of manual work has to be done, because the reused modules do not match 100% with the registry variables. After importing the smoking class into our ontology, we had to manually add the missing variables. Another limitation, is the manual approach by itself, considering the complexity of the ontological sources.

Referring to the issue of standardisation presented in the use case in chapter 3.2, the health registry ontology, based on PICO model and reusing other biomedical ontologies, would contribute to solving that issue. PICO would allow the registry data to be filtered with the correct tag, which could improve the quality of registry data and thus the quality of statistics. The ontology of registry variables could be used in an ontology driven system to produce new standardised registries.

Answering the research question of reuse of existing biomedical ontologies, we can say that it occurs through combining and merging ontologies and their modules, but this requires manual effort on the part of a knowledgeable ontology engineer, as well as a domain expert.

6.3. EVALUATION OF THE POPULATION ONTOLOGY

The evaluation of the Population ontology showed that the imported Smoking class successfully interacted with our ontology. We focused on queries that are related to smoking population to find out whether the ontology structure and reasoning were correct after we reused the CSEO ontology. Our results have certain limitations, because we evaluated the ontology with only smoking related questions due to reuse of the CSEO ontology. Evaluation with the whole list of questions would require the complete ontology to be created, which was not in the scope of this master thesis. Also, due to the size of the created ontology, we used only one method of evaluation. With the growth of the registry variables ontology and reuse of more resources, the need for more evaluation methods arises.

The evaluation results showed that before creating the competency questions it is important to understand the purpose of the ontology and how it will be used. Also, the competency questions have to be designed having the ontology structure and content in mind. Many of the evaluation questions that we created cannot be applied to evaluate a current ontology, due to its limited content, compared to the whole list of registry variables. However, they can be used to expand the current ontology and evaluate it later.

6.4. ISSUES WITH DESIGNING OTTR TEMPLATES

At last, we designed the OTTR templates to reflect how the health registry variables are modeled. See the detailed process of template design in Chapter 5.3. The results show that the use of OTTR templates might take us one step closer to the standardization of the registry variables. The limitation of the results is that the templates we created are not enough to satisfy the needs of non-ontology experts who are going to use the tool. They might need more variety in the templates for new types of variables. Also, it would be useful to include a PICO tag into the templates, so that the variables have already the correct category. However there is another issue here that will have to be faced - many variables have several tags. For example medication variables can be both a part of Population and Intervention. We have not included it in the scope of this master thesis as our goal was to explore how OTTR templates can be used in the creation of registry variables in order to standardize them.

Several issues occured when working on the templates. We started with creating the most general template that covers all the variables. That would be a variable that has a ?name which is both a class and an individual. And it is is an example of punning. At first we assumed that it would be correct if ?name was an individual, because the class Health Registry Variable (HRV) should contain not data, but registry variables. Also, a health registry ontologies is supposed to have HRVs as individuals. However having ?name as an individual does not allow subclass/superclass relationship with ?Parents/?Children. We are aware that punning is not the preferable way to express this, mostly because of the fact that the intended meaning can be difficult to be expressed in an engineered system. We believe that punning is a solution for that issue. In that we say that ?name is a member of HRV, but also that ?name can contain elements (namely data from the registry). The membership in HRV would be the only occurrence of punning in that case.

Another issue with OTTR templates is that they are purely syntactic, that is there is no reasoning when templates find occurences. Reasoning would allow to detect the semantic redundancies and improve maintenance of the templates.

The results of creating the templates showed that they can be used in the context of health registry variables and applied to a real life case, but the mentioned issues have to be taken into consideration. The templates can be further used in an application that creates registries, based on the variables generated by the templates. This is discussed in Chapter 6.5.

6.5. PRACTICAL USE OF THE OTTR TEMPLATES. REGISTRY BUILDER TOOL.

The created OTTR templates can be used in an ontology driven information system which helps medical professionals to create better registries. It is a tool that helps clinical researchers and medical personnel to define health registry variables by constructing the health registry ontology. That tool is based on the OTTR templates, which are the representation of different types of variables (Categorical, Continuous and Free Text).

The tool takes a variable, that is needed to create a registry, as an input. OTTR templates are the basis for creating the ontology. The output OWL file contains all the classes that are related to the registry ontology. The registry data storage owners import the health registry variables (the OWL file) from the registry builder to their registry in an ontology-driven system. The registry data, as well as the OWL files, can be exported to the Clinical Data Analytics Tool, where researchers can view data from various registries that have defined variables using the registry builder tool. This tool can output datasets needed for the statistical analysis.

For the templates to be used by all the medical workers and researchers they have to be centralized. That ensures that there will not be created an unlimited amount of templates, which does not solve the problem of standardization. Figure 18 gives a graphical representation of the potential registry builder and how it could be used below.

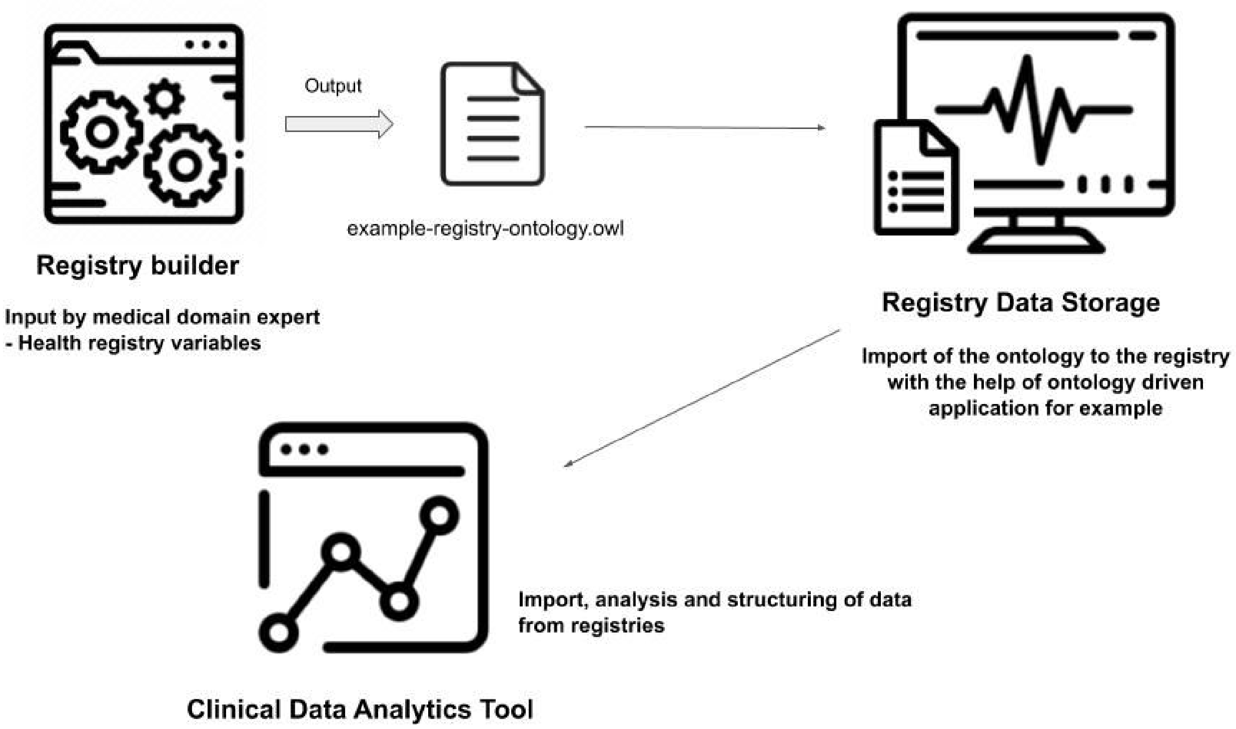


Figure 18. The graphical representation of the potential registry builder

7. CONCLUSION AND FUTURE WORK

In this thesis we explored the topics of ontology reuse and applying the OTTR templates in the context of health registries. We found out that:

- As discussed in Chapters 6.1 and 6.2, reuse of existing biomedical ontologies occurs through combining and merging ontologies and their modules, but this requires manual effort on the part of an experienced ontology engineer, as well as knowledge domain experts.
- As pointed out in Chapter 6.4, several issues occur while designing the OTTR templates, that need to be taken into consideration, for instance punning and no possibility for reasoning. But the results in Chapter 5.3 show that the OTTR templates can be applied to a real life case - health registry variables and be later used in registry builder application.

Future master students can use this research as a basis for future work. This is a theoretical master thesis, which has a potential for practical implementation. We contributed to the knowledge about reuse of biomedical ontologies and designing the OTTR templates in the particular use case of the Health Registries in Norway. There are several topics that could be further studied in the future:

1. Following Chapters 5.1 and 5.2, a bigger and more complete Health Registry Variables ontology could be created. In the future work, creating the full Health Registry variables ontology implies the following steps:

    - Reusing the whole PICO ontology;
    - More biomedical ontologies to be evaluated for reuse;
    - Reusing more ontologies;
    - Modeling all the registry variables in ontology

    Due to the size and complexity of the complete Health Registry Variables ontology, more automated methods have to be used. Pursuing this topic implies having an ontology engineering background, as well as knowledge about the biomedical domain. This would require a team of experienced ontology engineers with close collaboration with the medical domain specialists.

2. Following Chapters 5.3 and 6.5, the OTTR could be developed into a complete list of templates and applied in a registry builder application. We believe this tool has great potential and could bring us one step closer to a better and more standardized registries. However, continuing research in this direction implies having a good

> understanding of OTTR templates and their limitations. Creating a full functioning system would require a close collaboration of template developers, ontology engineers, medical workers and most importantly support from E-health Department. The registry builder tool would be able to address the standardisation issue, if the templates are centralised. That is why the approval and support of E-health Department would be necessary.

The target group that could find useful the outcome of this project includes medical workers, researchers and new master students. Clinicians who are the domain experts will benefit with better use of registries themselves. Better quality of registries will result in better health services and more accurate data.

APPENDIX A COMPETENCY QUESTIONS

Individuals/Population

1. All individuals with a specific ethnic background with a specific disease or type of disease.
2. All individuals with a certain innleggelse dato.
3. All individuals who are men and have a tobacco smoking addiction.
4. All individuals who are women on hormonal medication, who had a stroke and smoke cigarettes.
5. All individuals who are smoking tobacco.
6. Individuals who spent in the hospital (*innleggelse*) more than a certain amount of days.
7. All antibiotics which were prescribed to patients with a certain disease after being in the hospital.
8. All individuals with a certain disease wha have a relative with the same disease.
9. The list of diseases of people older than a certain age with some habit (smoking for example).
10. Which individuals should be checked for update (related to the variable query over which free text and structured are related)

Queries over variables:

1. What are all the demographic variables?
2. Variables connected with heart disease.
3. Which variables are associated with documenting patients who are
4. Which variables ask about non-white ethnic background?
5. How is ethnicity captured in variables?
6. Which variables are relevant for women only? Men only?
7. Which variables are related to genetics? Or are relevant for clinical phenotypes?
8. Which variables collect information about events that occur in certain locations? (e.g. outside of Norway?)
9. Which variables are relevant quality indicators for X department that can be used for reporting. Quality indicators should be based on guidelines- benchmarking that gives benefit. Things that are easy to measure are not necessarily good things. (Styringsdata).
10. Which variables can change? Smoking status, etc. (age started smoking is final for example)
11. Which variables have free text that needs to be processed?
12. Which free text variables and which structured variables have a relationship?

13. What are the values of the variables? Which variables have similar values? Which variables are overlapping? - which values include values under another variable?
14. Which of the variables can be used for predictions (events should have time stamped?) (add metadata about variables, the ones that are events are useful for patient trajectory)

Query over both variables and individuals data (at the time of filling out the form):

1. Which data is missing? (this is query for individual that occurs when people are filling out the forms) - continuous process, one person could fill out one thing and another person fills out another… (the last person to see the patient needs to know and take responsibility that the registry data is completed). (process help)

Query related to outside modeled info:

1. Who is the person who filled out this variable info. - process and provenance of data.

APPENDIX B OTTR TEMPLATES

- We created the template for the Mutually Exclusive Categorical HRV, which means as mentioned in Chapter 5.3, that only one value from ?vals can be chosen.
  hrv-template:MutuallyExclusiveCategoricalHRV[owl:NamedIndividual ?name, List<xsd:string> ?vals, List<xsd:string> ?SNOMEDCodes, ? xsd:string ?formText, ? List<owl:Class> ?Parents, ? List<owl:Class> ?Children, ? List<owl:Class> ?RelatedClasses] :: {

- The call of the more general template which is Categorical HRV. *hrv-template:CategoricalHRV(?name, ?vals, ?SNOMEDCodes, ?formText, ?Parents, ?Children, ?RelatedClasses),*

- The name has the type of a Mutually Exclusive Categorical HRV
  ottr:Triple(?name, rdf:type, hrv-class:MutuallyExclusiveCategoricalHRV),

- Name has precisely one value for hasValue (that value can be anything). The call to CategoricalHRV above ensures that this value must be from the list of legal values.

  o-owl-ax:SubObjectExactCardinality(?name, 1, hrv-prop:hasValue, owl:Thing).}.

- Continuous HRV. Value is given in a number, for example age, days spent in the hospital and the amount of cigarettes per day.
  hrv-template:ContinuousHRV[owl:NamedIndividual ?name, ? xsd:integer ?input, ? xsd:string ?formText, ? List<owl:Class> ?Parents, ? List<owl:Class> ?Children, ? List<owl:Class> ?RelatedClasses] :: {

- HRV has the type Continuous.
  hrv-template:HealthRegistryVariable(?name, ?formText, ?Parents, ?Children, ?RelatedClasses, hrv-class:ContonuousTextHRV),

- Value that is an input
  Ottr:Triple(?name, hrv-prop:hasValue, ?input).}.

- Free Text HRV. Value is given as a string in the form a free text, for example D-number, social security number, name etc.
  hrv-template:FreeTextHRV[owl:NamedIndividual ?name, ? xsd:string ?input, ? xsd:string ?formText, ? List<owl:Class> ?Parents, ? List<owl:Class> ?Children, ? List<owl:Class> ?RelatedClasses] :: {

- HRV has the type Free Text.
  hrv-template:HealthRegistryVariable(?name, ?formText, ?Parents, ?Children, ?RelatedClasses, hrv-class:FreeTextHRV),
  The ?name can have one value
  Ottr:Triple(?name, hrv-prop:hasValue, ?input).}.

BIBLIOGRAPHY

AberOWL (no date). Available at: http://aber-owl.net/ontology/#/ (Accessed: 12 November 2019).

Baader, F. et al. (2003) The Description Logic Handbook: Theory, Implementation and Applications. Cambridge University Press.

Baader, F. and Nutt, W. (2002) 'Basic Description Logics', The Description Logic Handbook, pp. 47–104. doi: 10.1017/cbo9780511711787.004.

Basic Formal Ontology (BFO) | Home (no date). Available at: http://basic-formal-ontology.org/ (Accessed: 12 November 2019).

Blomqvist, E. and Sandkuhl, K. (2005) 'PATTERNS IN ONTOLOGY ENGINEERING: CLASSIFICATION OF ONTOLOGY PATTERNS', Proceedings of the Seventh International Conference on Enterprise Information Systems. doi: 10.5220/0002518804130416.

Blomqvist, E. (2009). Semi-automatic ontology construction based on patterns (Doctoral dissertation, Linköping University Electronic Press).

Blomqvist E., Daga E., Gangemi A., Presutti V. (2010) Modelling and using ontology design patterns.

Blumenthal, S. (2018) 'Improving Interoperability between Registries and EHRs', AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science, 2017, pp. 20–25.

Bodenreider, O., Smith, B., Kumar, A. and Burgun, A. (2004) Investigating subsumption in DL-based terminologies: A case study in SNOMED CT.

Bontas, E.P., Mochol, M. and Tolksdorf, R. (2005) Case studies on ontology reuse. In Proceedings of the IKNOW05 International Conference on Knowledge Management (Vol. 74, p. 345).

Bontas, E. P., Schlangen, D. and Niepage, S. (2005) 'Ontology Engineering for the Semantic Annotation of Medical Data', 16th International Workshop on Database and Expert Systems Applications (DEXA'05). doi: 10.1109/dexa.2005.150.

Brochhausen, M. et al. (2013) 'Developing a semantically rich ontology for the biobank-administration domain', Journal of biomedical semantics, 4(1), p. 23.

Caldarola, E. G. and Rinaldi, A. M. (2016) 'An Approach to Ontology Integration for Ontology Reuse', 2016 IEEE 17th International Conference on Information Reuse and Integration (IRI). doi: 10.1109/iri.2016.58.

Caldarola, E. G. and Rinaldi, A. M. (2018) 'A Multi-strategy Approach for Ontology Reuse Through Matching and Integration Techniques', Advances in Intelligent Systems and Computing, pp. 63–90. doi: 10.1007/978-3-319-56157-8_4.

Ceusters W, Smith B, Flanagan J. (2003) Ontology and medical terminology: why description logics are not enough. TEPR; 2003. San Antonio;.

Chatterjee, A. (2017) Elements of Information Organization and Dissemination. Chandos Publishing.

Cigarette Smoke Exposure Ontology | NCBO BioPortal (no date). Available at: https://bioportal.bioontology.org/ontologies/CSEO (Accessed: 13 November 2019).

Cochrane (no date). Available at: https://www.cochrane.org/ (Accessed: 13 November 2019).

Doran, P. (2006) 'Ontology Reuse via Ontology Modularisation' In Proceedings of KnowledgeWeb PhD Symposium 2006 (KWEPSY2006). Budva, Montenegro

Doran, P., Tamma, V. and Iannone, L. (2007) 'Ontology module extraction for ontology reuse', Proceedings of the sixteenth ACM conference on Conference on information and knowledge management - CIKM '07. doi: 10.1145/1321440.1321451.

Dugas, M. et al. (2016) 'ODMedit: uniform semantic annotation for data integration in medicine based on a public metadata repository', BMC medical research methodology, 16, p. 65.

Egaña, M. et al. (2008) 'Applying Ontology Design Patterns in Bio-ontologies', Knowledge Engineering: Practice and Patterns, pp. 7–16. doi: 10.1007/978-3-540-87696-0_4.

Family Health History Ontology | NCBO BioPortal (no date). Available at: https://bioportal.bioontology.org/ontologies/FHHO (Accessed: 13 November 2019).

Gangemi, A. and Presutti, V. (2009) 'Ontology Design Patterns', Handbook on Ontologies, pp. 221–243. doi: 10.1007/978-3-540-92673-3_10.

Gastronet (no date). Available at: https://www.kvalitetsregistre.no/registers/543/resultater (Accessed: 13 November 2019).

Gene Ontology Resource (no date) Gene Ontology Resource. Available at: http://geneontology.org/ (Accessed: 12 November 2019).

Helsedataprogrammet (no date) ehelse.no. Available at: https://ehelse.no/helsedataprogrammet (Accessed: 7 May 2019).

Helseregistre og andre registre (no date) Folkehelseinstituttet. Available at: http://www.fhi.no/hn/helseregistre-og-registre/ (Accessed: 13 November 2019).

Hlomani, H. (2014) 'Approaches, methods, metrics, measures, and subjectivity in ontology evaluation: A survey' Semantic Web Journal

Hotz, L. et al. (2014) 'Configuration Knowledge Representation and Reasoning', Knowledge-Based Configuration, pp. 41–72. doi: 10.1016/b978-0-12-415817-7.00006-2.

Javor, A. et al. (2019) 'Stroke risk factors and treatment variables in rural and urban Austria: An analysis of the Austrian Stroke Unit Registry', PLOS ONE, p. e0214980. doi: 10.1371/journal.pone.0214980.

Kaljurand, K. (no date) OWL axiom examples. Available at: https://files.ifi.uzh.ch/cl/kalju/ontologies/OWL_spec/owl_spec_examples.html (Accessed: 12 November 2019).

KARTLEGGING AV VARIABLER I NASJONALE HELSEREGISTRE (no date). Available at: https://www.fhi.no/globalassets/dokumenterfiler/rapporter/2017/kartlegging-av-variabler-i-nasjonale-h elseregistre.pdf (Accessed: 13 November 2019).

Katsumi, M. and Grüninger, M. (2017) ‘Choosing ontologies for reuse’, Applied Ontology, pp. 195–221. doi: 10.3233/ao-160171.

Kindermann, C., Parsia, B. and Sattler, U. (2019) ‘Detecting Influences of Ontology Design Patterns in Biomedical Ontologies’, Lecture Notes in Computer Science, pp. 311–328. doi: 10.1007/978-3-030-30793-6_18.

KitAMO (no date). Available at: https://www.ida.liu.se/~patla00/research/KitAMO/ (Accessed: 13 November 2019).

Lantow, B., Sandkuhl, K. and Tarasov, V. (2013) Selecting Content Ontology Design Patterns for Ontology Quality Improvement. In ILOG@ BIR (pp. 68-79).

Lupp, D.P. (2019) A Higher-Level View of Ontological Modeling: Rule-Based Approaches for Data Transformation, Modeling, and Maintenance. pp. 4-5,17-18

Mavergames, C., Oliver, S. and Becker, L. (2013) Systematic Reviews as an Interface to the Web of (Trial) Data: using PICO as an Ontology for Knowledge Synthesis in Evidence-based Healthcare Research. In SePublica (pp. 22-26).

Masic, I. (2014) ‘Five Periods in Development of Medical Informatics’, Acta Informatica Medica, p. 44. doi: 10.5455/aim.2014.22.44-48.

Medisinske Kvalitetsregistre (no date). Available at: https://www.kvalitetsregistre.no/om-medisinske-kvalitetsregistre-0 (Accessed: 13 November 2019).

Min, H. et al. (2009) ‘Integration of Prostate Cancer Clinical Data Using an Ontology’, Journal of biomedical informatics. NIH Public Access, 42(6), p. 1035.

Nasjonale registervariabler - metadata (no date). Available at: http://hregrapport2017.kodemakeriet.no/Default.cshtml (Accessed: 15 November 2019).

Nasjonal strategi: Slik skal vi jobbe med kvalitetsregistrene fremover (no date). Available at: https://www.kvalitetsregistre.no/artikkel/nasjonal-strategi-slik-skal-vi-jobbe-med-kvalitetsregistrene-fremover (Accessed: 13 November 2019).

Nasjonalt kvalitetsregister for barnekreft (no date). Available at: https://www.kvalitetsregistre.no/registers/731/resultater (Accessed: 13 November 2019).

NCBO BioPortal (no date). Available at: https://bioportal.bioontology.org/ontologies (Accessed: 11 November 2019).

Noy, N.F. and McGuinness, D.L. (2001) `Ontology Development 101: A Guide to Creating Your First Ontology' Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880

Norsk hjerneslagregister (no date). Available at: https://www.kvalitetsregistre.no/registers/353/resultater (Accessed: 13 November 2019).

Obitko, M. (no date) Description Logics - Introduction to ontologies and semantic web - tutorial.

Available at: https://www.obitko.com/tutorials/ontologies-semantic-web/description-logics.html (Accessed: 12 November 2019).

OPTypes - Odp (no date). Available at: http://ontologydesignpatterns.org/wiki/OPTypes (Accessed: 12 November 2019).

OWL 2 Web Ontology Language Manchester Syntax (Second Edition) (no date). Available at: https://www.w3.org/TR/2012/NOTE-owl2-manchester-syntax-20121211/ (Accessed: 12 November 2019).

OWL 2 Web Ontology Language Profiles (Second Edition) (no date). Available at: https://www.w3.org/TR/owl2-profiles/#OWL_2_EL (Accessed: 12 November 2019).

OWL Web Ontology Language Overview (no date). Available at: https://www.w3.org/TR/owl-features/ (Accessed: 14 November 2019).

PICO | helsebiblioteket (no date). Available at: https://www.helsebiblioteket.no/kunnskapsbasert-praksis/sporsmalsformulering/pico (Accessed: 13 November 2019).

PICO ontology (no date). Available at: https://linkeddata.cochrane.org/pico-ontology (Accessed: 13 November 2019).

Prasser, F. et al. (2018) ‘Data Integration for Future Medicine (DIFUTURE)’, Methods of information in medicine, 57(S 01), pp. e57–e65.

Prosjekt Harmonisering i Helsedataprogrammet (no date) ehelse.no. Available at: https://ehelse.no/helsedataprogrammet/prosjekt-harmonisering-i-helsedataprogrammet (Accessed: 7 May 2019).

Prostate Cancer Life Style Ontology | NCBO BioPortal (no date). Available at: https://bioportal.bioontology.org/ontologies/PCAO/?p=summary (Accessed: 13 November 2019).

Punning - OWL (no date). Available at: https://www.w3.org/2007/OWL/wiki/Punning (Accessed: 14 November 2019).

Raad, J. and Cruz, C. (2015) ‘A Survey on Ontology Evaluation Methods’, Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management. doi: 10.5220/0005591001790186.

Read Codes, Clinical Terms Version 3 (CTV3) | NCBO BioPortal (no date). Available at: https://bioportal.bioontology.org/ontologies/RCD (Accessed: 13 November 2019).

Reasonable Ontology Templates (OTTR) (no date). Available at: http://ottr.xyz/#Introduction_and_Overview (Accessed: 13 November 2019).

Rebstock, M., Janina, F. and Paulheim, H. (2008) Ontologies-Based Business Integration. Springer Science & Business Media.

Rosse, C. and Mejino, J. L. V., Jr (2003) ‘A reference ontology for biomedical informatics: the Foundational Model of Anatomy’, Journal of biomedical informatics, 36(6), pp. 478–500.

Ruiz, F. and Hilera, J. R. (2006) ‘Using Ontologies in Software Engineering and Technology’, Ontologies for Software Engineering and Software Technology, pp. 49–102.

Saranto, K. (2009) Connecting Health and Humans: Proceedings of NI2009. IOS Press, p. 348.

Sidhu, A. S. (2009) Biomedical Data and Applications. Springer, p. 148,149.

Simperl, E. (2009) 'Reusing ontologies on the Semantic Web: A feasibility study', Data & Knowledge Engineering, pp. 905–925. doi: 10.1016/j.datak.2009.02.002.

Skjæveland, M.G., Forssell, H., Klüwer, J.W., Lupp, D.P., Thorstensen, E. and Waaler, A. (2017) Pattern-Based Ontology Design and Instantiation with Reasonable Ontology Templates. In WOP@ ISWC.

Smith, B. et al. (2007) 'The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration', Nature biotechnology, 25(11), pp. 1251–1255.

Smith, B., Ceusters, W., Klagges, B., Köhler, J., Kumar, A., Lomax, J., Mungall, C., Neuhaus, F., Rector, A.L. and Rosse, C. (2005) Relations in biomedical ontologies. Genome biology, 6(5), p.R46.

Smith, B. and Kumar, A. (2004) 'Controlled vocabularies in bioinformatics: a case study in the gene ontology', Drug Discovery Today: BIOSILICO, pp. 246–252. doi: 10.1016/s1741-8364(04)02424-2.

SNOMED Home page (no date) SNOMED. Available at: http://www.snomed.org/ (Accessed: 12 November 2019).

SOMER: Semantic Ontology Matching using External Resources (no date). Available at: http://somer.rd.ciencias.ulisboa.pt/ (Accessed: 13 November 2019).

Stuckenschmidt, H., Van Harmelen, F., Serafini, L., Bouquet, P. and Giunchiglia, F. (2004) Using C-OWL for the alignment and merging of medical ontologies. University of Trento.

Suárez-Figueroa, M. C., Gómez-Pérez, A. and Villazón-Terrazas, B. (2009) 'How to Write and Use the Ontology Requirements Specification Document', On the Move to Meaningful Internet Systems: OTM 2009, pp. 966–982. doi: 10.1007/978-3-642-05151-7_16.

Takai-Igarashi, T. and Tanaka, H., 'A Resolution of the Problem of Multiple Inheritance in Gene Ontology'.

Talebpour, M., Sykora, M. and Jackson, T. (2018) 'Ontology Selection for Reuse: Will It Ever Get Easier?', Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management. doi: 10.5220/0006937101080116.

Taye, M.M. (2011) 'Web-based ontology languages and its based description logics' The Research Bulletin of Jordan ACM, 2(1), pp.1-9.

The Galen Project | Universiteit Utrecht (no date) The Galen Project. Available at: https://projectgalen.sites.uu.nl (Accessed: 13 November 2019).

The OBO Foundry (no date). Available at: http://www.obofoundry.org/ (Accessed: 13 November 2019).

The TAMBIS Project (no date). Available at: http://www.cs.man.ac.uk/~stevensr/tambis/details.html (Accessed: 13 November 2019).

Torgersen, T.-V. (2003) Kroppen.

‘Unified Medical Language System (UMLS)’ (2009). U.S. National Library of Medicine. Available at: https://www.nlm.nih.gov/research/umls/index.html (Accessed: 12 November 2019).

Welcome to the NCBO BioPortal | NCBO BioPortal (no date). Available at: https://bioportal.bioontology.org/ (Accessed: 13 November 2019).

What is an Ontology? (no date). Available at: http://www.cs.man.ac.uk/~stevensr/onto/node3.html (Accessed: 13 November 2019).

Younesi, E. et al. (2014) ‘CSEO - the Cigarette Smoke Exposure Ontology’, Journal of biomedical semantics, 5, p. 31.

Yu, A. C. (2006) ‘Methods in biomedical ontology’, Journal of Biomedical Informatics, pp. 252–266. doi: 10.1016/j.jbi.2005.11.006.

Zagorulko, Y., Borovikova, O. and Zagorulko, G. (2018) ‘Development of Ontologies of Scientific Subject Domains Using Ontology Design Patterns’, Communications in Computer and Information Science, pp. 141–156. doi: 10.1007/978-3-319-96553-6_11.

Zulkarnain, N. Z., Meziane, F. and Crofts, G. (2016) ‘A Methodology for Biomedical Ontology Reuse’, Natural Language Processing and Information Systems, pp. 3–14. doi: 10.1007/978-3-319-41754-7_1.