Ontology-Based Modelling of Stroke Clinical Pathways

By

Matteo Crupi

A Thesis Submitted to Department of Electrical and Computer Engineering

in Partial Fulfillment of the Requirement for the Degree of

Master of Science

at

Lakehead University

Thunder Bay, Ontario, Canada

December 2019

Acknowledgement

I would like to thank my supervisor, Dr. Rachid Benlamri, whose guidance was much appreciated and crucial to the completion of this work.

I would also like to thank my colleagues, friends and family. Without their assistance, this work would never have been possible.

This work was conducted using the Protégé resource, which is supported by grant GM10331601 from the National Institute of General Medical Sciences of the United States National Institutes of Health.

Abstract

Healthcare spending in Canada is on the rise. One method to reduce healthcare spending is to reduce length of stay (LOS). Clinical Pathways (CPs) are one recommended management technique to reduce LOS. CPs are implementations of medical guidelines in a specific healthcare environment. This may include hospitals, clinics or other healthcare facilities. They represent an evidence-based patient care workflow for a specific disease. The adoption of CPs allows for easier continuity of care across different healthcare settings and medical teams. While the use of CP as part of standard patient care has grown considerably in the past decades, not much progress was made in CP representation and modeling to encode CP data properly within existing Health Information Systems (HIS). One proposed method to achieve this goal is ontological modeling. Ontology is a formal model that represents a certain subject matter. It not only communicates what things exist in a certain domain or field but also how those things relate to each other. This research proposes an ontological model for stroke CP representation and processing. Such a model would allow CPs to be sharable, extendable, and machine-readable, thus enabling greater patient management. The Systematized Nomenclature of Medicine – Clinical Terms (SNOMED – CT) is used to encode medical knowledge described within clinical pathways.

The stroke CP Ontology is an extension of a generic CP ontology, with new concepts introduced specific to the domain of stroke. It is able to represent different types of CP activities, occurring over a period of type, referencing medical knowledge contained in SNOMED CT. It is also able to infer new knowledge using the Semantic Web Rule Language (SWRL). This ontology is presented to users through a prototype Clinical Pathway Management System (CPMS). The CPMS is built using Java and the Eclipse IDE. The OWL and SWRL API are used to directly connect to and query ontology files. After completion of a CP, the CPMS generates new ontology files unique to each patient's CP execution as well as a general output file of patient activities and outcomes. Data analytics can be performed on this output file to determine the most common CP activities, levels of compliancy and similarities between patient CP progressions.

Table of Contents

1 Introduction 10
1.1 Problem Statement and Motivation 11
1.2 Research Challenges 13
1.3 Research Objectives 14
1.4 Applications and Benefits 14
1.5 Thesis Structure 15
2 Background and Related Work 16
2.1 Clinical Pathway 17
2.2 Ontology 17
2.2.1 Semantic Modeling and Basic Structure 17
2.2.2 Web Ontology Language (OWL) 18
2.2.3 Allen's Algebra 19
2.2.4 Semantic Web Rule Language (SWRL) 22
2.2.5 Protégé 24
2.3 Systematized Nomenclature of Medicine – Clinical Terms (SNOMED – CT) 24
2.4 Literature Review 25
3 CP Ontology and Semantic Modelling of Clinical Pathways 35
3.1 Conceptual Modeling of Clinical Pathways 36
3.2 Meta Clinical Pathway Ontology 36
3.3 Medical Knowledge Ontology 41
3.4 Time Ontology 43
3.5 Stroke CP Ontology Modelling 47
3.5.1 Event-Intervention and Observation-Outcome Execution Chains 54
3.5.2 The Duality of Tasks 56
3.6 Clinical Pathway Management System 56
3.6.1 Software Development Environment 56
3.6.2 System Architecture and Presentation 57
3.6.3 System Execution and CP Semantic Interoperability 59
3.6.4 Typical Use Cases and Scenarios 61
3.7 Discussion 68
4 Experimental Results 69
4.1 Generation of Patient Results and System Output 70
4.2 Data Analytics 73
4.2.1 Longest Common Subsequence 74
4.2.2 Spatial Analysis – Longest Common Subsequence in 3D Space 80

4.3 Discussion 82
5 Conclusion and Future Work 84
6 References 88

List of Figures

Figure 1: Healthcare spending in Canada as a share of GDP. 11
Figure 2: A generic RDF triple. 18
Figure 3: A potential CP scenario modelled using the components of OWL 2. 19
Figure 4: A standard time interval with endpoints denoted as '-' and '+'. 21
Figure 5: The 13 major relationships between temporal intervals [8]. 22
Figure 6: The main GUI for Protégé, with a hierarchical view of created classes. 24
Figure 7: The ID and term name for the concept Clinical Finding. 25
Figure 8: The ontology of [27]. 26
Figure 9: The ontology of [36]. 30
Figure 10: CP State evolution from admission to discharge [34]. 31
Figure 11: The ontological model of [33]. 32
Figure 12: The object properties linking CP Event, Event Path and Variance. 38
Figure 13: The definition of Event Path with reference to the research of [32] which served as major inspiration. 38
Figure 14: Event Path with children Branch Path and Common Path. 38
Figure 15: The Meta CP ontology. 39
Figure 16: The class definition for a member of the MK ontology. 41
Figure 17: A subsection of the MK ontology representing the ancestry of the class "Brushing of Teeth". 42
Figure 18: The relationship between Nicotine Replacement Intervention and Discussion about changes in Lifestyle. 43
Figure 19: Discussion about changes in lifestyle in the greater MK ontology. 43
Figure 20: Temporal Entity hierarchy. Interval relations are defined within class Proper Interval [24]. 45
Figure 21: The Temporal Duration hierarchy [24]. 45
Figure 22: The Temporal Entity hierarchy with Day Interval, Trace Interval, Admission Time and Discharge Time. 46
Figure 23: The relationship between intervals in the CP ontology. 47
Figure 24: Patient CP Education hierarchy. 48
Figure 25: Document hierarchy. 48

Figure 26: The CP Event component of the Stroke CP ontology. 51
Figure 27: The semantic relationships of the CP Assessment Complete Screening for Dysphagia. 53
Figure 28: The hierarchy of CP ontologies. 53
Figure 29: The interrelationship of CP Event, Observation, Intervention and Outcome. 54
Figure 30: MVC paradigm applied to the structure of the CPMS. 57
Figure 31: The GUI for Day 2 Nutrition Events in the CPMS. 58
Figure 32: CPMS View class diagram. 59
Figure 33: Screening for Dysphagia Procedure results prompt. 63
Figure 34: Screening for Dysphagia Procedure result options. 63
Figure 35: A prompt for the availability of the Speech Language Pathologist. 64
Figure 36: A prompt to consult the Speech Language Pathologist. 64
Figure 37: A prompt to repeat the Screening for Dysphagia Procedure (added to Day 2 of CP). 65
Figure 38: Repeated Screening for Dysphagia Procedure results prompt. 65
Figure 39: A prompt to consult with a physician because of the failed Swallowing Screening... 65
Figure 40: The CPMS Search Bar. 66
Figure 41: The CPMS Search Bar with a dropdown menu of all SNOMED CT Concepts. 67
Figure 42: The CPMS Search Bar with the results of the search for the term “Care”................. 67
Figure 43: The nicotine replacement event from the generic knowledge graph. No outcome created. 71
Figure 44: The discussion observation from the generic knowledge graph. No observation value created. 72
Figure 45: The nicotine replacement event from the patient knowledge graph. New outcome created. 72
Figure 46: The discussion observation from the patient knowledge graph. New observation value created. 72
Figure 47: The hierarchy of ontologies. 73
Figure 48: The interventions that occurred for each patient during the ischemic stroke CP. 74
Figure 49: Stroke Monitoring is performed instead of Screening for Dysphagia. 75
Figure 50: The output for intervention Screening for Dysphagia. 79
Figure 51: The output file for the Stroke CP. 79

Figure 52: Patient Groups by CP Longest Common Subsequence... 80
Figure 53: A potential CP cube with landmark point circled in red. 82

List of Tables

Table 1: The endpoint relationships that exist between interval t and s. 21
Table 2: A summary of the literature review with contributions of this research. 33
Table 3: The major Meta CP ontology classes and their children. 40
Table 4: The event breakdown for the Stroke CP. 48
Table 5: Event-Day Matrix categorized by Event type. 49
Table 6: The components of the Swallowing Screening Event. 62
Table 7: The workflow of the Swallowing Screen Event in the ischemic stroke CP. 62
Table 8: The solution to the LCS algorithm for “card” and “cord”. 76
Table 9: The solution to the LCS algorithm for 2 SNOMED CT ID chains. 77
Table 10: LCS Algorithm output. 79
Table 11: A subsection of a CP over a certain day. 81

1 Introduction

1.1 Problem Statement and Motivation

In 2019, it is predicted that healthcare expenditure in Canada will be approximately $264 billion, with about 25% of spending involving hospitals [1]. This represents a two percent increase from 2018. Furthermore, healthcare spending in Canada was a growing share of Canada's Gross Domestic Product (GDP). Only in the past decade has there been a decrease in the overall share of Canada's GDP, as seen in Figure 1. This decrease was a result in major investments in healthcare during the 1990's and 2000's [17]. Even with increased investments, approximately 20% of hospital spending will include nursing inpatient services [11], while conditions such as Chronic Obstructive Pulmonary Disease (COPD), heart failure, pneumonia and knee replacement were the most expensive hospital conditions of 2017 [12].

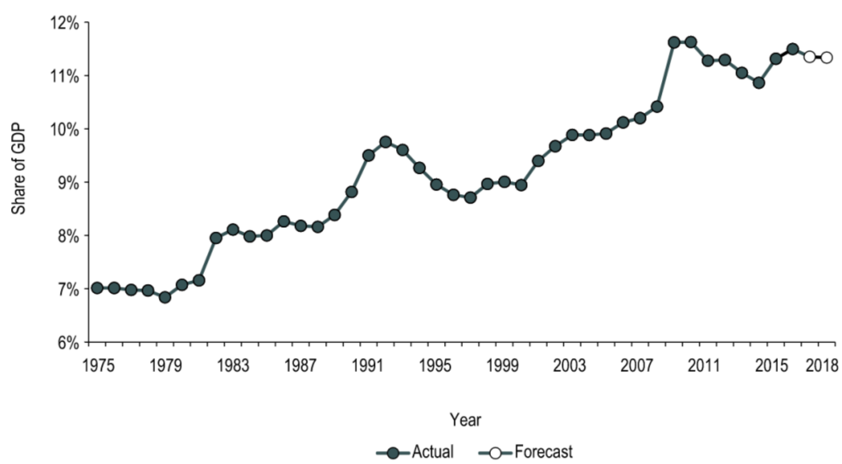

Figure 1: Healthcare spending in Canada as a share of GDP.

One approach to reduce hospital spending is to reduce patient length of stay (LOS). There are several methods recommended to reduce LOS, with one solution being to follow standardized practice guidelines and critical pathways during care. There is also a recent trend in Canadian healthcare to shift from inpatient to outpatient care. Since 2005, outpatient care has grown about 1.5 times as much as its inpatient counterpart [13]. With this movement towards out-of-hospital care, there is also a need to follow standardized practice guidelines. If such standardization of healthcare treatment can be achieved between different institutions, inter-professional medical teams and across different care settings, there should be an overall reduction in LOS [2]. Recently, this coordination and standardization has been facilitated using Clinical Pathways (CP) or Clinical Pathway Workflows (CPW). Historically, these CP have paper-based, institution specific instructions. However, paper-based CPs have several disadvantages.

As a physical document, there is limited space on a paper-based CPs for recording information. This not only limits the amount of shareable information on each CP, but also introduces space constraints as more CPs are created and stored in a physical location. Paper-based CPs also impede meaningful connections between different patient records. If a collection of patient data is spread across several different documents (or geographical locations), it becomes extremely difficult (or impossible) to reference such information on a patient CP. Furthermore, it is difficult to represent real-time changes in CP progression using paper-based documents. A paper-based CP can only give a static view of the current CP, limiting historical context and knowledge on previous patient condition. Semantic connections between different CP activities are also minimal when using paper-based CP. For example, if a certain activity should always directly follow another activity, the only way to express this relationship is through some written statement. There is no other mechanism linking these two events. In such a situation, the likelihood of missing or misunderstanding the instruction becomes more likely. Finally, there is no accepted standard for the creation of paper-based CP.

With the deficiencies of paper-based CPs, there has been a shift towards digitization. The proposed method of digitization in this research is ontology. Ontology represents what exists in a certain domain and how such things relate to one another. It can be considered a shared, standardized knowledge base. Among other benefits, digitization of CPs using ontology allows for greater reusability, extensibility, data visibility and semantic interoperability.

With digital CPs, healthcare professionals and different healthcare teams distributed geographically would be able to more easily share patient CP progression and potentially work concurrently. Presently, hospital specialized paper-based CP make it difficult for healthcare professionals to share CPs across institutions either regionally, nationally or internationally. A lack of standardized terminology and no semantic relationships between CP components, make it difficult for other healthcare teams to share knowledge or gather the full context of a patient's CP. With ontology digitization, the progression of a patient through a CP in one healthcare environment could be captured before being distributed to other stakeholders and continued. The resulting CP would be a dynamic reflection of the patient's healthcare journey pulling from several different data (and geographical sources).

Digitization also offers the potential of machine readability. CPs that are machine readable can be audited and analyzed in an efficient manner not possible in paper-based alternatives. This includes increased potential for data analytics on patient data, outcomes and overall pathways taken. This also facilitates the use of automated Natural Language Processing (NLP) not possible with paper-based CPs. From an auditing perspective, this includes detection of redundant CP activities and potential elimination of ineffective pathways. Therefore, the motivation of this research is to reduce healthcare spending in Canada by digitizing CP using ontology as a semantic framework.

1.2 Research Challenges

There are several main challenges present in this research. These vary in nature from potential technical challenges to more social challenges. These challenges include:

I. Developing a generic, yet extensible, ontology to represent CP. There is no accepted standard terminology for CP, causing difficulty when attempting to represent this domain. CP also vary greatly in construction depending on disease and location

II. Extending such a generic CP to a specific disease CP in a routine manner. This will require the use of a recognized hospital's CP, domain expert collaboration and the establishment of modelling steps.

III. Understanding complex (sometimes obscure) medical data and knowledge. This will again require the opinion of domain experts and also self-education. The author has a limited understanding of the medical domain, which poses obstacles in modelling.

IV. Representing and standardizing medical knowledge in a widely accepted format. There is no standard for the representation of medical knowledge provincially, nationally or internationally. A validated, consistent standard should therefore be adopted.

V. Adopting a consistent, non-ambiguous temporal system. Written and spoken instructions during CP execution can be vague or ambiguous. For example, an event happening on day 2 of a CP could be recorded to the hour, minute or second.

VI. Capturing variance in CP execution. It is common for a CP to change based on patient experiences and outcomes. There is no current standard to identify and record these causes of variance. The challenge of identifying these sources of variances and then capturing them within ontology should be overcome.

VII. Representing CP in a format that will be supported and endorsed by medical professionals. A proposed CP ontology is valuable only if it is endorsed by its user base. These users would have a history with other management tools and software. Therefore, providing a simple, user-friendly, logical interface for medical professionals could be a challenge.

1.3 Research Objectives

The objectives of this research include:

1) The creation of a generic CP Ontology that can easily be extended to represent a hospital and disease specific CP. This generic CP ontology will be extended to represent an Ischemic Stroke CP. It should describe both conceptual knowledge about stroke treatment artifacts (such as literature and assistive tools), and detailed treatment plans for ischemic strokes (such as activities and outcomes).
2) The generic CP ontology should also incorporate SNOMED-CT for the standardization of any medical knowledge associated with a hospital or disease specific CP. SNOMED-CT should be integrated in a logical and semantically rich fashion.
3) Furthermore, this CP ontology should adopt a consistent and correct temporal system such that the timing of any CP is unambiguous.
4) The use of the proposed ontology with a corresponding management system would allow generation of CP-tailored medical records for stroke patients, on which data analytics can be easily performed.
5) These data analytics should provide inferences and insights that would not be possible using paper-based CP or other types of representation
6) The proposed ontological model should provide a mechanism for CP semantic interoperability, thus allowing different medical communities to exchange their CPs and learn from each other’s CP treatment experiences.

1.4 Applications and Benefits

Benefits of this research include standardized coordination between healthcare environments; increased patient understanding of care processes, which is especially important considering the movement towards outpatient care; reduction in LOS; fraud detection and increased potential for e-Healthcare (web-based medicine). If a consistent, correct CP ontology model is developed and applied in a healthcare setting, issues arising from redundancy, ambiguity and incorrectness should be theoretically minimized.

1.5 Thesis Structure

The structure of this thesis will be as follows. In Section 2, pertinent background topics will be discussed such as Clinical Pathways, Ontology and SNOMED-CT. This will be followed by a general literature review of other research performed in CP modeling, stroke modelling and standardization. In Section 3, the details of CP modeling techniques in this research will be explained. Section 4 will cover the experimental results of the CP Ontology as well as applicable data analytic techniques. Finally, Section 5 will summarize the completed research and briefly discuss future work that should be completed.

2 Background and Related Work

2.1 Clinical Pathway

Clinical Pathways are implementations of medical guidelines in a specific healthcare environment [18]. This may include hospitals, clinics or other healthcare facilities. They represent an evidence-based patient care workflow for a specific disease [6]. In general, CPs can be considered a set of best practices for the successful treatment of a patient suffering from a disease. They are a widely accepted tool for clinical management that detail expected actions, progress and outcomes during the healthcare of a patient [7]. The adoption of CPs allows for easier continuity of care across different healthcare settings and medical teams within a hospital or other healthcare facility. This is a major benefit of CP in reducing patient LOS. Furthermore, the provincial, national or global acceptance of standardized CPs would allow for improved patient care across hospitals.

First introduced in the 1980's, the use of CP as part of standard patient care has grown considerably in the past decades [19]; however, these CP documents remain mostly paper-based. Digitalization of CP is the next obstacle in the standardization and improvement of healthcare. CPs that are digitally visible and machine readable would allow for even greater standardization. One proposed method for the digitization and standardization of CP is the use of ontology and ontological modelling.

2.2 Ontology

2.2.1 Semantic Modeling and Basic Structure

Ontology is a model that represents a certain subject matter [9]. It not only communicates what things exist in a certain domain or field, but also how those things relate to each other. Ontology can be used to represent any type of domain, field or subject matter. This can include topics such as education, food, healthcare, finance and natural sciences. In this research, ontology is required to express the medical field, specifically Clinical Pathways. Representing CP as ontology would allow for the sharing and reuse of domain knowledge [20], since ontology are machine understandable when compared to paper-based documents. The many concepts and inter-relationships of Clinical Pathways will be modelled in a semantically rich way using the power of ontology. Ontology can be built using the Web Ontology Language (OWL).

2.2.2 Web Ontology Language (OWL)

OWL is a family of related knowledge representation languages. It is built upon the W3C XML standard for objects known as the resource description framework (RDF) [21]. RDF is the main syntax for OWL. At the core of RDF is the concept of RDF triples. Any expression made in RDF is a triple. It has a structure containing a: a) subject, b) predicate and c) object. The predicate indicates a relationship between the subject and object. If the predicate does in fact hold between the subject and object, the triple has been asserted [22]. This relationship is shown in Figure 2.

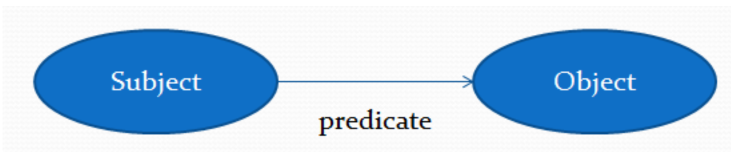

Figure 2: A generic RDF triple.

Each component of an RDF triple is assigned a Uniform Resource Identifier (URI). This URI uniquely identifies each part of a triple even when it may be distributed across a network, or more commonly the internet. OWL adopts a very similar identification framework except that it uses Internationalized Resource Identifiers (IRI). IRIs offer more flexible naming conventions as they can draw from a larger valid character set. An example of a valid IRI used as part of OWL is: www.w3.org/2006/time#TemporalEntity. This IRI denotes a class known as Temporal Entity in the OWL-Time ontology, which will be discussed in greater detail in Section 3.4. All web-published ontology will have a unique base IRI, which identifies the location at which the ontology can be found.

The Web Ontology Language release adopted for this research is known as OWL 2 and it is the current version of OWL. The major entities of OWL are Classes, Individuals, Object Properties and Data Properties. Together, these entities compose triples.

2.2.2.1 OWL Entities

OWL classes provide an abstraction mechanism for grouping concepts with similar characteristics. OWL Individuals are instances of these classes. In a proposed CP ontology, a relevant class may be ‘Patient’ or ‘Patient CP Education’. Furthermore, an individual of the class ‘Patient’ may be ‘Jon Jones’, while an individual of class ‘Patient CP Education’ may be ‘A Guide for People Living with Stroke’.

OWL properties consist of two groups, object properties and data properties. Object properties link one class/individual to another class/individual. They describe the relationship between these two classes or individuals. An object property will have a domain and a range, equivalent to the RDF concepts of subject and object respectively. One class will function as the domain, while the other will function as the range. To extend the previous example, the object property 'provided Education' could relate 'Jon Jones' (the domain) to 'A Guide for People Living with Stroke' (the range).

Alternatively, data properties link a class/individual to a data value. The domain and range structure still applies to data properties. In this case, the domain is the individual, while the range is the data value. Data values are the literal data belonging to a certain date type. These data types can belong to a number of data schemas defined in XML, known as XML schema definition (XSD). Some examples include xsd:int, xsd:date or xsd:string. Completing the previous example, the data property 'DOB' can be created between individual 'Jon Jones' and data value 'July 19 1987' of data type xsd:date. This example is documented in Figure 3 below.

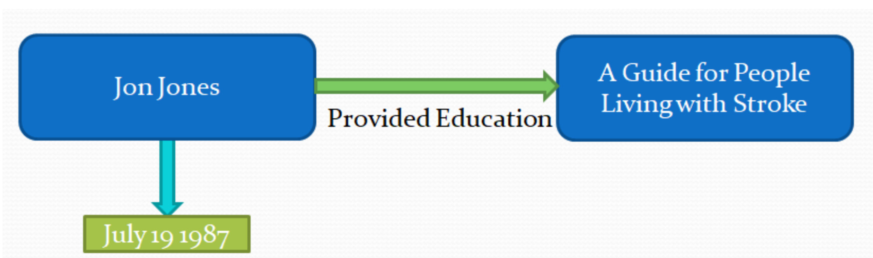

Figure 3: A potential CP scenario modelled using the components of OWL 2.

2.2.3 Allen's Algebra

Properly modelling temporal knowledge, constraints and relationships is crucial to developing an ontology-based Clinical Pathway. The proper timing of events in the execution of a Clinical Pathway, such as tests, medication administration, and consultations are fundamental to its successful execution. It is not only important to represent current events, but also accurately represent historical events and information. As a result, it is necessary to adopt a strong approach in modelling time. Temporal modelling in this research was heavily influenced by the work of James F. Allen, specifically his research on time points, time intervals and non-ambiguous temporal inferences in his development of Allen's Algebra,

Several key considerations were made by Allen in his research. These include that temporal information is often relative; the exact relationship between two times is not always explicitly known; temporal knowledge varies in granularity and that temporal events should be persisted. One example of persistence is that if a patient is moved to a different medical unit on June 1st 2000 at 3:00 pm EST, they should still be in this unit at 3:30 pm EST (assuming no other actions have been taken). These considerations should also be accounted for within this research. Before the work of Allen, temporal research was focused in four main categories: State Space Approaches, Date Line Systems, Before-After Chaining and Formal Models. These main groups will be summarized below.

State Space solutions can express time in simple situations [14]. The state is a representation of the current environment at a point in time. It is a collection of facts that is updated by a specific action. Once an action has occurred, new facts are added and pre-existing facts may be deleted if they are no longer true. This method can be effective but it is often expensive to retain all previous states [10]. Therefore, the concept of persistence is not followed.

Dateline systems assign a date to each fact [15]. The format of this date can vary, but in general, it is straight forward to apply computations to these dates and determine the temporal ordering of a series of facts. Problems arise in this method if an exact date is not known for a fact. It then becomes impossible to determine whether one fact precedes another.

The method of Before-After Chaining allows temporal events to be directly linked together [10]. This method is effective but becomes computationally expensive as the size of the chain increases. However, the work of Allen can be considered an extension and improvement of the Before-After Chaining methodology [10].

The Formal Methods proposed before the work of Allen focus mostly on temporal information described as an instantaneous, point-based situation [16]. In these approaches, only one event can occur at a time and no knowledge can be inferred about the transitions between situations [10]. It is the aim of this research to avoid these 4 types of temporal knowledge systems, even where a simple representation may be convenient, and instead represent the temporal information of CP in a semantically rich way.

In natural language, it is common to refer to time as both instants and intervals. For example, in the domain of healthcare, a patient might describe chest pain as occurring after their favourite television show but before bedtime. In this case, the words ‘after’ and ‘before’ indicate a certain interval of time in which the chest pain could have occurred (likely late evening). Alternatively, a patient could describe their chest pain as occurring at noon. This description would indicate a point-based, instance of the occurrence. However, the granularity of such a description can always be increased. Occurring at noon may indicate the event occurred precisely at 12:00:00 pm or within a certain amounts of minutes around 12:00:00 pm. It is due to this varying level of granularity that Allen's focus is on time intervals and furthermore, on time instances as 'very small' intervals [10].

The concept of a time interval is central to Allen's Algebra [10]. Time intervals are modelled through expressing their endpoints. Therefore, an interval is an ordered pair of points with the first time point being less than the second time point [10]. This approach is expressed using a plus-minus notation. For example, an interval 'I' will have its first (lesser) endpoint denoted as 'I-' and its second endpoint (greater) denoted as 'I+'. This is shown in Figure 4 below. Relationships can also be defined between time intervals, which are fundamental components in Allen's algebra and allow for strong temporal reasoning.

Figure 4: A standard time interval with endpoints denoted as '-' and '+'.

Considering two intervals, t and s, the relationships between their endpoints are described in the table below. Of these relationships, 'during' is the most significant in temporal reasoning. Extending the example of chest pain above, if a patient describes chest pain as occurring during their favourite television program and that television program occurs during 7:00 pm to 8:00 pm EST, it can be established that the chest pain occurred during 7:00 pm and 8:00 pm. Therefore, this 'during' relationship, allows an inheritance hierarchy between intervals [10].

Table 1: The endpoint relationships that exist between interval t and s.

Interval Relation	**Equivalent Relations On Endpoints**
t before s	t+ < s-
t equals s	(t- = s-) & (t+ = s+)
t overlaps s	(t- < s-) & (t+ > s-) & (t+ < s+)
t meets s	t+ = s-
t during s	((t- > s-) & (t+ =< s+)) or ((t- >= s-) & (t+ < s+))

Considering the relationships established between endpoints above, as well as the 'starts' relationship, 'finishes' relationship and each relationship's inverse, there are 13 major relationships that are defined between intervals. These 13 relationships are denoted in the Figure below and are at the core of the temporal system proposed in this research.

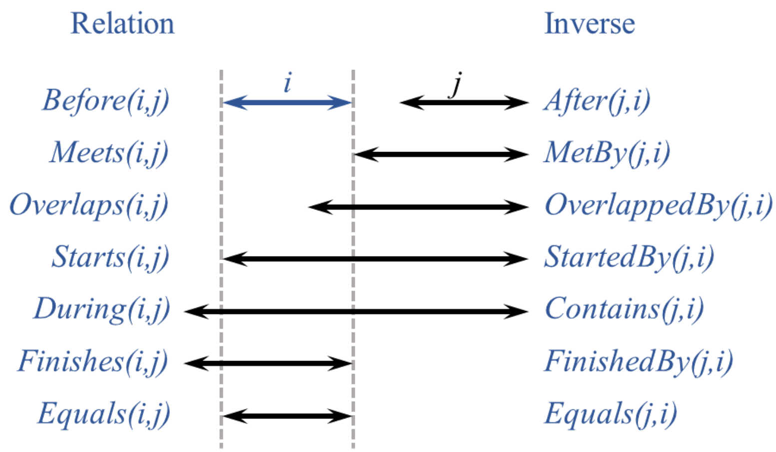

Figure 5: The 13 major relationships between temporal intervals [8].

In Allen's approach, when a new interval is added, new relationships may be asserted between intervals. For example, if a patient experiences arm pain during a physical examination and the physical examination happens before day 2 of a patient's hospital stay, then the relationship that arm pain occurred before day 2 of a patient's stay can be inferred. This relationship is established even without explicitly denoting that the arm pain occurred before day 2. This is a crucial benefit in CP modelling, where temporal relationships between specific events may be expressed without explicitly placing these events on a larger timeline. Using these techniques, each event's temporal location can be effectively inferred.

The application of Allen's Algebra in an ontological format is accomplished through the OWL-Time ontology. This ontology contains the 13 fundamental temporal relationships as well as many other temporal tools. Therefore, OWL-Time is used, and extended, in this research to model temporal knowledge in Clinical Pathways. A summary of OWL-Time will be provided in Section 3.4.

2.2.4 Semantic Web Rule Language (SWRL)

While ontology is an effective method to model a domain and express a multitude of inter-relationships, there is also a need to codify a rule set for a domain. This rule set should be able to infer new knowledge dynamically as an ontological model is updated. Specifically, during the execution of CP ontology, new knowledge should be inferred as events occur. As actions are performed during the progression of a CP, new individuals and relationships may be created. These new additions may have a cascading effect on the rule set in that multiple rules are 'fired' and additional relationships are created. The Semantic Web Rule Language (SWRL) is the rule language selected in this research to model CP-specific logic.

SWRL is a combination of the OWL and the Rule Markup Language (RuleML). The format of SWRL consists of an antecedent and consequent, where whenever the antecedent holds (is true), the consequent also holds. Both the antecedent and consequent can consist of multiple atoms. The general structure of SWRL is therefore as follows:

$$Antecedent \rightarrow Consequent$$

$$Atom\ 1\ \wedge\ Atom\ 2\ \wedge\ Atom\ 3 \rightarrow Atom\ 4\ \wedge\ Atom\ 5$$

The atom notation of SWRL varies for classes, object properties and data properties. An OWL class denoted in SWRL takes the form:

$$C(?x)\ where\ ?x\ \equiv a\ class\ variable\ or\ individual\ name$$

An OWL object property in SWRL takes the form:

$$O(?x, ?y)\ where\ ?y\ \equiv a\ class\ variable\ or\ individual\ name$$

An OWL data property in SWRL takes the form:

$$D(?x, ?v)\ where\ ?v \equiv a\ data\ value\ variable\ or\ literal\ data\ value$$

Using SWRL, new knowledge can be added to ontology. Anytime an antecedent holds, the atoms of the consequent will be created within the ontology. This results in the creation of new object or data properties, which in turn may cause additional antecedents to become true.

An example of a concrete SWRL statement is shown below. Its purpose is to infer if a certain person has a nephew. If three individuals of class Person exist (x, y and z), person x has a brother y and person y has a son z; therefore, person x has a nephew z. This rule is able to infer additional domain knowledge when not explicitly stated. This ability is crucial in the development of CP ontology, where certain relationships may not be stated explicitly but still exist. An example of this may be where a certain event occurs past the normal LOS for a CP. In this situation, the 'during' relationship between this event and a certain day might be explicitly created; however, a SWRL statement would have to be created that infers that any events occurring during that day are 'overdue' or 'late'.

$$Person(?x)\ \wedge\ Person(?y) \wedge\ Person(?z)\ \wedge\ hasBrother(?x, ?y)\ \wedge\ hasSon(?y, ?z)$$
$$\rightarrow hasNephew(?x, ?z)$$

2.2.5 Protégé

Protégé was selected as the ontology building software for this research. Protégé is an open-source framework and ontology editor developed by the Stanford Center for Biomedical Informatics Research at the Stanford University School of Medicine [5]. It includes full support for OWL 2, and connections to several description logic reasoners. Furthermore, Protégé offers a customizable UI with tools for axiom, class, individual, object property and data property creation. Additional features include ontology merging, visualization and SWRL development. These features made Protégé a suitable choice for ontology creation during this research.

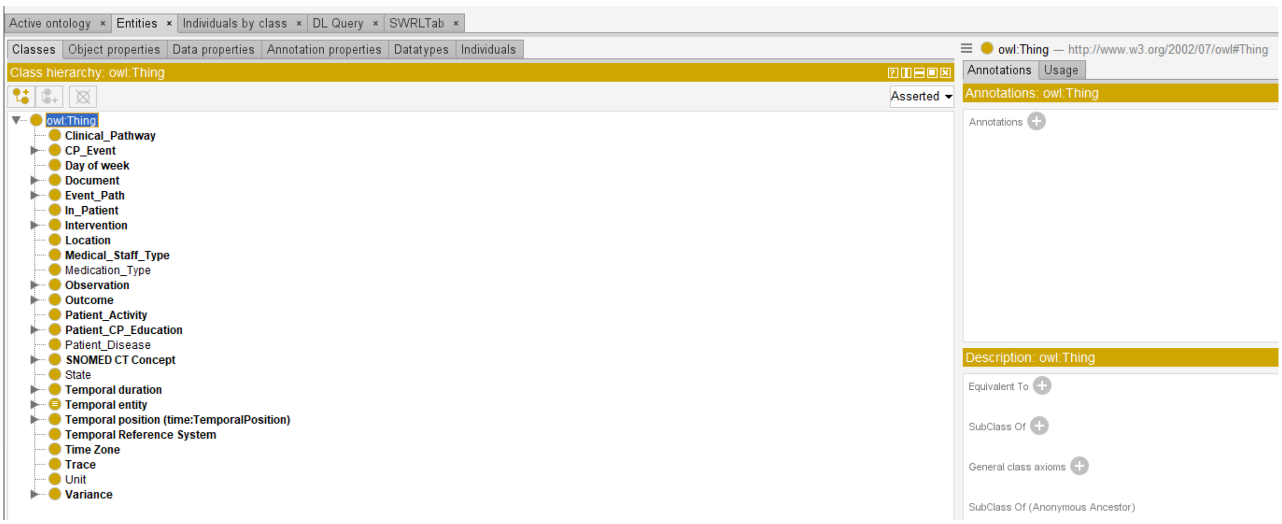

Figure 6: The main GUI for Protégé, with a hierarchical view of created classes.

2.3 Systematized Nomenclature of Medicine – Clinical Terms (SNOMED – CT)

In the development of a generic CP ontology, medical knowledge must be expressed in a standardized, widely recognized format. Currently, CPs do not follow a standardized terminology system in expressing medical terms. The nomenclature paradigm selected in this research to accomplish this goal is the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT). SNOMED CT is the most comprehensive, multilingual clinical healthcare terminology in the world. It contains comprehensive, scientifically validated clinical content. Furthermore, it enables consistent representation of clinical content in electronic health records. Finally, it is used in more than eighty countries and mapped to other international standards [23]. Due to these factors, SNOMED CT is a suitable option to represent medical concepts and knowledge in CP ontology.

A concept in SNOMED CT consists of a standardized term and ID. The ID is a numeric code uniquely identifying that SNOMED CT concept. An example of a SNOMED CT term and ID is shown in Figure 7 below. It denotes a Clinical Finding with the ID 404684003. Using SNOMED CT, a paper-based CP can undergo standardization for many of the written terms.

404684003 |Clinical finding (finding)|

Figure 7: The ID and term name for the concept Clinical Finding.

2.4 Literature Review

Extensive research has been conducted in the field of semantic modelling of clinical pathways. This research spans the past decade and covers concepts in knowledge modelling, real-time updates, task handling and temporal ordering.

Research on the implementation of CP for urological operations was performed through 1997 and 1998 by Chang et al. at Chang Gung Memorial Hospital. After implementation, the results were compared to urological operations performed in the previous years (April 1996 to March 1997), which did not adhere to clinical pathways. It was found that patient's undergoing urological operations post-implementation of CP experienced a reduction in LOS. LOS was reduced from 5.5 days to 4.9 days, with a p-score less than 0.01 [25]. This indicates less than a 1% chance that the LOS remained 5.5 days, given this observation. Furthermore, the rate of surgical complications was also found to be reduced [25]. While this research indicates following CP can lead to a reduction in LOS, there were also variations from the implemented urological operation CPs in 39.3% of cases [25]. The frequency of variation from these CP displays the importance of properly handling variance during CP execution. The ability to document and adjust to variance during CP execution should be a focus of this research.

Focused on creating more standardized CP management, with emphasis placed on tracking and managing variance, Wakamiya and Yamauchi deployed a new system into pre-existing hospital management systems [26]. The benefits of this system include low development cost, free access through the internet and automatic analysis of variance. At present, no issues have been reported on this software [26]. Wakamiya's and Yamauchi's focus on developing a low-cost, open-access and variance-aware system serves as a benefit of this research.

While the research performed by both Chang as well as Wakamiya focused on the implementation and standardization of CP, neither considered the use of ontology. Conversely, the research of Ye et al. proposes ontology for the modelling of CP workflows. A workflow model will usually consist of a number of activities and their dependencies [30]; however, the model proposed by Ye et al. must also represent knowledge on patient outcomes, resource use, variance and time constraints. The overall architecture recommended by this research is a Clinical Pathway Ontology (CPO) as a common semantic foundation [30]. The CPO is built in part using a process-based ontology (OWL-S) and “subontology of time” [31] for temporal modelling.

Afandi et al. focus on the development of ontology for stroke rehabilitation. In the rehabilitation process for the Hospital Universiti Sains Malaysia (HUSM), patient assessment records and other information are stored in a Patient Information System (PIS), which has a relational database structure. The PIS draws from various sources and is potentially inconsistent. This may lead to the completion of assessments that are not important for the patient [27]. Afandi et al. argue that sematic descriptions of assessments are only available in the PIS when queries are made. Otherwise, the schema of the data may not be obvious. They consider ontology as an alternative to PIS relational databases when storing patient information.

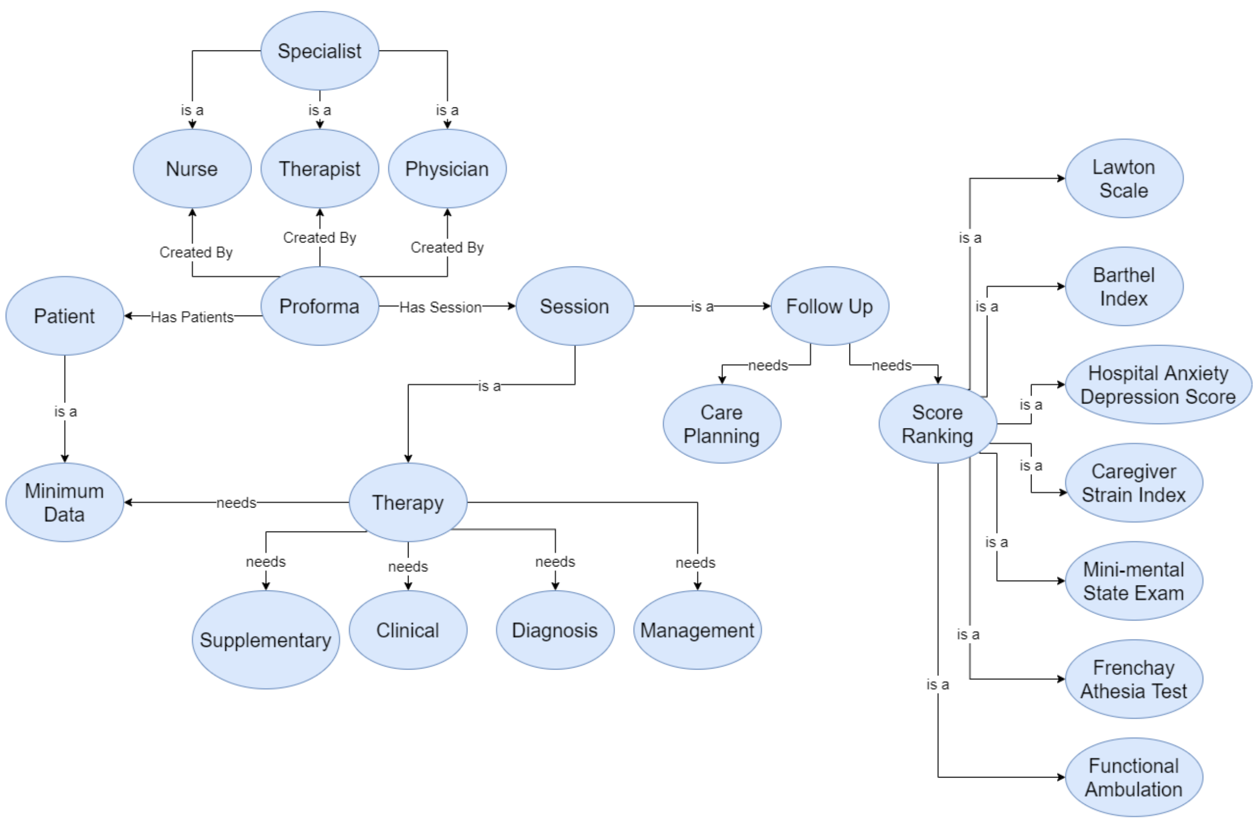

Figure 8: The ontology of [27].

During elicitation of ontology components and requirement analysis, this research considers 10 key objects of HUSM. These objects are Physician, Therapist, Nurse, Minimum Data, Clinical, Diagnosis, Management, Supplementary, Care Planning and Score Ranking. Additional work was performed to group and capture these major objects in the stroke rehabilitation ontology. The common characteristics of these objects were considered and generalizations where made to construct ontology classes and class relationships (object and data properties). From this further analysis, six main classes were developed: Specialist, Patient, Proforma, Session, Therapy, and Follow-up. The ontological model proposed by Afandi et al. is shown in Figure 27 above.

These classes highlight the importance of both Therapy and Specialists in the rehabilitation process. As patient rehabilitation progresses at HUSM, weekly assessments, therapy and follow-ups are conducted by specialists on recovering patients. Since limb disability is a common characteristic of stroke survivors, it is crucial to be able to model the role of therapy and specialist groups in rehabilitation ontology. Furthermore, when considering larger stroke ontology, it remains important to properly express the roles of assessments, therapy and specialists.

A centralized class of the rehabilitation ontology is Proforma. The Proforma can be considered the rehabilitation process as a whole. It is created by specialists (object property created By), has patients (object property has Patients) and has sessions (object property has Session). During the first day of the rehabilitation process (proforma), initial patient data is loaded into the ontology. As sessions and therapies are completed, additional data pertaining to these sessions are added to the ontology. This ontology is implemented for users through a PHP and JavaScript based interface. Validation and evaluation of the prototype system will be performed by domain experts and 10 specialists at HUSM [27]. It is important to consider that no temporal aspects are explicitly modelled by this research. It is possible that temporal knowledge will be added in future iterations of the ontology; however, this is currently a weakness of the rehabilitation ontology.

In the development of the Neurological Disease Ontology (NDO), Jensen et al. focused on the clear and unambiguous annotation of data for Alzheimer’s disease, Multiple Sclerosis and stroke [28]. Of special interest to this thesis, is the research performed on stroke. The NDO was developed using the Ontology for General Medical Science (OGMS), which contains terms describing core concepts of medicine. A main objective of the NDO is extension of the OGMS classes Disease and Disease Course to neurological diseases, such that a domain-encompassing ontology is developed. In OGMS, a Disease is defined as "A disposition to undergo pathological processes that exists in an organism because of one or more disorders in that organism." [28]. NDO extends this class by defining child classes as members of their parent class with a differentiating criterion. As an example, Neurodegenerative Disease is "A neurological disease that is characterized by atrophy or death of neurons or related structures progressively affecting the functioning of the nervous system". This process of definition is a useful methodology to develop a single hierarchy of semantically linked diseases.

Another contribution of this research is the development of relationships between diseases, syndromes and disease courses; specifically, a disease course being the realization of a disease. This manifestation of a disease, expressed through disease course, involves many related ontology entities such as body processes, infections and disorders. Furthermore, a disease may result in a syndrome. These syndromes, usually presented as clinical observations, do not necessarily result from the same underlying disease. For example, this allows a syndrome such as dementia, although often stemming from Alzheimer's, to exist independently from the Alzheimer's disease. This approach of expressing disease as not only a stand-alone concept, but as a grouping of disease, disease course and syndrome is important in increasing the semantic meaning and overall flexibility of diseases in the context of ontology. A concrete example of this is the creation of different disease types for Multiple Sclerosis (MS) in other terminologies. In the NDO, based on background research, it is more appropriate to represent these unique MS diseases as components in the disease course for MS [28].

While extensive in its development of neurological disease, the NDO does not consider the patient progression through a specific neurological disease. Therefore, this ontology lacks important aspects of a greater CP ontology including the progression of time, the occurrence of events and the activities of both medical staff and patients. These shortcomings should be considered when developing CP ontology, while not ignoring the widespread value of NDO's disease structuring.

The work of Liu et al. focuses on the modeling of clinical pathways as well as monitoring their execution. It argues for real-time monitoring of CP, instead of just post-execution monitoring based on different CP outcomes. Its real-time monitoring approach of CP is realized using a Clinical Data Ontology, Clinical Knowledge Ontology and SWRL Ontology [36]. Together, these ontologies allow quick response to deviations in the execution of a CP. This approach was adopted because of its encapsulation of data, knowledge and rules separately; its abstract, general design; and its focus on real-time application.

Some challenges that Liu et al. faced were 1) retrieving up-to-date medical data for patients; 2) representing medical knowledge in a semantically rich formalism; 3) Storing the retrieved data accurately; 4) abstracting and representing practice oriented knowledge from paper based CP; 5) and monitoring CP execution in real-time [36].

This research proposes receiving up-to-date data from an EMR passively, allowing the EMR to 'push' new data as it receives it. The proposed ontological model for this research is depicted in Figure 9 below. Its purpose is to represent the domain of CP in an abstract and generic fashion. The main classes of this ontology are Trace, Event and Intervention. A Trace records the actual treatment behaviours in CP execution. An Event is connected to certain medical knowledge and Intervention represents a certain medical activity [37]. Another key class is Observation. An observation will arise from some Intervention. Also considered in this ontological model are the concepts of Date, LOS and Time. These limited temporal classes are used to assign admission and discharge times as well as LOS.

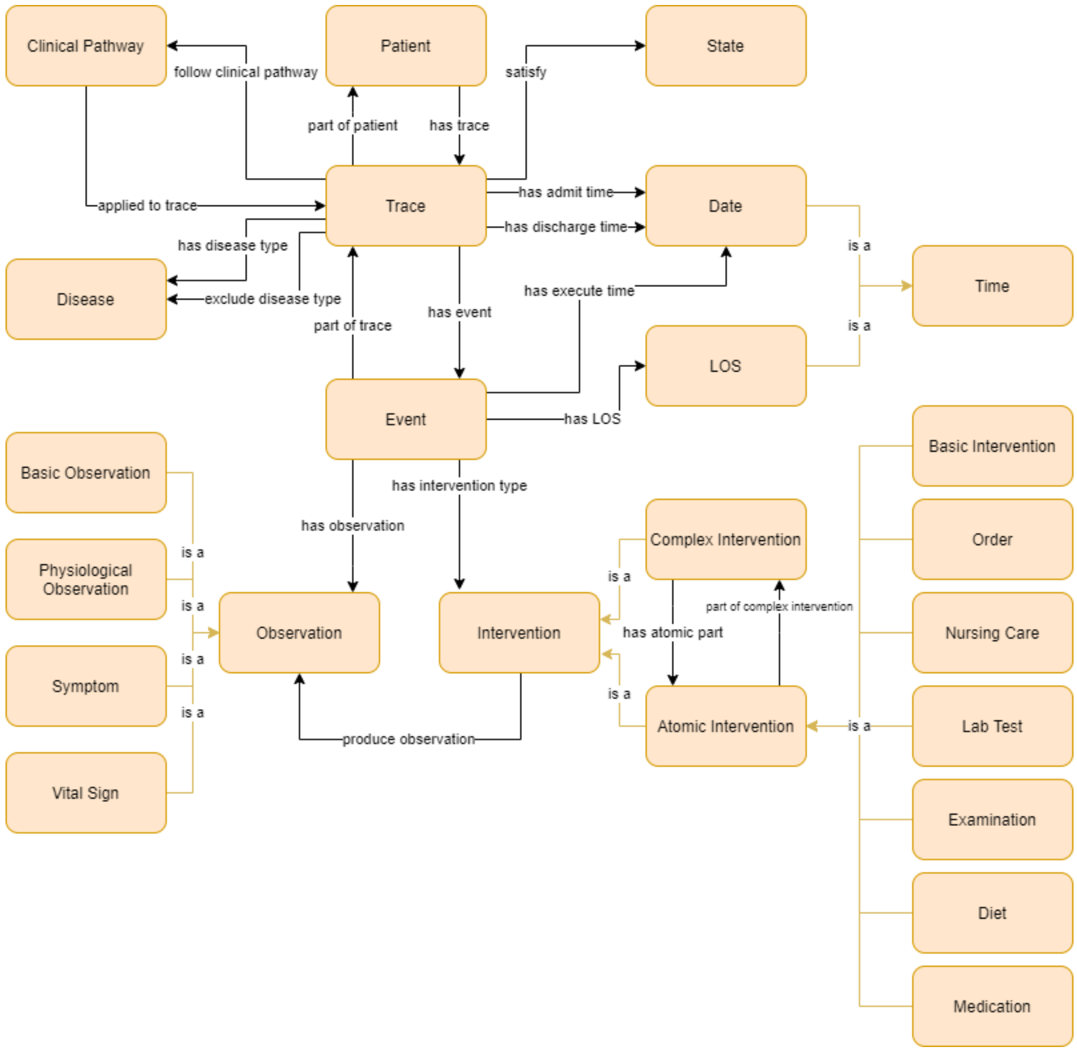

Figure 9: The ontology of [36] with major classes trace, event, and intervention.

To determine a patient's state within a particular trace, Liu et al. uses SWRL statements. These SWRL statements will contain the atoms that a state should satisfy within the antecedent and an atom (object property) assigning this state to a trace in the consequent. In general, "a state or condition will be satisfied only when the state or condition can be inferred and connected to the trace" by the object property of the consequent [38]. Specific sets of SWRL statements can be developed for different CP. Individual sets can then be used by the Clinical Data Ontology to extend that ontology for use in different CP.

By using an ontological approach, this research can describe both CP domain concepts and their interrelationships in a generic way. This generic approach allows for reusability and extensibility of the original design. The use of SWRL statements for disease-specific CP further enhances the generic nature of this design.

During the progression of a CP, there is a need to continuously determine patient condition. This is a necessary step that will aid in decision making and selection of the most appropriate future interventions. As a patient's characteristics and condition will evolve over time, the progression of CP interventions should also be dynamic in order to offer the most appropriate treatment process. The research of Fan et al. focuses on the representation, determination and analyzation of underlying patient state in the integrated treatment process.

Fan et al. proposes emphasis of several crucial medical activities that affect patient state. These activities are admission, pre-operation, operation, post-operation and discharge [33]. This research considers the entire CP to be contained between admission and discharge. A patient will travel through a series of states (S) during the progression of a CP, which are connected through a set of interventions (I). These interventions will serve to send a patient from one state to another. This relationship is depicted in Figure 10.

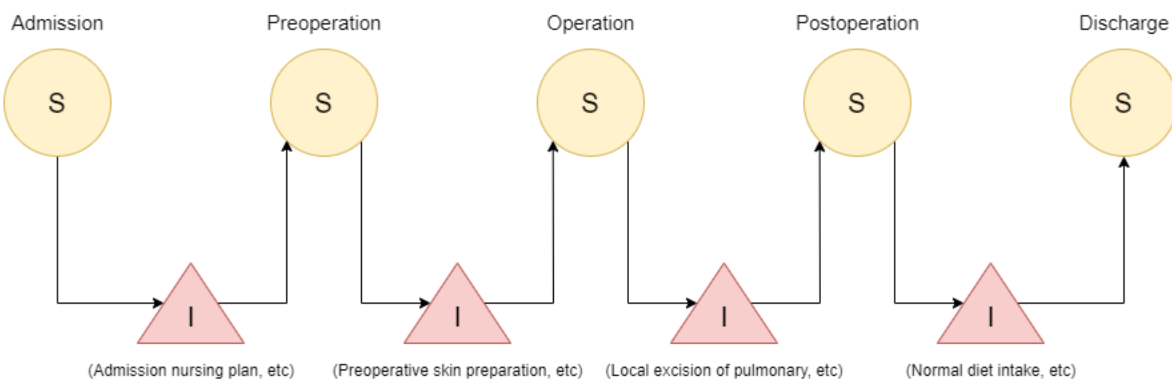

Figure 10: CP State evolution from admission to discharge [34].

In the construction of the CP ontology of Fan et al., there is a focus on the representation of medical knowledge, interventions and temporal constraints. This research recommends the creation of two major components, atomic state base and atomic intervention base, each with several children. Atomic state base gives rise to basic, quantitative, qualitative and miscellaneous state children. Atomic intervention base, gives rise to medication, nursing, physical examination and miscellaneous intervention children. From these bases, new states and interventions can be created. Temporal conditions are represented in this research through the use of Time Point, Time Interval and Time Hybrid [35]. These temporal entities can be assigned on patient states, such that a patient state can occur over an interval or only at a certain instance.

The method of ontological construction for Fan et al. is based on collaboration with medical experts to organize pre-existing medical knowledge and CP knowledge as individuals of either the atomic intervention base class or atomic state base class. Once these individuals are generated, a state individual will be associated to a certain intervention individual through the object property “hasIntervention”. The state individual will also be associated to temporal data through data properties such as “hasIntervalBegin” and “hasIntervalEnd”. The ontology of this research is shown in Figure 11 below.

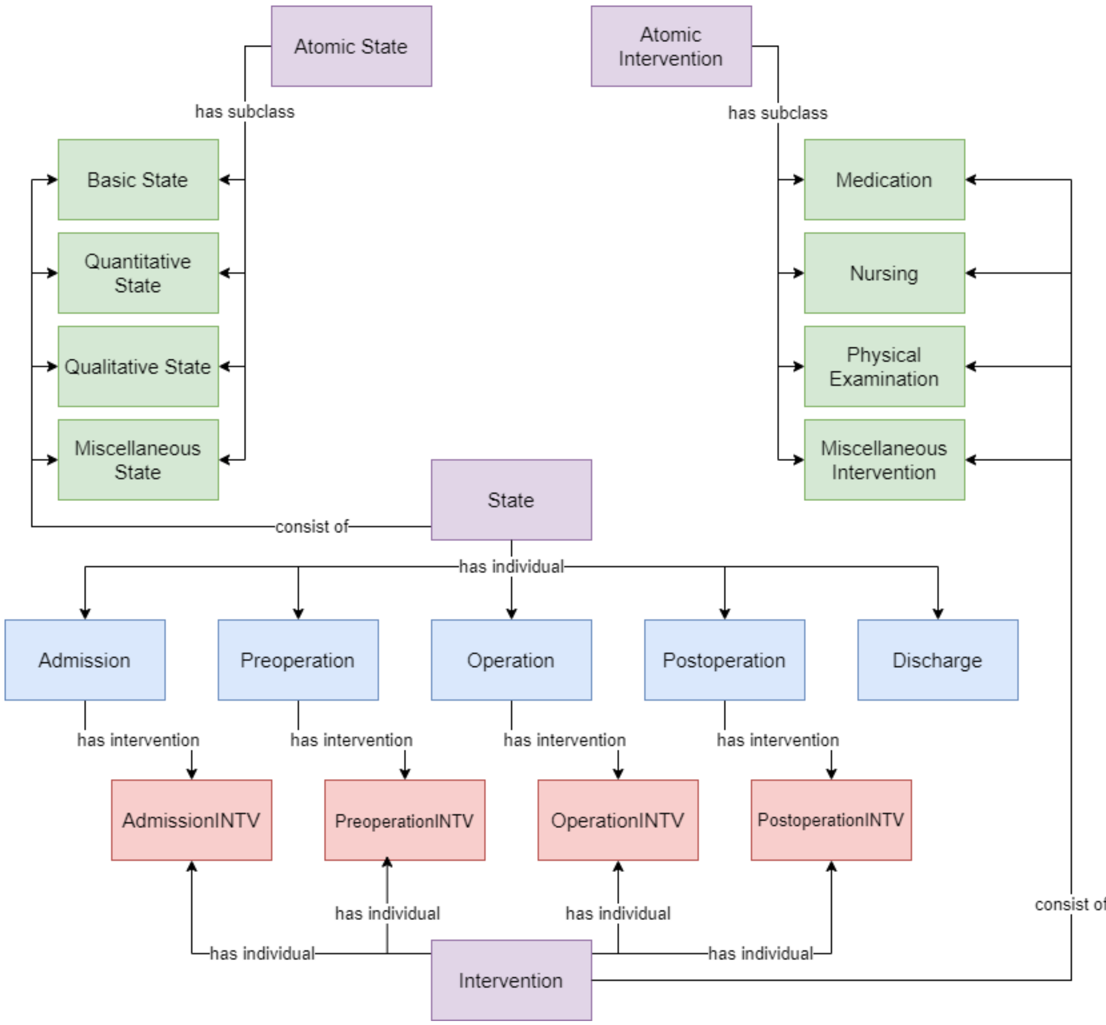

Figure 11: The ontological model of [33].

Representing patient state in this thesis research is an important step towards creating a semantically correct, generic CP Ontology. It is also crucial to accurately modelling temporal restraints as patient state dynamically changes over time. Concepts such as outcomes, observations and patient data can all assist in the proper inference of patient state during CP execution. Furthermore, representation of underlying patient state in this research is a useful tool for increased interoperability of CP, another research objective. Patient states created, influenced or derived from standardized CP interventions, events and outcomes are shareable between healthcare organizations and groups. In the case where a patient arrives at a certain uncommon state, medical professionals would be able to reference the standardized interventions and outcomes that lead to this uncommon state. It would be feasible therefore, for another group of professionals to quickly understand these standardized processes as the other group shares a similarly standardized group of processes. The group being shared with would then be able to check their own patient states derived from the interventions and outcomes in question, and communicate with the original group if such an uncommon patient state has ever been encountered. Through the process of determining patient states based on standardized pre-conditions (interventions, outcomes, patient data, observations, etc.), knowledge sharing may become easier and more effective among different institutions.

Table 2: A summary of the literature review with contributions of this research.

Research Title	**General Findings**	**Deficiencies**	**Research Contribution**
A Holistic Environment for the Design and Execution of Self-adaptive Clinical Pathways [42]	Uses a workflow engine with a SWRL statement base. Encapsulates medical, financial and operational knowledge.	No explicit representation of time except for modelling using "next steps"	Time modelling is considered using the Time Ontology
An Ontology-based Hierarchical Semantic Modeling Approach to Clinical Pathway Workflows [20]	Uses both a temporal ontology and process based ontology. Incorporates Allen's Algebra.	Overly generalized representation of CP. No use of standardization.	Standardization is introduced using SNOMED CT
Ontology-based Clinical Pathways with Semantic Rules [40]	Uses CP Meta Ontology to ensure consistency and stability of model. Uses connection layer between CP and EMR for independence and reusability.	Terminology of the Meta Ontology is not a widely recognized standard	Meta ontology terminology is corroborated using SNOMED CT
An Ontological Modelling Approach to Align Institution Specific Clinical Pathways: Towards Inter-institution Care	The use of merge and branch nodes allows for greater flexibility in modelling Care plans are aligned at a high level	Limited temporal considerations and time-based classes.	Expansion of temporal modelling using the Time Ontology. Tailoring the Time Ontology to the domain of CP.

Standardization [43]	Care tasks are aligned at a lower level		
Ontology-based Computerization of Acute Coronary Syndrome Clinical Guideline for Decision Support in the Emergency Department [44]	Modularization of concepts for reusability. Prioritizes clinical actions based on institutional resources/policy.	No temporal considerations. A greater focus is placed on patient resource management than medical knowledge	Medical knowledge is considered using the Medical Knowledge Ontology.
An Ontology-based Real-time Monitoring Approach to Clinical Pathway [36]	Up to date storage of medical data within the ontology. Use of SWRL. Class Event stores medical knowledge.	Timing mechanisms are unclear. Medical knowledge ontology is small	Standardization and expansion of medical knowledge using the Medical Knowledge Ontology.

In general, the results of the literature review show significant progress in the field of ontological modelling of CP; however, several deficiencies and challenges still exist. Where research was performed in the modelling of CP processes, such as [33] and [38], there is no apparent standardized modelling techniques for medical knowledge. Medical knowledge is represented through some combination of assistance from domain experts and background research. While likely correct, these terms may not be widely recognized or accepted. Where research was performed in modelling of medical knowledge, such as [28], very little consideration is given to modelling of CP processes, patient progression and variance. Finally, even when present in the research of the literature review, temporal modelling is over-simplified and lacking. It becomes difficult to represent complicated temporal knowledge in these conditions. Furthermore, none of the research of the literature review provides user-friendly, logical interfaces for utilization of their proposed models. Where these user interfaces exist they are underdeveloped or difficult to navigate. It is the aim of this research to improve upon these deficiencies and overcome the challenges outlined in Section 1.2.

3 CP Ontology and Semantic Modelling of Clinical Pathways

3.1 Conceptual Modeling of Clinical Pathways

This research proposes a method of modelling CP that draws from three other ontologies. These ontologies are the

1) Meta CP ontology
2) Medical Knowledge ontology
3) Time ontology.

Individually, these ontologies capture CP procedural knowledge (Meta), medical knowledge (Medical) and temporal knowledge (Time). Together these ontologies form a greater CP ontology that is generic, yet extensible for modelling of specific disease CP. A recurring theme of the literature review is the lack of any work to combine all three of these knowledge components. Where one or two of these knowledge bases were captured and modelled, the third was usually missing (or lacking). This architecture allows any clinical pathway to be expressed as a series of activities over a certain time, drawing from a knowledge base.

The Meta CP ontology is inspired by the work of Hu et al. in “Modeling of Clinical Pathway based on Ontology” [40] and Liu et al. in “An Ontology-Based Real-Time Monitoring Approach to Clinical Pathway” and will be covered in Section 3.2. In this research, the Medical Knowledge ontology is based on SNOMED-CT. The medical knowledge base is represented using SNOMED CT concepts for standardization, integrity and consistency. This ontology will be covered in Section 3.3. Finally, the Time ontology is based on the OWL-Time ontology, which will be covered in Section 3.4. In general, the development of the CP ontology is performed in a top-down fashion. This allows for the identification of the most generalized classes and properties.

3.2 Meta Clinical Pathway Ontology

A Meta CP ontology was developed by Hu et al. through analysis of 140 CP in collaboration with domain experts. Its aim is to depict a generic model that is consistent, without redundancies and contradictions [41]. The Meta CP Ontology of this research borrows concepts from both Hu’s Meta CP ontology and the ontology of Liu et al. The semantic structure of the new Meta CP ontology, with additions proposed by this research, can be seen in Figure 30. The major classes of the Meta Clinical Pathway Ontology are:

a) Trace – a trace contains all treatments during a clinical pathway for a specific patient and disease. A patient may have several traces, spanning different diseases;

b) CP Event – a medical event that is part a trace for a patient;

c) Intervention – a medical intervention performed during a CP Event. Interventions can be either atomic or complex. An atomic intervention is typically a single process/action/task, while a complex intervention is composed of atomic interventions;

d) Observation – an observation made during a CP Event or because of an intervention;

e) Outcome – the result of a certain CP Event, dependent on an Observation;

f) State – the position that a CP is currently in, depending on the progression of the patient.

Several changes were made to the Meta CP ontology for use in this research. This includes storing patient information on the Patient class instead of the Trace class, CP Event representing a process/activity instead of certain medical knowledge and completely removing temporal classes in favour of the Time ontology. These changes were made with consideration that a patient's data should be stored directly with a patient, medical knowledge would be represented by the Medical Knowledge ontology of Section 3.3 and that the temporal modelling of the Meta CP ontology was limited. Several additions were also made to the Meta CP ontology.

While the Meta CP ontology is a strong base for ontological modelling of processes, several additions were made to the Meta CP ontology for incorporation into the CP ontology. These additions include the creation of a Variance class, Event Path class and Recommendation class. The Variance class is meant to represent and describe the possible types of variance in CP. Through domain analysis and collaboration with domain experts, the children of Variance were determined to be: Declined by Family, Declined by Patient, Facility Resource Restrictions, Modified based on Medical Evidence and Modified due to Comorbidity.

In collaboration with the Variance class, the Event Path class is designed to denote whether any CP event is part of the normal execution of a CP or has arisen due to variance. The definition of Event Path is shown in Figure 13 and its hierarchy is shown in Figure 14. Two children exist, Common Path and Branch Path. The relationships between CP Event, Variance and Event Path (and their inverses) are shown in Figure 12 below. The object property 'classifies Event' connects Event Path and CP Event, such that an Event Path classifies event CP Event. The inverse of 'classifies Event' is 'has Path Type', such that a CP Event has path type Event Path. The object property 'creates Branch' connects Variance to Branch Path, such that a

Variance creates a Branch Path. The inverse of ‘creates Branch’ is ‘caused By Variance’, such that a Branch Path is caused by a Variance.

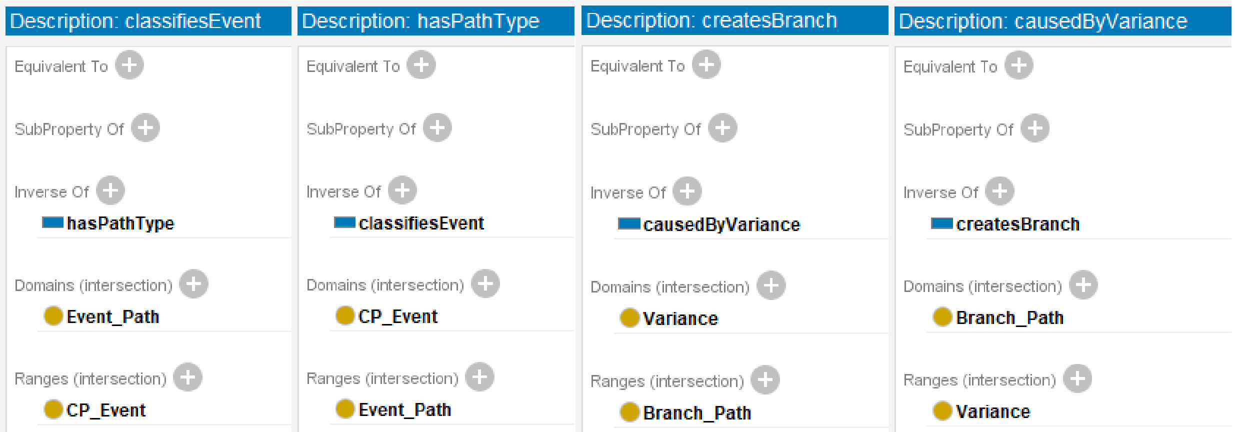

Figure 12: The object properties linking CP Event, Event Path and Variance.

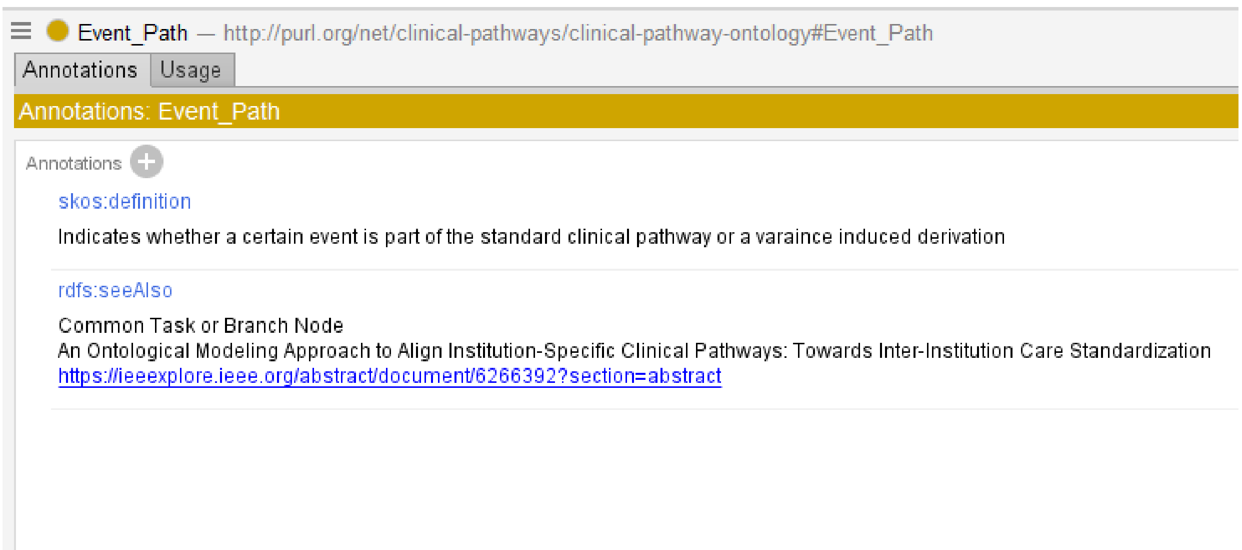

Figure 13: The definition of Event Path with reference to the research of [32] which served as major inspiration.

Figure 14: Event Path with children Branch Path and Common Path.

The Recommendation class was also added to the Meta CP ontology. It was developed to accommodate such interventions that are passive in nature. That is, the interventions that do not have a specific timing associated with them (performed continuously or latently). Through the Recommendation intervention class, instructions that are given in a CP that are temporally ambiguous or continuous can be easily represented. One example of this is the instruction to monitor the IV of a patient. Since this instruction is continual in nature, with no concrete timing, it is considered a Recommendation Intervention. The Recommendation class proposed by this research can be used to classify CP instructions that are intended more as reminders, hence the name, than as concrete instructions to be performed at a specific time or over a specific duration.

Figure 15: The Meta CP ontology.

The major classes, their children, and major inter-relationships of the Meta CP ontology are documented in Table 2 below. This includes a brief description of the children of the Intervention and Observation Class. These child classes are generic enough to represent a wide array of potential CP individuals but can also be further extended depending on the level of granularity needed. An example of greater detail being required would be when extending the CP ontology to create disease specific ontology. This increase of granularity is achieved through either the creation of additional child classes or by the instantiation of more detailed OWL individuals.

Table 3: The major Meta CP ontology classes and their children.

Class Name	Children	Description	Major Properties
Trace	N/A	A collection of all activities in a CP	Follows Clinical Pathway Has Disease Type Satisfies Has Encounter
Intervention	Basic Intervention	Any other type of intervention	Produces Observation Requires Activity
	Doctor Order	General categories of doctor's orders	
	Injection Order	Defines a type of injection	
	Lab Test	Patient lab tests	
	Medication	Medication to be administered	
	Nursing Task	Healthcare tasks performed by nurses	
	Nutrition Diet	Dietary restriction or instruction	
	Physical Examination	Assessment of patient	
	Recommendation	Defined above	
Observation	Basic Observation	Any other type of observation	Has Unit
	Symptom	Departure from normal function	
	Physiological Observation	Other functional observable	
	Vital Sign	Indicates Patient Condition	
CP Event	N/A	Activity that is part of a Trace	Has Intervention Has Observation Has Outcome

			Part of Trace
Outcome	N/A	The result of a certain Event	N/A
Clinical Pathway	N/A	Represents the CP being followed	Applied to Trace

3.3 Medical Knowledge Ontology

The Medical Knowledge (MK) Ontology is a representation of SNOMED-CT. The root class of the ontology is “SNOMED-CT Concept”, which is the same as the root of SNOMED-CT. All other terms in the SNOMED-CT hierarchy are children (subclasses) of this root. A typical class of the MK ontology will be defined similar to the class shown in Figure 16. Its class name will be the standardized SNOMED CT ID (223488008) and it will be labeled with the standardized SNOMED CT Term (Discussion about changes in lifestyle).

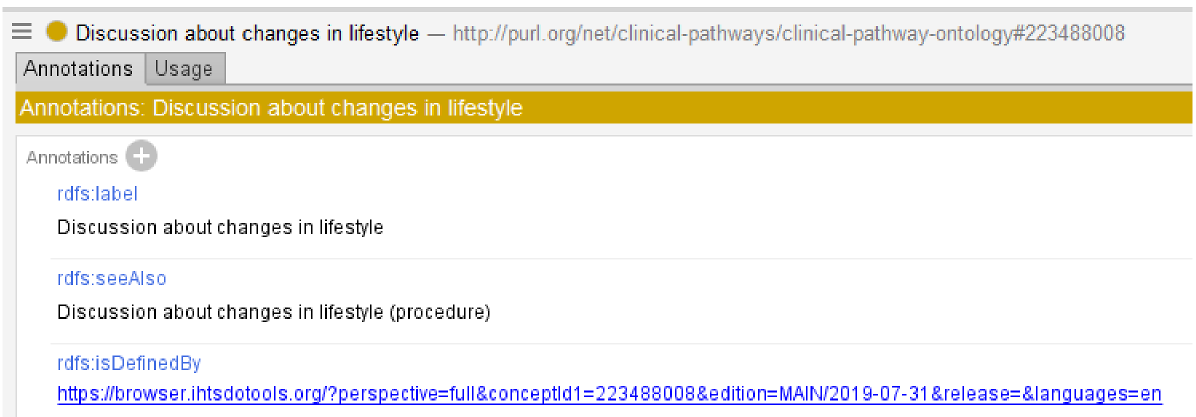

Figure 16: The class definition for a member of the MK ontology.

A major strength of the MK Ontology is this ordering of subclasses, such that a class in the MK ontology is not only denoted by a standardized ID and term but also by its placement in the larger hierarchy. This greatly increases the semantic strength of each MK class. For example, the class “Brushing of Teeth” is a descendant of the class “Procedure on Head” as seen in Figure 17. Therefore, a relationship between the concept of teeth and head is established that would not be explicitly clear if the class “Brushing of Teeth” was isolated from the greater ontological structure. This merit of this approach is two-fold. It can provide additional context to a human user who may be confused by the meaning of a term. It can also allow easier machine-based inferences when new classes (terms) are introduced. To continue the previous example, if a new SNOMED-CT term is added to the MK Ontology that involves ‘teeth’, an inference could be made that this new class is a possible child of “Procedure on Head”.

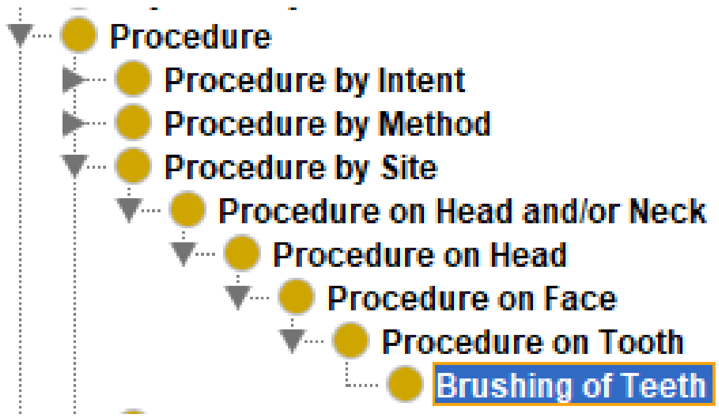

Figure 17: A subsection of the MK ontology representing the ancestry of the class “Brushing of Teeth”.

The MK ontology in its current form is only a subsection of the full SNOMED CT knowledge base. Due to memory considerations, only major SNOMED CT concepts as well as stroke CP specific concepts (during creation of the Stroke CP ontology) were created. However, due to its hierarchical structure, it can be expanded as additional SNOMED CT concepts are needed for reference in other disease CP ontology or when memory constraints are not a concern. These new classes can simply be inserted into the greater hierarchy as necessary. The MK ontology can therefore be grown systematically as paper-based CP terms are investigated and matched to the proper SNOMED CT concept.

Each Intervention and Outcome of the CP ontology will reference the appropriate class/individual of the MK ontology, for standardization purposes. This connection is created using the object property “references SNOMED CT Concept”. This property will have a domain of Intervention or Outcome and a range of “SNOMED CT Concept”, the root class of the MK ontology. Therefore, an Intervention will be standardized to SNOMED CT concepts through its relationship with classes in the MK ontology. This relationship for a specific Intervention is highlighted in Figure 18 below. This intervention, Nicotine Replacement Intervention, is at its core a conversation with a patient about eliminating smoking or other nicotine use. As a result, it is connected to the SNOMED CT concept “Discussion about changes in lifestyle”, which is part of the MK ontology shown in Figure 19.

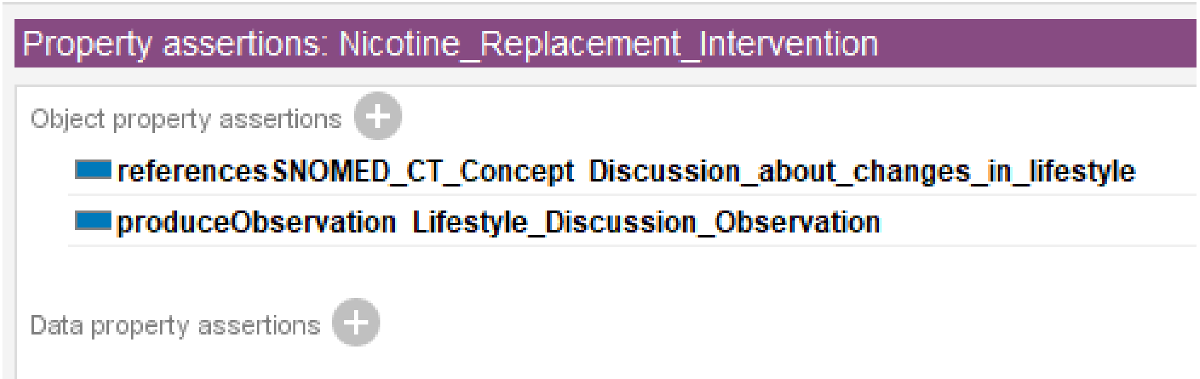

Figure 18: The relationship between Nicotine Replacement Intervention and Discussion about changes in Lifestyle.

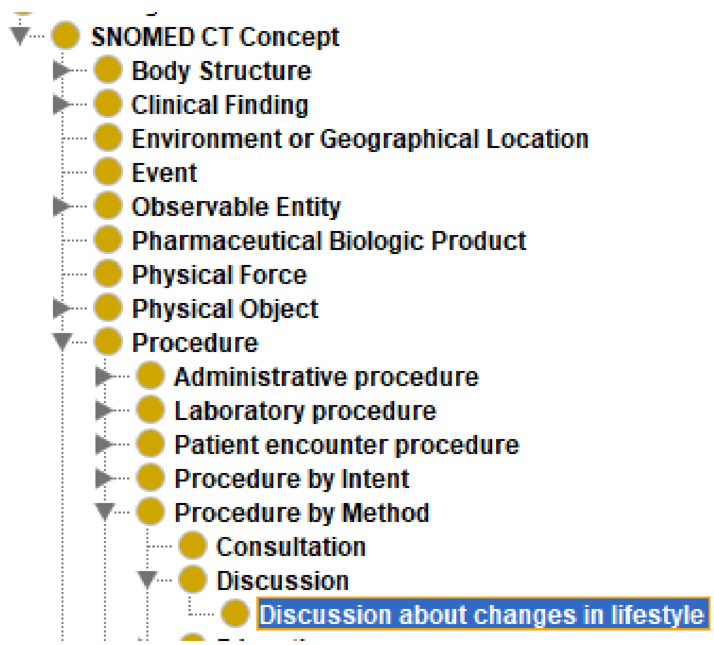

Figure 19: Discussion about changes in lifestyle in the greater MK ontology.

3.4 Time Ontology

The Time Ontology, formally known as OWL-Time, is an ontological model of temporal concepts [24]. It is used in this research to model temporal aspects, relations and constraints in Clinical Pathways. OWL-Time represents temporal intervals based on Allen's Algebra and the corresponding 13 fundamental interval notations. Moreover, OWL-Time allows the description of time positions, time durations and temporal reference systems in numerous formats. Due to its flexibility, extensibility and strong mathematical foundations, OWL-Time is a valuable semantic tool for thoroughly modelling temporal aspects in clinical pathways. Furthermore, as OWL-Time is a generic ontology for time in any real-world application, it can be extended for use in the domain of CP.

The essential root class of OWL-Time is Temporal Entity. Any individual in the CP ontology that inherits from the OWL-Time ontology is a child of Temporal Entity. More directly, an individual in the CP ontology that is associated in some form with temporal information will, at its highest level, be an individual of the Temporal Entity class. Temporal Entity has two children, Instant and Interval. As in Allen's research, an Interval has two endpoints while an Instant is point-based. Furthermore, an Interval has a child Proper Interval, which is the same as Interval except that its endpoints (beginning and end) are distinct. It is the Proper Interval class that is both the domain and range of the fundamental interval relationships of Allen's Algebra. These relationships, expressed as object properties in OWL-Time, are shown in Figure 20 below. The final child in the Interval hierarchy is Date-time Interval. This Interval type is equivalent to its parent Proper Interval except that its duration (interval length) must be an element of General Date Time Description (a specific day, month, year, etc.). OWL-Time provides the object properties 'has Temporal Duration', 'has Beginning', 'has End' and 'has Time' to allow entities in the OWL-Time ontology to be associated with timing-related entities of other ontologies.

Class diagrams of both Temporal Entity and Temporal Duration are shown in Figure 20 and 21 respectively below. The Temporal Duration class supports the description of duration of any temporal entity. These durations can be expressed using data formats such as simple decimals (2.6 days, 3.0 hours, etc.) or with data properties corresponding to second, minute, hour, day, week, month and year (2 day 1 hour 3 minute for example). The final class of this hierarchy, Duration Description, is the same as General Duration Description except that the data properties of Duration Description are fixed to the Gregorian calendar.

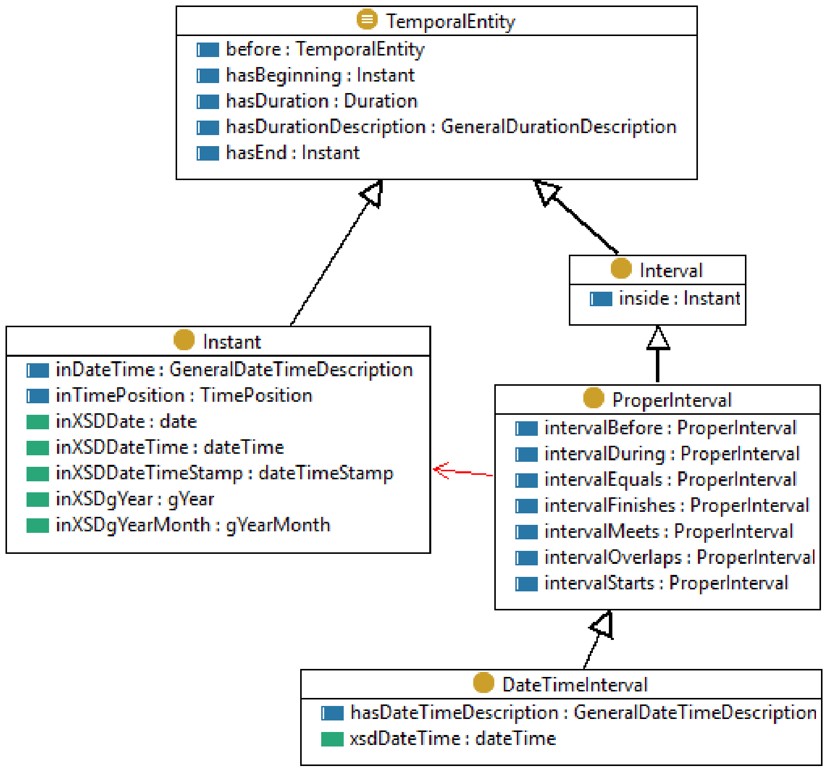

Figure 20: Temporal Entity hierarchy. Interval relations are defined within class Proper Interval [24].

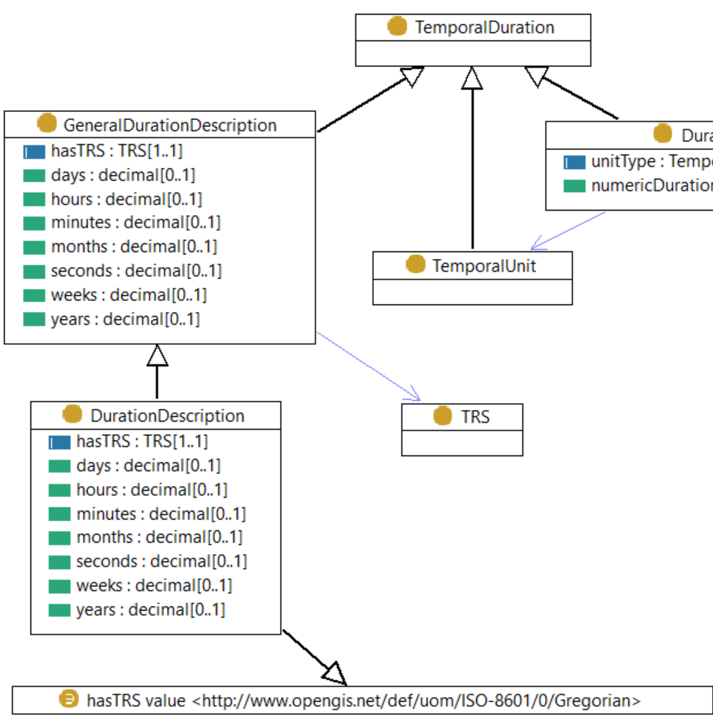

Figure 21: The Temporal Duration hierarchy [24].

As stated above, the Time ontology can be extended for use in CP. These include several additions that were made to the Time ontology for use in concert with the CP ontology. Namely, Day Interval, Trace Interval, Admission Time and Discharge Time were created. These concepts are so important to the temporal modelling of CP, that they were added to the ontology as new OWL classes. While it was also possible to add these concepts as individuals, they were considering generic enough and significant enough to exist as classes within the CP ontology. Admission Time and Discharge Time are children of the Time Instant class, being specific times of entry and exit from the CP. Day Interval and Trace Interval are both children of the Proper Interval class except that Day Interval is also a child of Date-time Interval, while Trace Interval is not. Day Interval is a child of Date-time Interval because a Date-time Interval must have a duration equivalent to a day, month, year or other Gregorian calendar element. This same relation does not exist for a Trace Interval as it can be any arbitrary length of time. The class structure of these additions is shown in Figure 22 below.

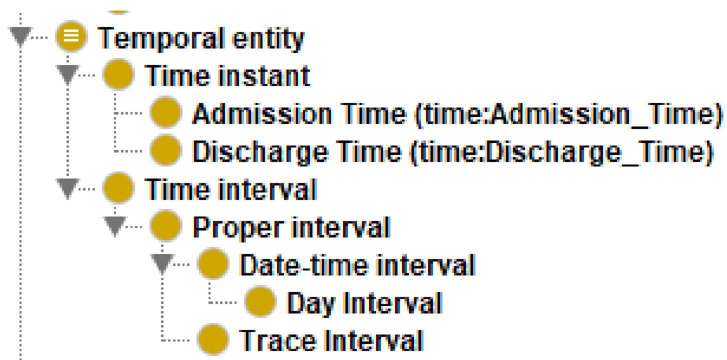

Figure 22: The Temporal Entity hierarchy with Day Interval, Trace Interval, Admission Time and Discharge Time.

Figure 23 below highlights the relationship and temporal ordering of the intervals in the CP ontology. Three interval classes are shown in this figure, Trace Interval, Day Interval and Event Interval. The Trace Interval contains all other temporal components within the CP ontology occurring from Admission Time (the start of the interval) to Discharge Time (the end of the interval). Each Day Interval occurs during the Trace Interval; they are 24 hours in duration and have a Date-Time description of exactly one Day a Year. In general, the duration of any Time Description class will have duration of its Unit Type. In this case, the unit type is Day. Therefore, each Day Interval has the 'meets' object property with the next Day Interval. Finally, any Event Interval (the temporal representation of a CP Event) will occur during a Day Interval. Anyone of the fundamental interval relationships can exist between two Event Intervals. For example, two CP Events can overlap, occur during the exact same time, or one may start the other. This temporal flexibility is a major benefit of the Time Ontology. Using this format, any CP Event can be modelled and placed in the larger temporal structure of the CP trace.

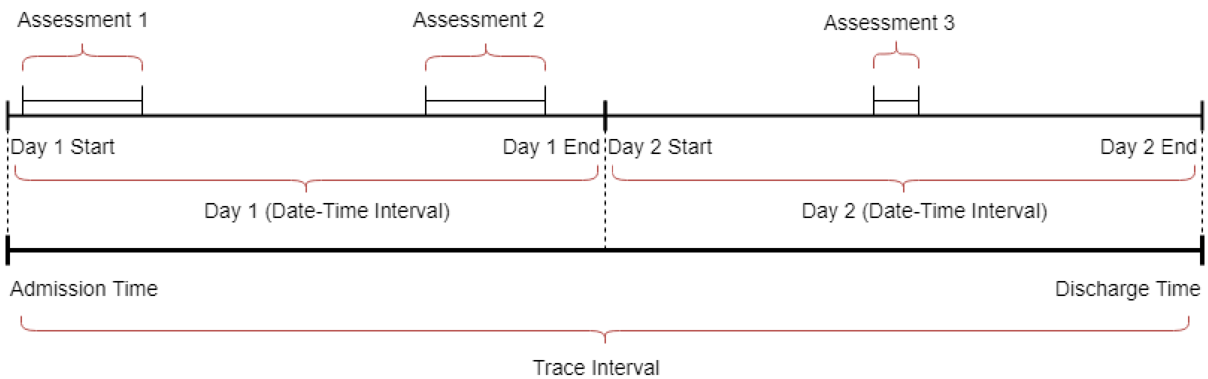

Figure 23: The relationship between intervals in the CP ontology.

3.5 Stroke CP Ontology Modelling

With the creation of the CP ontology complete, it is now possible to customize and extend this ontology, creating disease specific CP ontologies. While this customization can be completed for any disease, this research has selected ischemic stroke. A paper-based ischemic stroke CP was provided by Ottawa Hospital and studied, in collaboration with domain experts, to elicit new classes and properties required for ischemic stroke CP ontology. This bottom-up approach is valuable to identify concepts that are important to stroke, which may have been omitted in the top-down development of the CP ontology.

One ontology class central to the execution of stroke CP, is Patient CP Education. Since educating patients and their families on the rehabilitation process is crucial to a successful recovery, special attention was paid to the development of this class. Two children were added to this class after review of the ischemic stroke CP. These two children reflect the literature received by the patient after beginning the stroke CP. Furthermore, because they are also documents, these two educational booklets (A Guide for People Living with Stroke and TOH Care Companion Booklet) are also children of the class Document. The class structures of Patient CP Education and Document are shown in Figure 24 and Figure 25 respectively. The power of multi-inheritance OWL classes is demonstrated by this example and is another benefit of ontology in modelling CP.

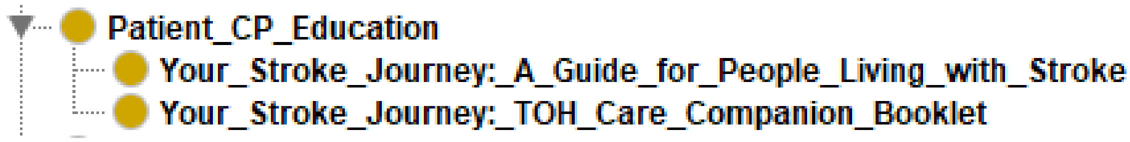

Figure 24: Patient CP Education hierarchy.

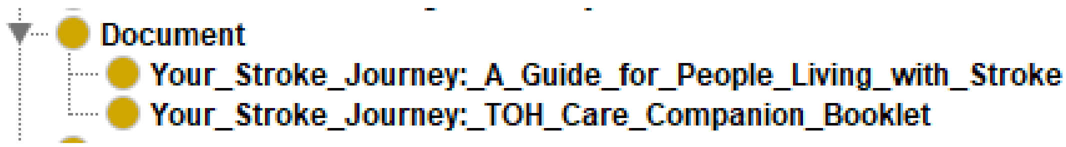

Figure 25: Document hierarchy.

During design of the stroke CP ontology, additions were also made to the CP Event class. The CP Event class of the CP Ontology is further expanded into five subclasses in the Stroke CP Ontology. These subclasses (or CP Event categories) include Activity, Assessment, Elimination, Nutrition and Teaching. These classes are derived from the instruction headings of the paper-based Stroke CP and are an efficient way of classifying different types of CP Events that occur during the execution of the Clinical Pathway. In short, an Activity is a specific action to perform on, or with, a patient; an Assessment is an action resulting in a concrete measurement or observation; Elimination is an action related directly (or indirectly) to toileting; Nutrition is an action related to dietary intake and outtake; and Teaching is an action directly associated with educating the patient or family members.

By expanding the CP Event class using these categories, the expected events of the CP execution can be organized and codified. Each day of the CP can easily be described in terms of which Activities, Assessments, Eliminations, Nutrition and Teachings should be performed. This breakdown is recorded in Table 3 below. Each specific event is described by an alpha-numeric code.

Table 4: The event breakdown for the Stroke CP.

Day	**CP Event Category**				
	Assessment	Nutrition	Activity	Elimination	Teaching
Admission	A0, A0a,A00, A000				T0, T00, T000
1	A1, A2, A3, A4	N1, N2, N3, N4, N5	AC1, AC2, AC3	E1, E2, E3	T1, T2
2	A1, A2, A3, A4	N6ab, N3, N4, N5, N2, N1	AC1, AC2, AC3	E1, E2, E3	T1, T3
3	A1, A2, A3, A4	N2, N3, N4, N5, N1	AC1, AC2, AC3	E1, E2, E3	T1, T4, T5
4	A1, A2, A3, A5	N2, N3, N4, N7	AC1, AC2,	E1, E2, E3	T6, T9

			AC3		
5	A1, A2, A3, A5	N2, N3, N4, N7	AC1, AC2, AC3	E1, E2, E3	T7, T10, T8

Organized in this fashion, the semi-structured written instructions of the paper-based CP have been translated to repetitive, codified processes. This representation also facilitates CP interoperability, a major objective of this research. In the exchange of medical practices between different healthcare organizations, it would be more effective to share a table of well-defined, reoccurring events than unorganized paper-based instructions.

Using paper-based CP, Hospital X would share their Stroke CP with Hospital Y through some combination of discussion, electronic transfer of a pdf copy, or mail transfer of the physical document itself. However, once grouped in this fashion, Hospital Y would be able to view a succinct overview of Hospital X's stroke CP processes, organized by general CP Event categories. Furthermore, CP Event patterns become more noticeable in this format, such as the repetition of Nutrition Events N2, N3, N4 and N7 on both Day 4 and Day 5. In the development of a Stroke CP Ontology, standardization and organization of CP Events becomes essential.

Once the possible CP Events of the Stroke CP are clearly defined and codified, their hierarchy, interrelationships and properties can effectively be explored. When presented in a different view as an Event-Day matrix (Table 4), the frequencies of certain CP Events are shown in a direct manner. In this format, both the most common and least common CP Events are easier to identify. Once again, this step towards the development of stroke CP ontology can be an effective tool for interoperability.

Table 5: Event-Day Matrix categorized by Event type.

CP Event Subclass	Name	Admission	Day 1	Day 2	Day 3	Day 4	Day 5
Assessment	A0	x					
	A0a	x					
	A00	x					
	A000	x					
	A1		x	x	x	x	x
	A2		x	x	x	x	x
	A3		x	x	x	x	x
	A4		x	x	x		
	A5					x	x
Nutrition	N1		x	x	x		
	N2		x	x	x	x	x
	N3		x	x	x	x	x
	N4		x	x	x	x	x
	N5		x	x	x		

	N6a			x			
	N6b				x		
	N7					x	x
Activity	AC1a		x	x	x	x	x
	AC1b		x	x	x	x	x
	AC2a		x	x	x	x	x
	AC2b		x	x	x	x	x
	AC3		x	x	x	x	x
Elimination	E1		x	x	x	x	x
	E2		x	x	x	x	x
	E3		x	x	x	x	x
Teaching	T0	x					
	T00	x					
	T000	x					
	T1		x	x	x		
	T2		x				
	T3			x			
	T4				x		
	T5				x	x	x
	T6					x	
	T7						x
	T8						x

As stated previously, each alpha-numeric code represents a specific CP Event within the Stroke CP ontology. For example, CP Event ‘A0’ is the paper-based CP instruction “Complete Swallowing Screen”. This instruction is contained within the Stroke CP Ontology as an OWL individual of the class Assessment. Furthermore, it is standardized using the SNOMED CT Term ‘Screening for Dysphagia’. Therefore, the Assessment Event Intervention takes the more standardized name ‘Complete Screening for Dysphagia’. Its position, as a member of the greater ontological hierarchy is shown in Figure 26 below. The class CP Event is a direct child of the class OWL Thing (the predefined root class of any OWL ontology). This CP Event class has five children, as discussed previously. The individual named Complete Screening for Dysphagia is an instantiation of one such child class, Assessment.

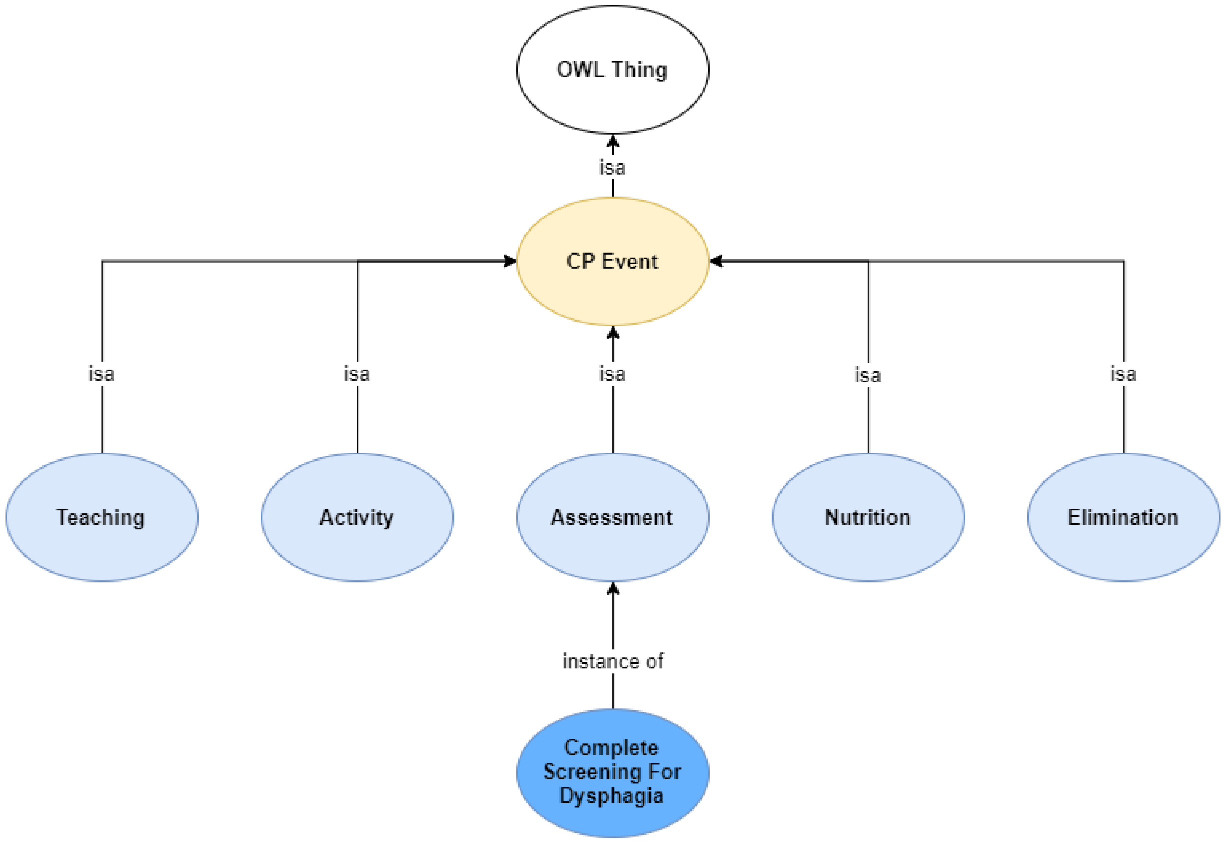

Figure 26: The CP Event component of the Stroke CP ontology.

This process continues for each CP Event identified in the paper-based ischemic stroke CP. Once a CP Event has been identified, it is related to a specific Intervention child class of the CP ontology depending on its characteristics, domain research and the opinion of domain experts. Once a child of Intervention has been selected, an individual of this class (Nursing Task for example) is created. The individual's name will typically be related to the overall action of the CP Event. For example, if the CP Event is to "Assess for signs and symptoms of nicotine withdrawal", the intervention individual will be named "Nicotine Withdrawal Assessment Intervention". Next, the library of SNOMED CT Concepts will be reviewed in order to find a standardized name and ID for this intervention. If this term does not already exist, it will be added to the MK ontology.

In the previous example, "Nicotine Withdrawal Assessment Intervention" is equivalent to the SNOMED CT Concept "Assessment of substance withdrawal" and has ID 711008001. Next, an Observation individual is created in a similar way to Intervention individuals. Any individual of the Observation class will have the data property "observation value", which is used to assign a literal value for that observation. Finally, the potential Outcome individuals for the CP Event in question are determined again through review of the paper-based CP and consultation with domain experts. These outcomes undergo a process of standardization and will be further discussed in Section 3.5.1. In general, the process for bottom-up construction of the Stroke CP ontology is as follows:

1) Isolate CP Event
2) Determine the Intervention type of the CP Event
3) Equate the Intervention to a standardized SNOMED CT individual of the MK ontology
4) Determine the Observation type of the CP Event
5) Determine all possible Outcomes of the CP Event
6) Equate the Outcome to a standardized SNOMED CT individual of the MK ontology
7) Relate all individuals of steps 1) – 6) using the appropriate object properties of the CP ontology

Figure 27 below depicts the full structure and interrelationships of a single Stroke CP Event (Complete Screening for Dysphagia). Following the process outlined above, and extending the original CP ontology, individuals of the Assessment, Physical Examination, Physiological Observation and Day Interval classes have been created. The Assessment individual Complete Screening for Dysphagia is related to the Physical Examination individual, Screening for Dysphagia Procedure by the object property "has Intervention Type". Screening for Dysphagia Procedure references the SNOMED CT Concept Screening for Dysphagia (ID 431765005) and produces the observation Screening for Dysphagia Observation. This observation can have a value of either Pass or Fail.

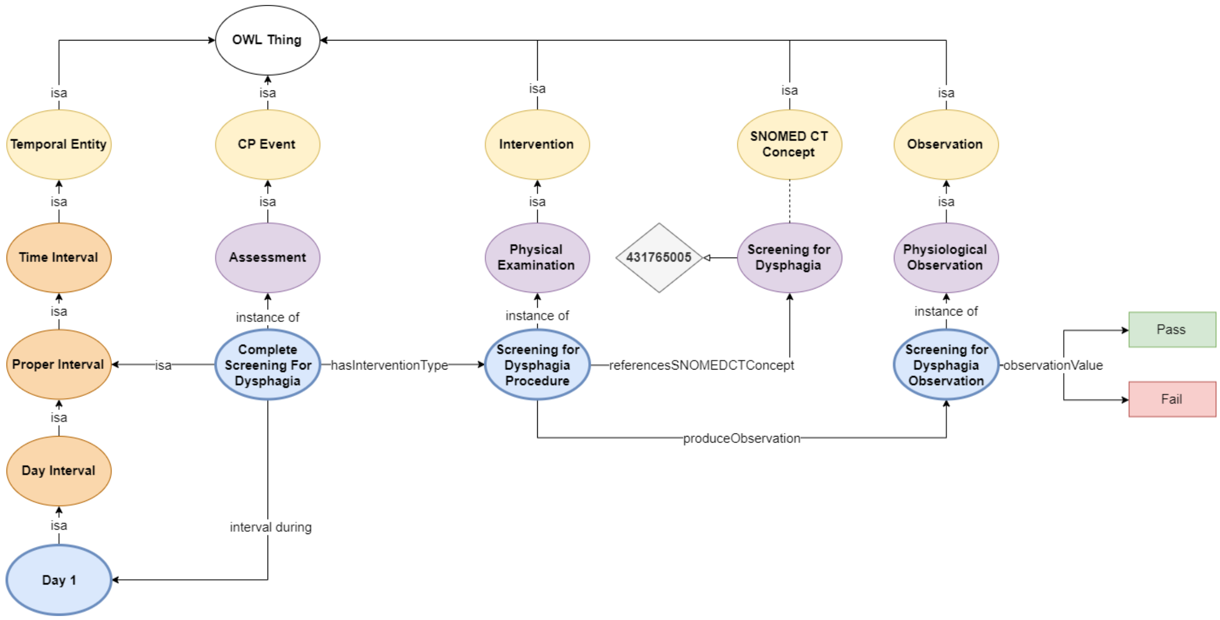

Figure 27: The semantic relationships of the CP Assessment Complete Screening for Dysphagia.

This method of design for the ischemic stroke CP ontology can be applied to any disease CP. Regardless of the CP chosen, the design steps will remain the same, which are creation of disease specific ontology classes, followed by elicitation of CP Event, Intervention, Observation and Outcome individuals. In this fashion, any number of disease CP ontology can be created using the CP ontology as a semantic base. This relationship, of X number of disease specific CP ontology developed from the single CP ontology is shown in Figure 28 below.

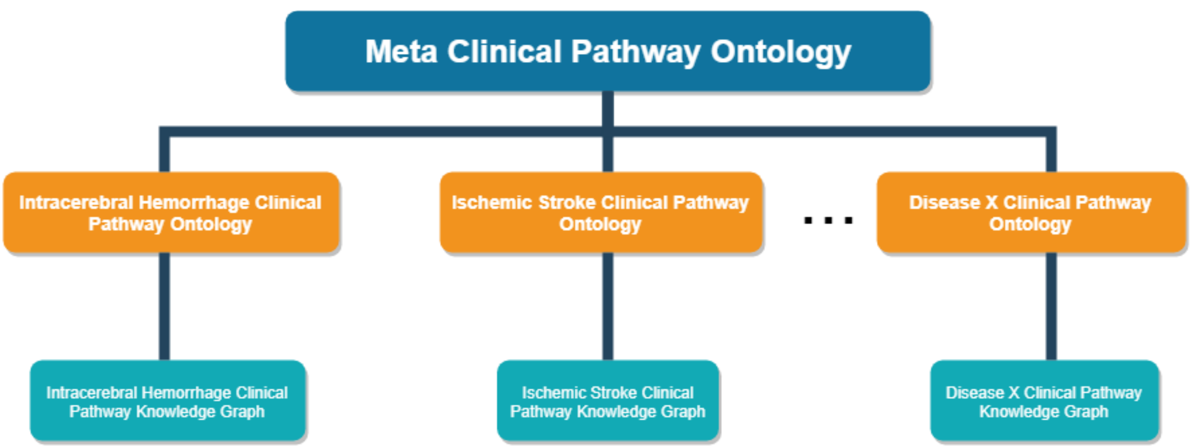

Figure 28: The hierarchy of CP ontologies.

The final step in development of the stroke CP ontology is the design of SWRL statements to infer new knowledge as it arises during the execution of a CP. These statements should be able to infer the Outcome of any CP Event based on the results observed by health care professionals. These observation results are stored using the data property 'observation value'. The SWRL statements of any disease ontology should also be generic enough as to limit the total number of statements. Development of generic SWRL statements will allow for the number of SWRL statements to be related to the number of unique interventions (dozens in the stroke CP) instead of the number of CP Events (hundreds or even thousands depending on the size of the CP). The SWRL statement design and development process is documented in Section 3.5.1.

3.5.1 Event-Intervention and Observation-Outcome Execution Chains

At the time of completion of an event in the progression of a CP, a CP Event will be related to a specific Outcome through the execution of SWRL statements. This is possible due to the semantic relationship between CP Event, Intervention, Observation and Outcome. The object properties defined between these classes allow for SWRL statements to be crafted in a generic way, but still infer correct knowledge for the outcome of a specific event. The properties relating CP Event, Intervention, Outcome and Observation are shown in Figure 29 below.

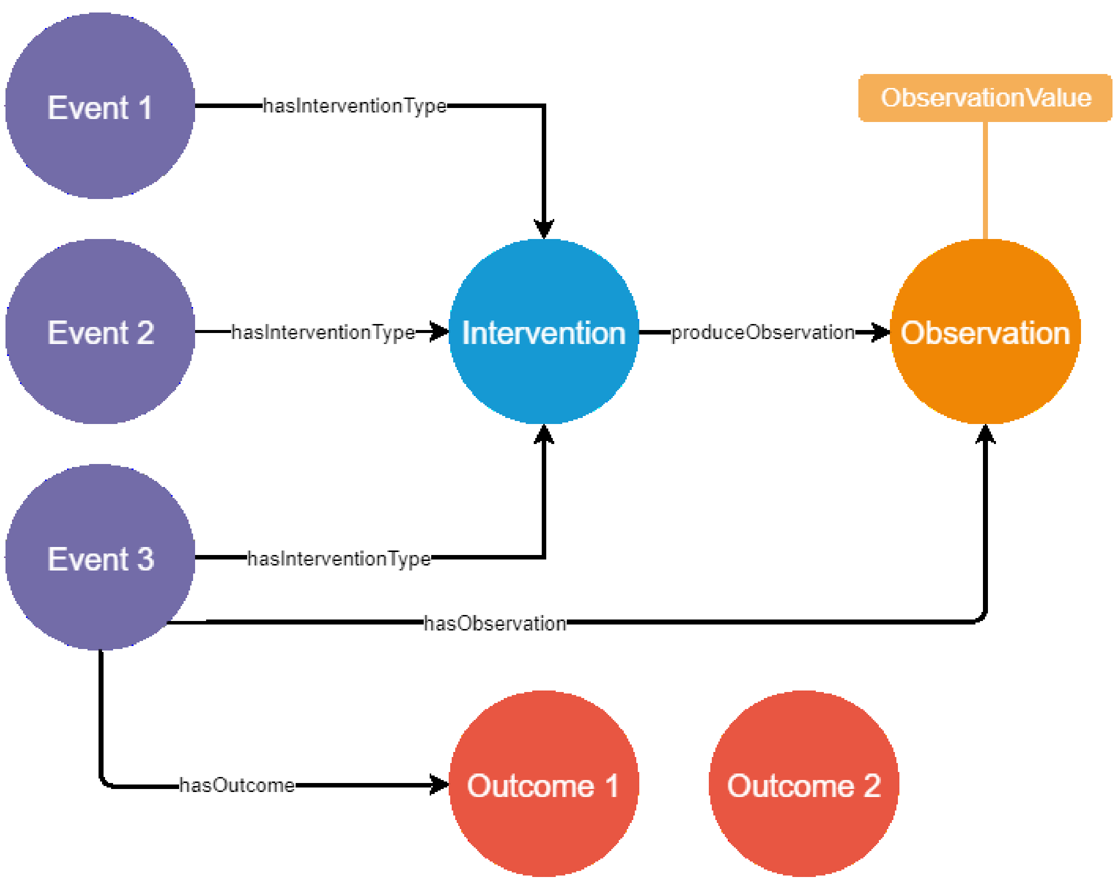

Figure 29: The interrelationship of CP Event, Observation, Intervention and Outcome.

The key to this design not changing all CP Events that have the same Intervention type after execution of SWRL statements is the “has Observation” relationship, which will exist between only a single CP Event and Observation at a time. Therefore, only a single CP Event will be related to an Outcome. The key to not having to uniquely specify a CP Event Individual name and therefore, not repeating the same SWRL rule multiple times is due to the “has Intervention Type” relationship. Only CP Event’s that are of the stated Intervention type will be affected by the SWRL rule. As a result, a CP Event can be referred to generically as ‘?x’ because it is identified uniquely by an Intervention individual. An example of a SWRL statement found in the stroke CP ontology is shown below:

CP_Event(?x) ^ intervention(DVT Assessment Intervention) ^ observation(DVT Observation) ^ hasInterventionType(?x, DVT Assessment Intervention) ^ hasObservation(?x, DVT Observation) ^ observationValue(DVT_Observation, “red skin”) ^ outcome(Deep Venous Thrombosis)
→ hasOutcome(?x, Deep Venous Thrombosis)

In this SWRL rule, if an Intervention, Observation and Outcome all exist pertaining to Deep Venous Thrombosis (DVT Assessment Intervention, DVT Observation, Deep Venous Thrombosis) and an Observation Value of “red skin” is recorded then the outcome of the CP Event will be Deep Venous Thrombosis. Therefore, a new relationship (new knowledge) is created called has Outcome between that specific Event and the Outcome Deep Venous Thrombosis.

SWRL statements to infer the next state of a CP should also be developed. An example of such a rule is shown below. In this example, if the outcome of a Nursing Task, “Nicotine Withdrawal Assessment” is the Outcome, “Nicotine Withdrawal Symptoms” then the Trace of this CP now satisfies “Consult with Physician about Nicotine Replacement Options”. In simpler terms, if a nurse finds that a patient is experiencing nicotine withdrawal symptoms during their assessment, the next step in the CP should be to consult with a physician about nicotine replacement.

Trace(Stroke Trace) ^ State(Consult with Physician about Nicotine Replacement Options) ^ Nursing Task(Nicotine Withdrawal Assessment) ^ Outcome(Nicotine Withdrawal Symptoms) ^ hasOutcome(Nicotine Withdrawal Assessment, Nicotine Withdrawal Symptoms)

→ satisfy(Stroke Trace, Consult with Physician about Nicotine Replacement Options)

These two categories of SWRL statements can be used in concert to first, infer what Outcome has occurred for a certain CP Event and secondly, what new state of the CP arises from this Outcome.

3.5.2 The Duality of Tasks

A key focus in the design of the stroke CP ontology is the seamless integration between the three ontologies of Section 3.2, Section 3.3 and Section 3.4. In order to achieve this goal, the concept of the duality of tasks is introduced. A task in the stroke CP ontology is considered any explicit instruction from the paper-based CP. These tasks were previously organized and denoted using alpha-numeric codes (Section 3.5). For example, the instruction "Assess for signs and symptoms of DVT" is considered a task. Any task in the stroke CP ontology exists as an OWL individual. Individuals are assigned types that correspond to the OWL class to which each individual belongs. In this ontology design, task individuals are assigned 2 types, CP Event and Proper Interval. This indicates that any task to be completed in the stroke CP ontology belongs to both the CP Event class of the Meta CP Ontology and the Proper Interval class of the Time Ontology. The duality of tasks allows any instruction in the paper-based CP to be expressed in the stroke CP ontology both procedurally as a Clinical Pathway concept and temporally as a Time concept.

3.6 Clinical Pathway Management System

The stroke CP ontology is presented to users through a prototype Clinical Pathway Management System (CPMS). This CPMS is designed to aid healthcare professionals in the progression of a patient through a CP. It provides a mechanism for the viewing, updating and editing of the stroke ontology through a simple UI.

3.6.1 Software Development Environment

The CPMS was developed in Java using the Eclipse IDE. Java was selected as the programming language of choice for the CPMS because of its popularity and the existence of several useful APIs. The Eclipse IDE was also selected because of its popularity and overall ease of use. Connecting the stroke ontology to the GUI of the CPMS are the OWL and SWRL APIs. The OWL API is used for creating, manipulating and serializing OWL ontologies [32] with support for OWL2. It provides components for work with OWL classes, individuals, object properties and data properties. It also provides Java classes for managing, reasoning over and creating OWL ontologies. The OWL API and its SWRL counterpart bridge the gap between ontology and the interface of the CPMS. Without the OWL API, a simplified user experience would be much more difficult.

3.6.2 System Architecture and Presentation

The CPMS is designed using the Model-View-Controller (MVC) paradigm. The MVC design of the CPMS is shown in Figure 30 below. MVC was considered the most appropriate design paradigm for this system because of the nature of ontology, the OWL API and GUI. The ontology is the CPMS model. It contains all data related to the execution of the CP. The functions of the OWL API are called by the controller of the CPMS. Actions such as updating the model or inferring new knowledge from the model will be driven by the OWL API based controller. Finally, the entire GUI of the CPMS will be contained within the view. It will present the data of the ontology in a simple, logical manner. A tabbed, checklist-centered GUI design is selected for the CPMS.

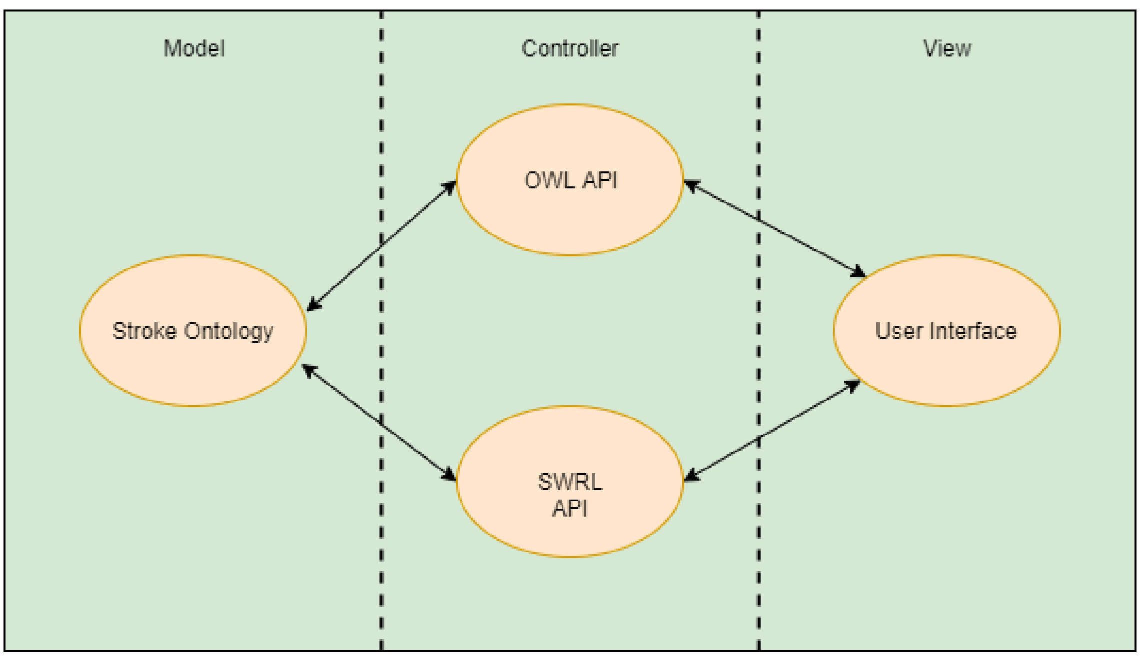

Figure 30: MVC paradigm applied to the structure of the CPMS.

The GUI of the CPMS is based on a two-tiered tab presentation. The first layer of tabs will display each day of the CP. In the case of the stroke CP, this will be five days. The second layer of tabs will display each CP Event category, in the context of ontology these are the children of CP Event. Therefore, a user of the system who would typically be a doctor, nurse, specialist or other healthcare professional will be able to easily navigate between the events of different days and categories without needing to understand the deeper semantic structure of ontology.

Each category tab will contain all CP Events of that category for that specific day, presented as a checklist. As events are completed, they can be checked off on the present checklist. The two-tiered checklist GUI is displayed in Figure 31 below.

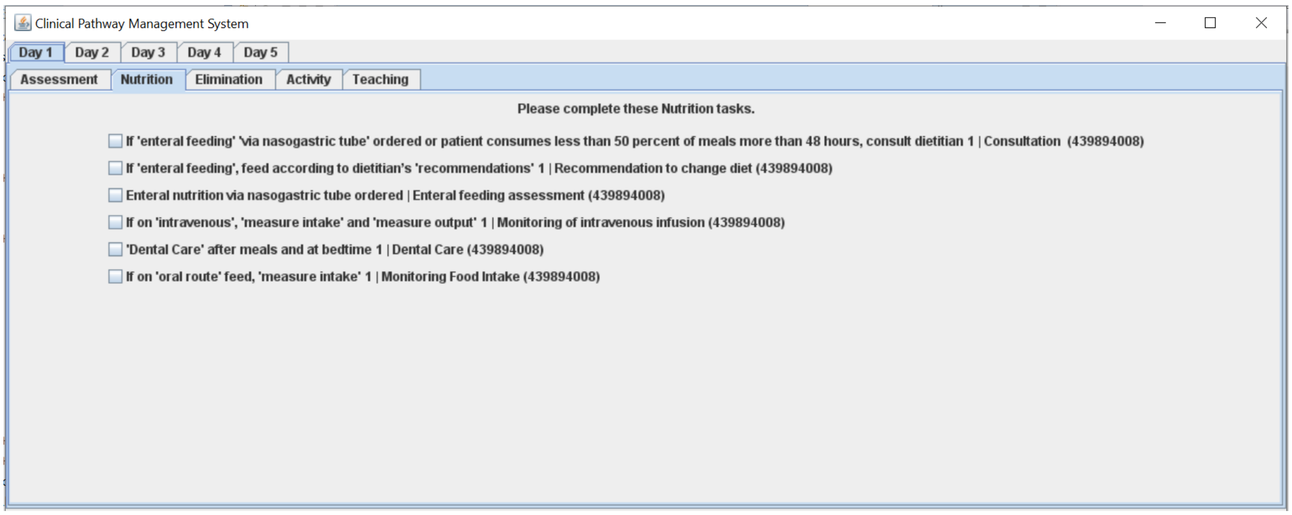

Figure 31: The GUI for Day 2 Nutrition Events in the CPMS.

The components of the CPMS view presented as a class diagram are shown in Figure 32 below. System classes inherit from the major Java UI classes JFrame, JTabbedPane and JPanel. The one-to-many relationship between Progress Menu and Day Screen as well as the one-to-many relationship between Day Screen and Encounter Screen highlight the flexibility of this GUI design to accommodate any number of days or CP event categories for each CP.

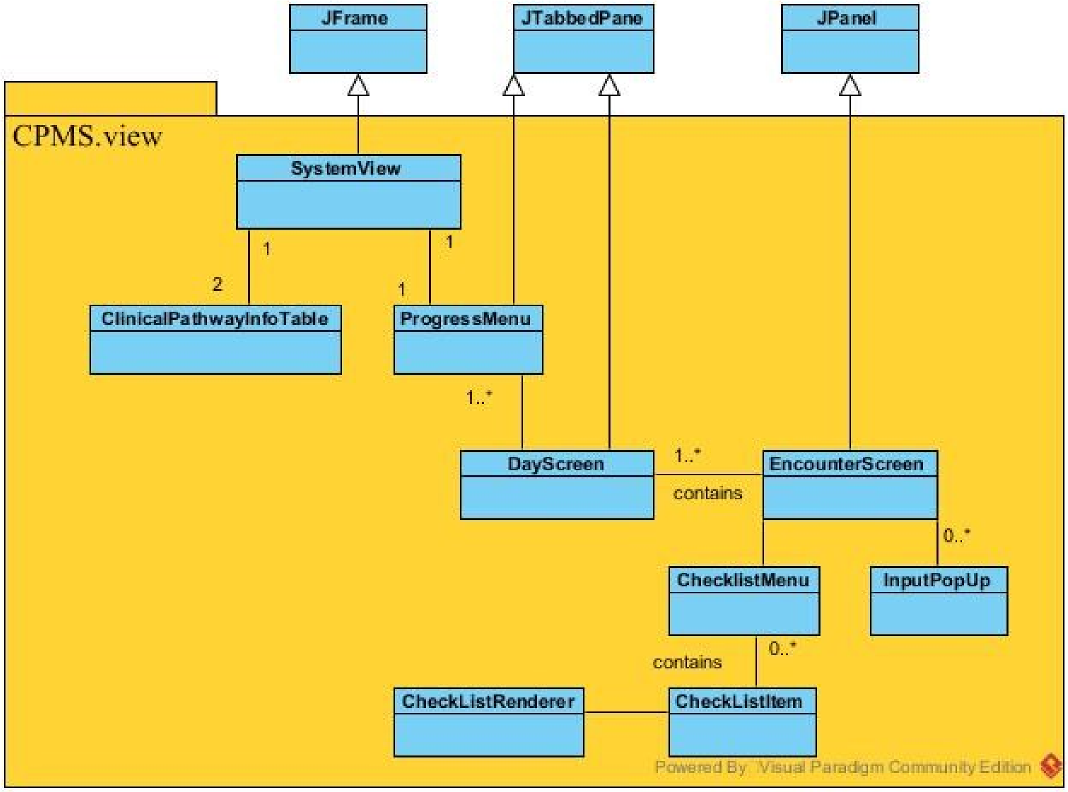

Figure 32: CPMS View class diagram.

3.6.3 System Execution and CP Semantic Interoperability

During the execution of the ischemic stroke CP, the CPMS will begin with an ideal version of the stroke CP Ontology. This will contain all of the expected Interventions, initial CP Events, possible Outcomes and SWRL statements. Patient data will be read by the system from an Electronic Medical Record (EMR) system or other Electronic Record System. In this research, a csv file simulates an EMR. Based on the patient ID, a new Patient Individual was created or a previously existing Patient is appended with any new data. These steps are documented in the pseudocode below.

FUNCTION ReadEMR

COUNT *rows* in EMR
GET *Column Names* in EMR
FOR EACH *Column Name*
IF *Column Name* IS NOT a data property of the Ontology Patient Class THEN
CREATE a new Patient Class data property
SET the range of the new data property to string
ENDIF
END FOR

```
    FOR EACH row in EMR
        GET Patient ID and Patient Name
        DISPLAY Patient ID and Patient Name to the user
    END FOR
    PROMPT user to make Patient selection
    GET Patient EMR data from the selected patient EMR row
    IF Patient DOES NOT exist THEN
        CREATE new Patient Individual in Ontology
        FOR EACH data property in Patient
            SET data property value to the corresponding EMR row value
        ENDFOR
    ENDIF
    ELSE
        FIND existing Patient Individual
        FOR EACH data property in Patient
            UPDATE data property value to the corresponding EMR row value
        ENDFOR
    SAVE Ontology
ENDFUNCTION ReadEMR
```

Once an event in the Stroke CP has been complete it will be checked off in the appropriate checklist. The user will then be prompted for input based on their observations during the current event. The process of user interactions with the CPMS is documented in Section 3.6.4. Next, the CPMS will update that CP Event. Specific data pertaining to timing, observations and outcomes will be linked to that CP Event within the CPMS model (stroke CP ontology). New knowledge can be inferred based on this additional data. The process of updating a CP Event is documented in the pseudocode below.

FUNCTION UpdateEvent

Input: *observation result, event, intervention term, intervention ID, observation, start time, end time, patient ID*

```
    GET the CP Event identified by event
    SET observation data property observationValue = observation result
    SET CP Event start to start time
    SET CP Event end to end time
    CALL fireRules
    CALL saveOntology
```

ENDFUNCTION UpdateEvent

The function fireRules is designed to load and execute all SWRL statements created for the current ontology of the CPMS. This process is completed in three steps.

1) Import all SWRL statements into the model
2) Run all imported statements on the ontology
3) Export all inferred knowledge back to the ontology

The final result of this updating process is that the CP Event is now semantically linked to certain Outcome, based on the observations that occurred during the CP Event and the SWRL statements specific to this CP.

3.6.4 Typical Use Cases and Scenarios

In order to demonstrate the merit of the CPMS, two scenarios will be demonstrated below. The first scenario documents the CPMS experience for a user (likely a nurse in this situation) during the execution of a Swallowing Screen in the stroke CPMS. The second scenario documents the CPMS experience for a user, likely a consulting medical professional or medical information system (MIS) administrator, attempting to standardize another paper-based CP.

3.6.4.1 Scenario 1: Screening for Dysphagia Procedure

On admission, an ischemic stroke patient is expected to undergo a Swallowing Screen. This instruction, "Complete Screening for Dysphagia", is considered a CP Event of the category Assessment. Anytime this instruction is completed, a new Assessment Event is created within the Stroke CP Knowledge Graph with details uniquely related to the new Event. However, any Complete Screening for Dysphagia Event within the Stroke CP Knowledge graph will have the same Intervention type. In this scenario, any Complete Screening for Dysphagia Events will belong to the Intervention "Screening for Dysphagia Procedure". Therefore, a many-to-one cardinality will exist between CP Event classes and an Intervention class. A CP Event can be considered an instantiation of an Intervention, with corresponding knowledge about patient outcome, timing and location.

By relating multiple Events to a single Intervention, standardization also becomes simpler. In this architecture, the relationship between an Intervention and SNOMED CT Concept, not the relationship between a CP Event and SNOMED CT Concept, is at the core of CP standardization. Each Intervention, not Event, will reference a standardized SNOMED CT Concept and ID. This eliminates the need to create a new Intervention individual if (or when) an Intervention should be repeated. If there are a series of CP Events that are closely related but not identical, for example "Consult with Physician about Symptoms", "Consult with Physician about

Treatment” and “Consult with Radiologist”, all three may reference the same standardized SNOMED CT Concept “Consultation”. In this architecture, it is possible to relate all three to the same concept, while conserving the uniqueness of three distinct interventions. All that would be necessary in this situation is semantically linking all three interventions to the same SNOMED CT Concept. This will reduce redundancy and improve scalability in larger Management Systems.

In this scenario, the Intervention “Screening for Dysphagia Procedure” references the Concept “Screening for Dysphagia” with ID 431765005. Table 5 below summarizes the different related OWL entities for a single Swallowing Screen Event in the ischemic stroke clinical pathway.

Table 6: The components of the Swallowing Screening Event.

Event Name	Event Type	Intervention	Day	SNOMED CT Concept	SNOMED CT ID
Complete Swallowing Screening Procedure	Assessment	Screening for Dysphagia Procedure	1	Screening for Dysphagia	431765005

The overall workflow of the Screening for Dysphagia, focusing on its progression from an uncompleted event to a completed event with a certain outcome is documented in Table 6 below. Where possible, these entities are linked to a SNOMED CT Concept using the object property references SNOMED CT Concept.

Table 7: The workflow of the Swallowing Screen Event in the ischemic stroke CP.

<table>
<tr><th>OWL Entity</th><th colspan="3">OWL Individual Name</th><th>SNOMED CT Concept</th></tr>
<tr><td>Event</td><td colspan="3">Complete Swallowing Screen Procedure
↓</td><td>N/A</td></tr>
<tr><td>Intervention</td><td colspan="3">Screening for Dysphagia Procedure
↓</td><td>431765005 | Screening for Dysphagia</td></tr>
<tr><td>Observation</td><td colspan="3">Screen for Dysphagia Observation</td><td>258149004 | Swallowing Finding</td></tr>
<tr><td>Observation Value</td><td>↓
Pass</td><td colspan="2">↓
Fail</td><td>N/A</td></tr>
<tr><td>Outcome</td><td>↓
Successful</td><td>↓
Failed</td><td>↓
Repeat</td><td>385669000 |Successful
27582007 |Repeat</td></tr>
</table>

				103709008 |Failed attempted procedure
State	↓ Continue CP Execution	↓ Consult SLP	↓ Repeat Screening Procedure	N/A

During the completion of a Screening for Dysphagia Procedure, a CPMS user will be prompted to perform actions and make several decisions. These actions will be registered by the CPMS, dynamically affecting the flow of the CP and future prompts displayed to the user. A typical execution of a Screening for Dysphagia Procedure is shown below.

1) On admission, a Screening for Dysphagia Procedure will be completed on the patient. The user will be prompted to enter the results of the Swallowing Screening Procedure (Figure 33). The 2 possible results for this procedure are Pass and Fail. These options are offered in a dropdown menu (Figure 34).

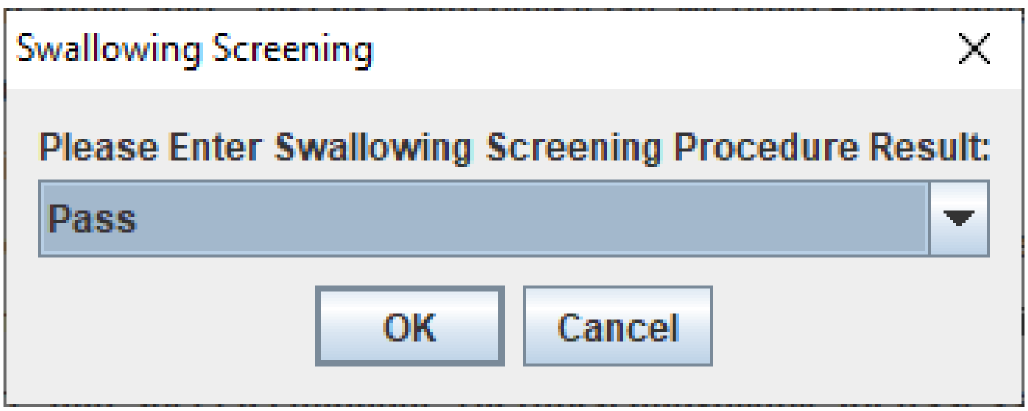

Figure 33: Screening for Dysphagia Procedure results prompt.

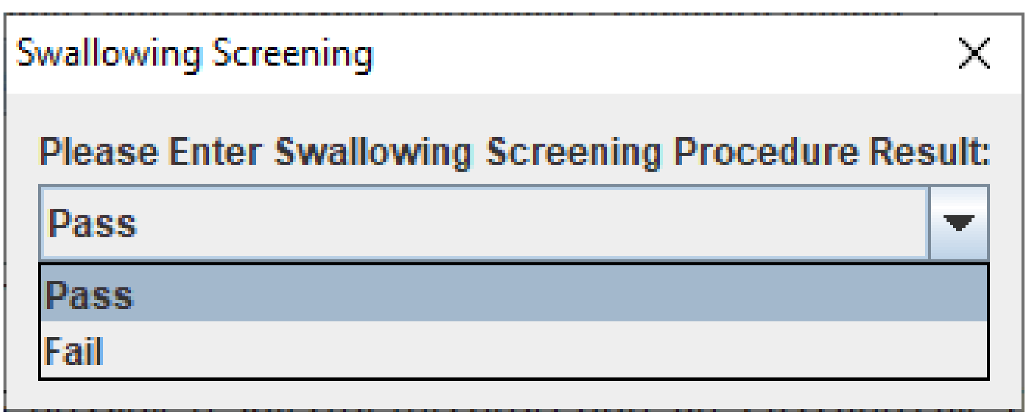

Figure 34: Screening for Dysphagia Procedure result options.

2) The user will select either Pass or Fail for the patient. Regardless of the observation value entered, the CPMS will now call the updateEvent function of Section 3.6.3. However, the result of this function, and its corresponding effect on the stroke CP ontology, will be dependent on the observation value.
    a. If the patient passes the Screening for Dysphagia, this task is complete and the CP will be continued.
    b. If the patient fails the Screening for Dysphagia, the user will be prompted about the availability of the Speech Language Pathologist (Figure 35).

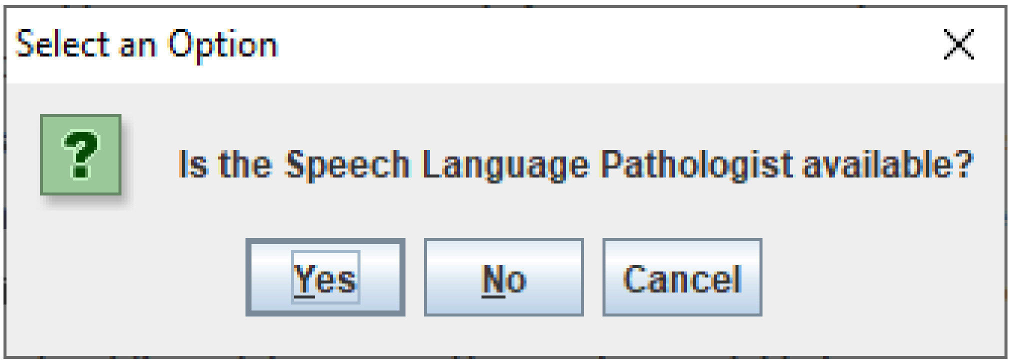

Figure 35: A prompt for the availability of the Speech Language Pathologist.

3) The user will select either Yes or No for the availability of the Speech Language Pathologist.
    a. If the Speech Language Pathologist is available, they should be consulted about the patient (Figure 36).
    b. If the Speech Language Pathologist is not available, the patient should repeat the Screening for Dysphagia Procedure within 24 hours (Figure 37). This CP Event will be added to Day 2 of the ischemic stroke CP.

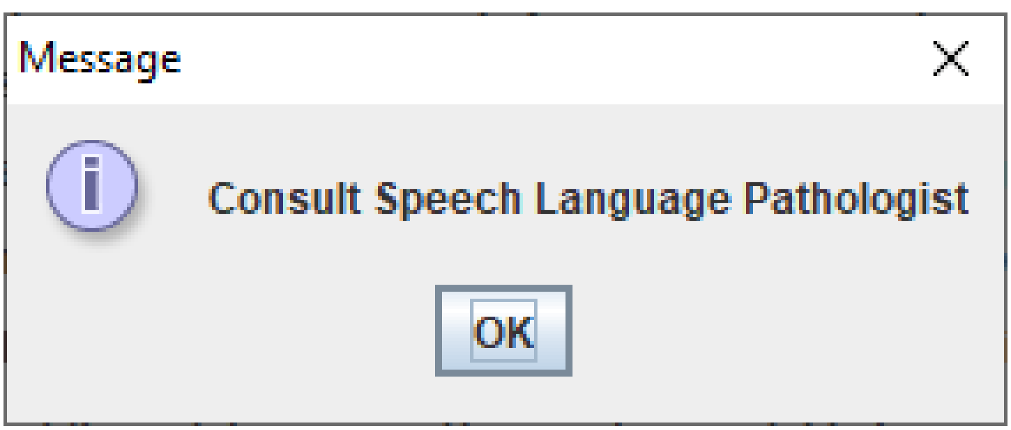

Figure 36: A prompt to consult the Speech Language Pathologist.

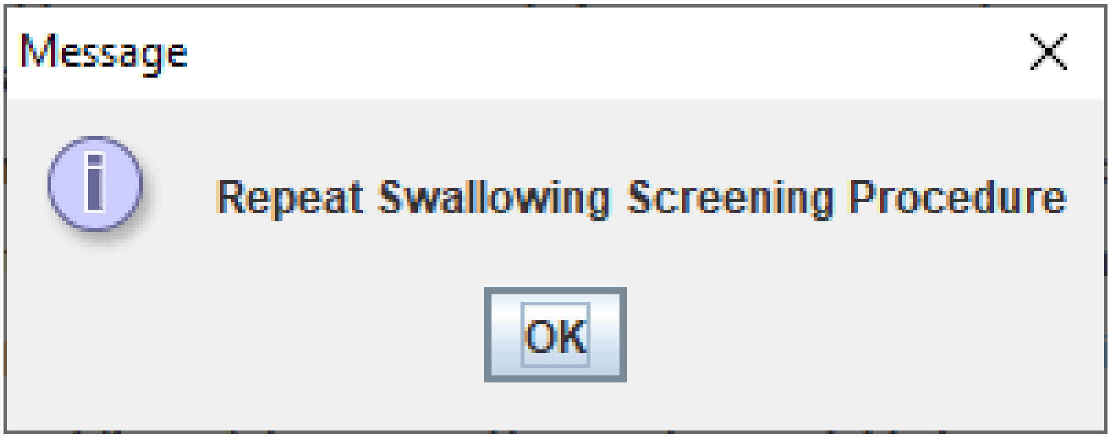

Figure 37: A prompt to repeat the Screening for Dysphagia Procedure (added to Day 2 of CP).

4) On completion of the repeated Screening for Dysphagia Procedure, the user will be prompted to enter the new results (Figure 38).

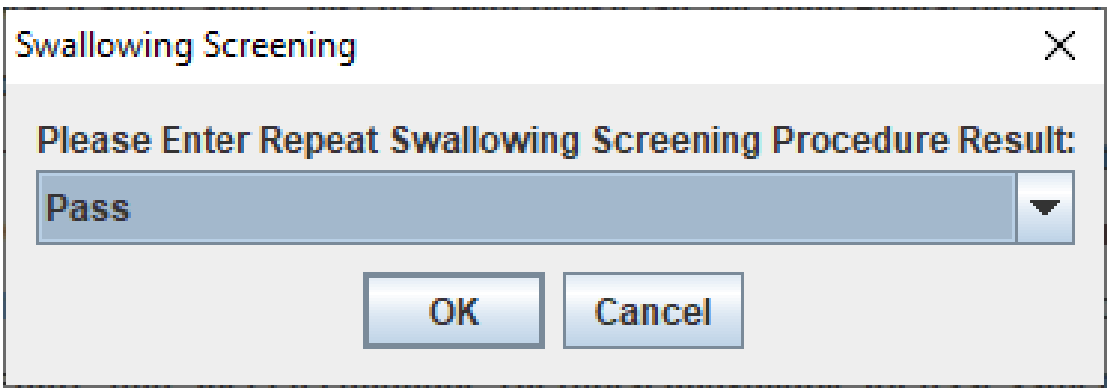

Figure 38: Repeated Screening for Dysphagia Procedure results prompt.

5) The user will select either Pass or Fail for the patient.
    a. If the patient passes the Repeat Swallowing Screening, the CP will be continued.
    b. If the patient fails the Repeat Swallowing Screening, the user will be prompted to consult with the physician (Figure 39).

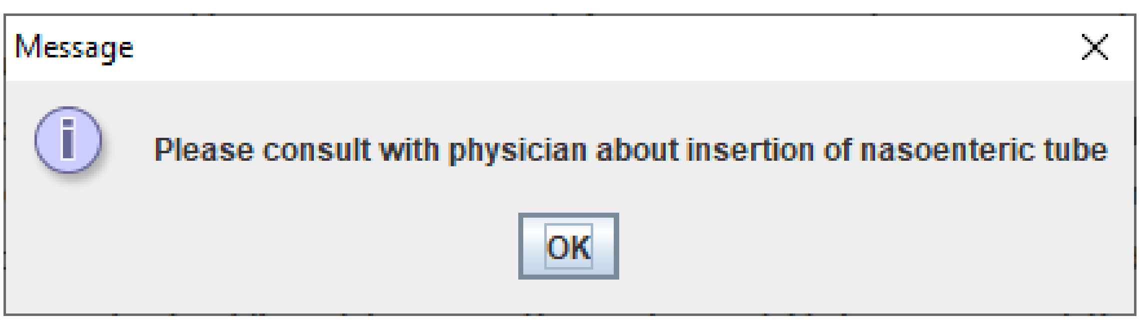

Figure 39: A prompt to consult with a physician because of the failed Swallowing Screening.

3.6.4.2 Scenario 2: Standardization Tool

The CPMS also allows users to search the collection of SNOMED CT Concepts. This is a powerful tool for individuals that have a new non-standardized paper-based CP and require standardization. In this case, a user can enter any term into the CPMS and be provided with a list of possible matching standardized SNOMED CT Concepts. If a non-standardized term is entered into the search bar and a complete (or partial) match is found, a user could substitute the non-standardized term with this match. This assistive tool, in collaboration with domain experts and critical analysis of potential matches, can aid future standardization of other disease specific CPs. Through this feature, a user can quickly check for standardized terms using the CPMS instead of other external software. The search bar interface of the CPMS standardization tool is shown in Figure 40 below.

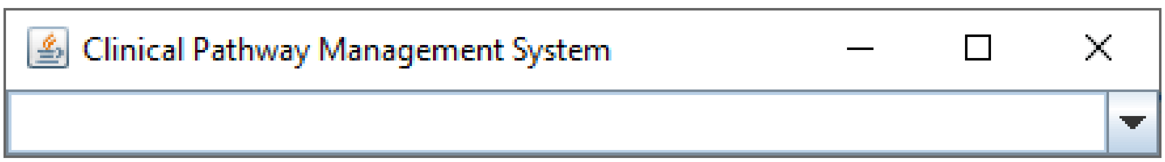

Figure 40: The CPMS Search Bar.

To begin, all available (previously created) SNOMED CT Concepts are shown in the search interface's dropdown menu (Figure 41). These concepts are listed in alphabetical order by default. They can be browsed manually if the user is unsure of what search term to enter or would simply like to browse the catalog of available terms. When the user does enter a search term, or part of a search term, the dropdown is updated to display the closest matching SNOMED CT Concepts. For example, in Figure 42, a user enters the search term "Care". There are 7 possible matching SNOMED CT Concepts, all including the term "Care". At this time, the user can continue to edit the search term or select one of the options in the dropdown. It is important to note that the first term displayed in the dropdown will always be the current search term.

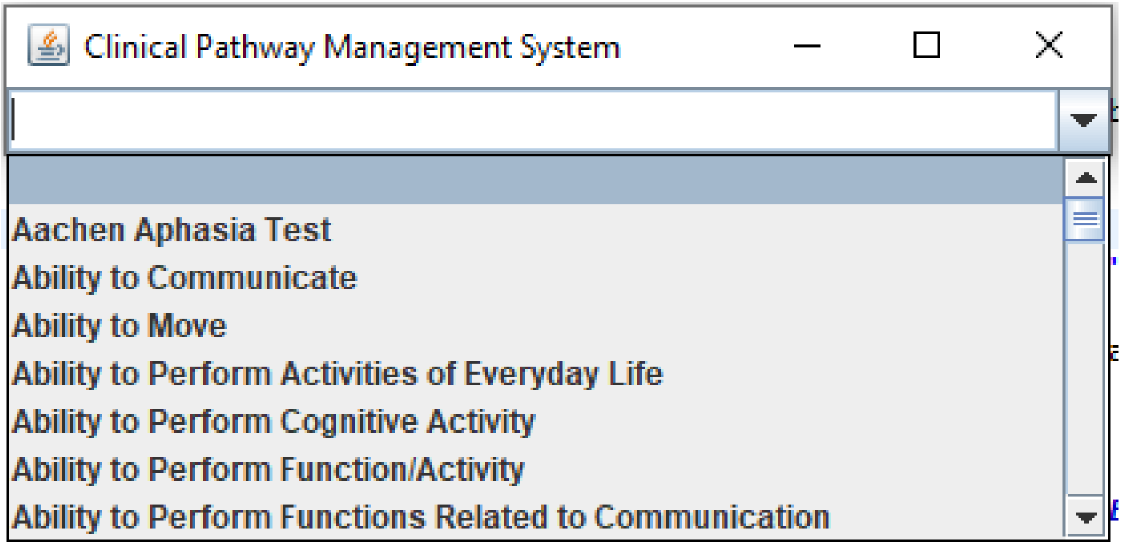

Figure 41: The CPMS Search Bar with a dropdown menu of all SNOMED CT Concepts.

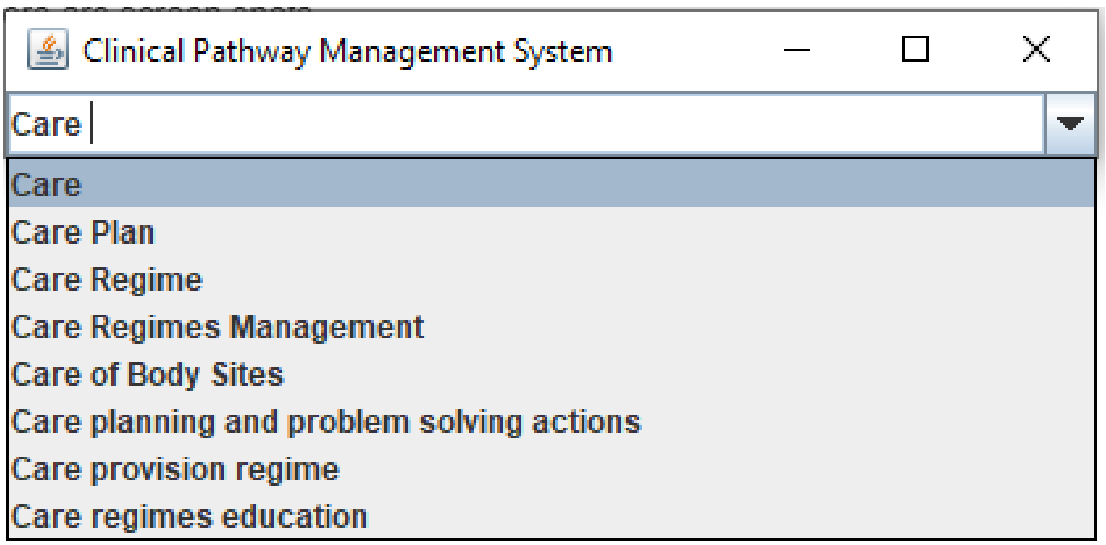

Figure 42: The CPMS Search Bar with the results of the search for the term "Care".

Guiding patient care through a CP and assisting medical professionals in the standardization of other CP are two major functionalities of the CPMS thus far. As the prototype is developed other features can be introduced to the CPMS. This may include graphical representation of a CP ontology, pathway recommendation, export of standardized terminology (SNOMED CT terms) and download/upload of other CP ontology.

3.7 Discussion

Through a three ontology design paradigm, a generic CP ontology was developed. The complete CP ontology draws from three separate domains represented by the Meta CP ontology, Medical Knowledge ontology and Time ontology. Individual additions were made to the ontologies to eliminate or reduce deficiencies identified in the literature review. These contributions include expanding the detail and semantic meaning of the variance class; the development of a relationship between event path and variance; the integration of admission, discharge and LOS within the time ontology; and the encapsulation of medical knowledge that may be referenced during a CP by semantic modelling of the SNOMED CT standard. These contributions accomplish objectives 2 and 3.

This generic CP ontology was then extended for use as a stroke CP ontology. This process included expanding the CP event class and Outcome class. This allowed specific stroke artifacts to be modelled not as generic concepts but as disease specific entities. For example, an instruction to teach a patient and their family about risk factors is captured not just as an event but a teaching event. This accomplishes, in part, objective 1. This process also included the development of a disease specific rule base using SWRL. These rules are used to infer new knowledge about the CP during its progression. This contribution also accomplishes objective 1.

Finally, a management system was developed to present the stroke CP ontology to potential users in a clear and intuitive way. This system allows for the progression of a CP in real-time, while inferring new knowledge about outcome and CP state. It should be a future aim of this research, to present this prototype system to stroke professionals for input and feedback.

4 Experimental Results

4.1 Generation of Patient Results and System Output

At the beginning of the execution of a clinical pathway, the system loads the generic knowledge graph for the CP in question. This knowledge graph varies from ontology in that it contains all initial individuals that should be present within the CP. They can be considered tangible objects or concepts present for all CP ontology for a specific disease. This should include at least one instance of each event occurring in the CP, all initial temporal instances (days, admission time, etc.) and instances of all referenced SNOMED CT classes. This disease CP knowledge graph is common to all patients undergoing that CP and will not have any patient specific (identifying) information.

Once execution is underway, specific events will occur and new knowledge will be added to the CP knowledge graph. This will include the creation of new data properties, object properties, running SWRL rules and the creation of new individuals. This process is covered in Section 3.6.3. At such time that any of the previous actions occur, the initial disease CP knowledge graph would be altered and therefore, customized to a patient specific execution of the CP. As a result, instead of overwriting the initial disease CP knowledge graph, a new version of the CP knowledge graph is created. This new knowledge graph will contain information pertaining only to a specific patient and their individual progression through a CP.

This method allows for a generic, but disease specific, CP framework to be preserved, while generating any number of new knowledge graphs containing knowledge unique to individual patients. This approach facilitates certain disease specific CP individuals to be created for all patients in a non-repetitive manner. For example, if it is common in a Stroke CP for all patients to receive a certain piece of literature, an instance of this literature can be created as part of the generic CP knowledge graph instead of during each specific execution.

The reusability of a general knowledge graph framework for a disease CP would be beneficial for both information system professionals altering the system, such that changes do not have to be repeated and also to medical professionals that may want to review pre-existing CP individuals in a single, concise file. From an auditing perspective, generating and preserving unique knowledge graphs for patient specific executions of CP allows for a richer history of the CP journey to be documented. Questions can be asked about the historic execution of a CP, such as the specific order of events, that otherwise may be difficult to answer. These generated files can be used in concert with EMR records and other CP output files to thoroughly capture a specific patient's experiences during a CP.

The process concerning the creation of new knowledge graphs from a generic framework is outlined in the pseudocode below. In summary, when changes occur to the original CP knowledge graph a unique knowledge graph is created and changes are then saved to this new knowledge graph.

```
FUNCTION saveOntology
Inputs: patient ID
    IF a new knowledge graph HAS NOT been created
        CREATE a new knowledge graph and APPEND the patient ID to the file
    ENDIF
    SAVE knowledge graph changes TO the new knowledge graph file
ENDFUNCTION saveOntology
```

A comparison of an original generic knowledge graph and a newly created knowledge graph is shown below. In these figures, a specific event "Discuss with patient nicotine replacement therapy" is completed. It is the case for this scenario that the replacement therapy has been refused (seen in the new observation value) and the overall outcome of the event is that there is a "Refusal of treatment by patient" (seen in the new "hasOutcome" object property).

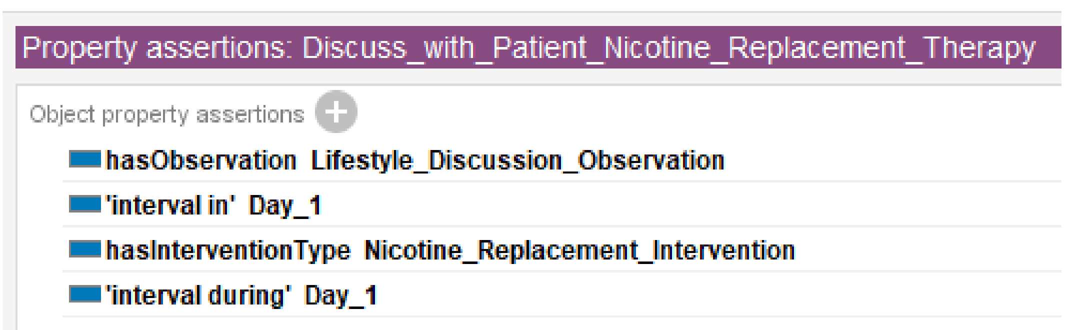

Figure 43: The nicotine replacement event from the generic knowledge graph. No outcome created.

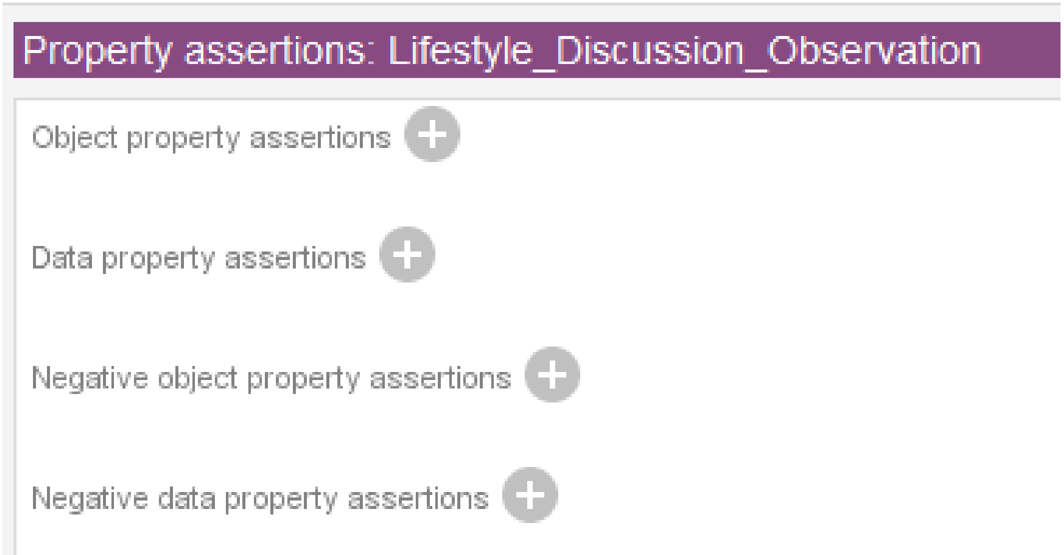

Figure 44: The discussion observation from the generic knowledge graph. No observation value created.

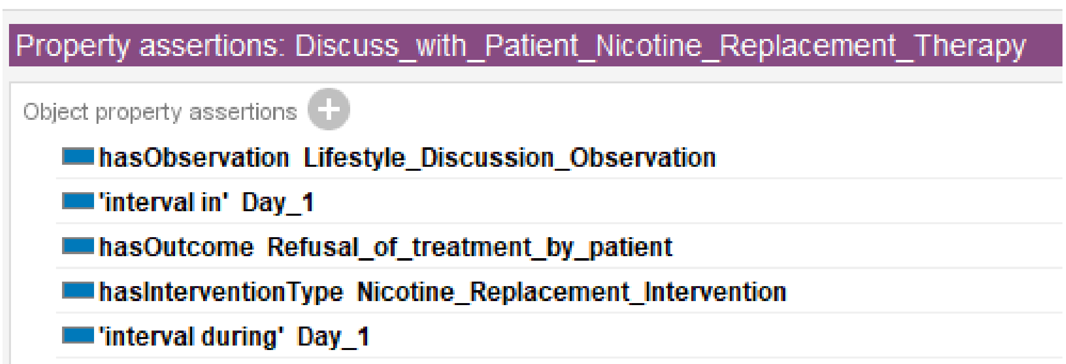

Figure 45: The nicotine replacement event from the patient knowledge graph. New outcome created.

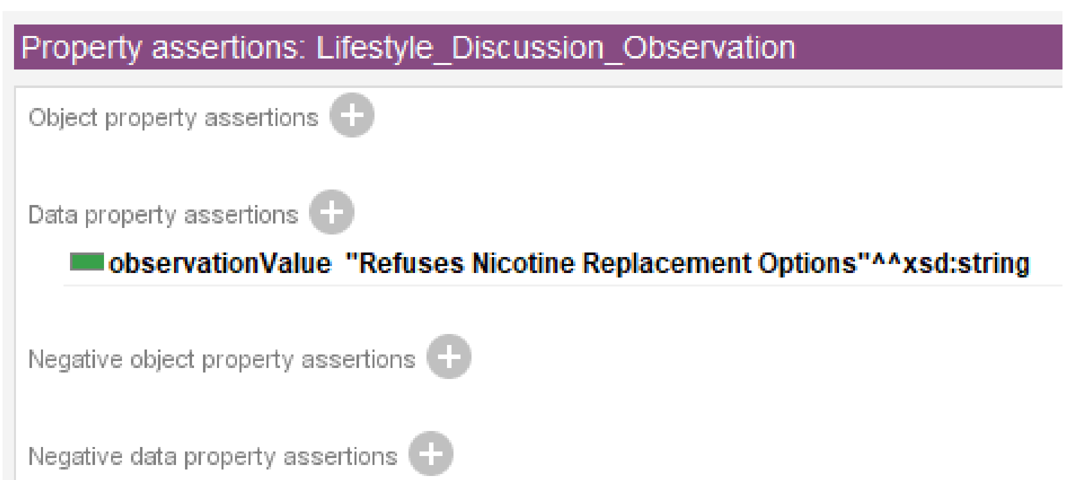

Figure 46: The discussion observation from the patient knowledge graph. New observation value created.

The hierarchy of ontologies, from the generic CP ontology to the patient specific knowledge graphs for stroke CP, is displayed in Figure 47. The base ontological structure is the CP ontology. This ontology is extended for stroke CP, specifically ischemic stroke. This stroke CP ontology introduces new classes and/or object properties but does not instantiate any individuals. It is the role of the Stroke CP Knowledge graph to instantiate initial, base individuals as explained above. Finally, a number (n) of patient specific stroke CP knowledge graphs are created, where n is the number of patients progressing through the stroke CP.

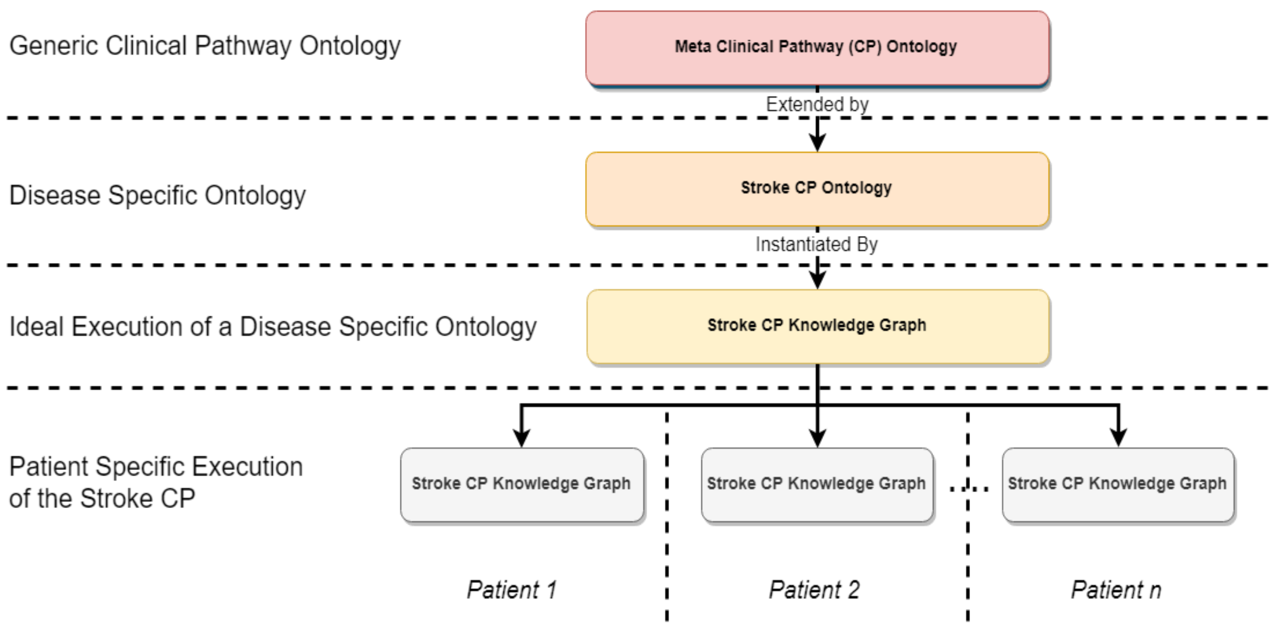

Figure 47: The hierarchy of ontologies.

Additional to the generation of unique patient knowledge graphs, an output file will be created for the execution of a disease specific CP knowledge graph. These output files will contain data for each patient that underwent that disease’s CP. Therefore, an output file for the ischemic stroke CP will be created with each row corresponding to data found in a certain patient’s specific knowledge graph. These output files will be further explained in Section 4.2.

4.2 Data Analytics

Data analytics can be performed on CP ontology is a number of ways. The focus of data analytics at present however, will focus on trend analysis in the output files generated for disease specific CP ontology. These output files, pertaining to interventions and outcomes for patients completing CP, will be used for application of the Longest Common Subsequence algorithm and spatial analysis.

4.2.1 Longest Common Subsequence

Screening Procedure	422504002-20135006	Functional Independence Measure	422504002-273469003	Tolerance Test Oral	422504002-90663002
Screening Procedure	422504002-20135006	Functional Independence Measure	422504002-273469003	Assessment of Substance Withdrawal	422504002-711008001
Screening Procedure	422504002-20135006	Functional Independence Measure	422504002-273469003	Tolerance Test Oral	422504002-90663002
Screening Procedure	422504002-20135006	Liaising with Speech and Language Therapist	422504002-225984004	Functional Independence Measure	422504002-273469003
Screening Procedure	422504002-20135006	Screening Procedure	422504002-20135006	Consultation	422504002-11429006

Figure 48: The interventions that occurred for each patient during the ischemic stroke CP.

Figure 48 is a subsection of an output file generated on completion of the Ischemic Stroke CP. Each row of the file belongs to a single patient's CP execution, for a total of 5 patients in this figure. Each row can also be considered the results of a specific patient knowledge graph as described in Section 4.1. It documents the Interventions completed during the execution of the CP using each Intervention's unique SNOMED CT Concept Name and corresponding hyphenated SNOMED CT ID. It follows the structure 'Concept Name, Hyphenated ID' such that each Hyphenated ID corresponds to the previous Concept Name. Taken as a pair, these items describe a single intervention. Therefore, there are 3 interventions documented per patient in this subsection.

The first section of each hyphenated ID, 422504002, is the SNOMED CT ID for the Concept 'Ischemic Stroke' because all of the above Interventions occur during the Ischemic Stroke CP. Uniformity of the first section in the hyphenated ID should always occur within the same output file because each csv file will correspond to a single CP type. Therefore, from an auditing perspective, if the first sections of the hyphenated IDs are not identical in a single output file, there are three potential situations. One, an error has occurred in recording the CP Interventions. Second, multiple CP files have been incorrectly merged. Third, during the execution of a specific CP, an intervention not typically part of the normal CP execution was performed. This case can be a result of variance due to several reasons including medical evidence, resource availability or patient/family input. If such a situation occurs, the output format proposed by this research allows for efficient identification of variant (non-CP) interventions. Therefore, when auditing a CP for standardization, it would be an effective first step to review the file of patient CP executions. A scenario in which a non-CP intervention may be included in a Stroke CP execution is shown in the figure below. The CP ID appending to the intervention 'Stroke Monitoring' is for the CP 'Thrombotic Stroke'. Due to one of the variance sources above, it may be decided to perform ‘Stroke Monitoring’ instead of ‘Screening for Dysphagia’.

Stroke Monitoring	371040005-170600009
Screening for Dysphagia	422504002-431765005

Figure 49: Stroke Monitoring is performed instead of Screening for Dysphagia.

This output format allows for simpler data analytics. When comparing the progression of two patients in the same CP, only the differences in the two specific output rows need to be noted. In Figure 48, after manual inspection, it can be seen that the patient of row 1 (Patient 1) and the patient of row 3 (Patient 3) have experienced the same progression of the Ischemic Stroke CP. Patient 1 and 3 both underwent the Swallowing Screening Procedures. They were then administered the Alpha Functional Independence Measure Assessment to document their ability to perform activities of daily living. Finally, both Patient 1 and 3 underwent Oral Tolerance Tests during the consumption of their first meal. In contrast, it is clear that Patient 1 and Patient 5 experienced significantly different progressions of the Ischemic Stroke CP. Both Patient 1 and 5 underwent the Swallowing Screening Procedure, however, after this Intervention their progressions diverge. Patient 5 undergoes a second Screening Procedure, indicating the first Swallowing Screening Procedure was failed and the Speech Language Pathologist was unavailable (see scenario above). A Consultation is then performed in Patient 5’s CP execution. This indicates a repeated failure of the Swallowing Screening Procedure and as a result a consultation with a physician about the insertion of a nasoenteric tube. Therefore, through the use of ‘Concept Name, Hyphenated ID’ pairs in CP execution output files; the progressions of different patients through the same CP can be easily compared and analyzed.

These output files can also be used for data analytics. One dynamic programming algorithm that can be applied effectively to reveal deeper trends in the CP output files is Longest Common Subsequence (LCS). The goal of the LCS algorithm is to find a common subsequence between two strings, which has a maximum length. The characters of this subsequence do not have to consecutive but they must be in order. For example, the strings “Huron” and “Heron” have a LCS of “Hron”. Formally, given two strings X and Y such that:

$$X = (x_1, \dots, x_m)$$

$$Y = (y_1, \dots y_n)$$

And let,

$$C[m, n] \equiv an\ array\ that\ contains\ the\ length\ of\ the\ LCS\ of\ X\ and\ Y\ at\ index\ (m, n)$$

The value of each index of array C can be found using the piecewise function,

$$C[i,j] = \begin{cases} 0 & if\ i = 0\ or\ j = 0 \\ C[i-1, j-1] + 1 & if\ i,j > 0\ and\ x_i = y_j \\ \max(C[i-1,j], C[i,j-1]) & if\ i,j > 0\ and\ x_i \neq y_j \end{cases}$$

Using this function, the LCS of any 2 strings can be found in a recursive way. This algorithm can be applied to the strings "card" and "cord". Given that X = "card" and Y = "cord", the LCS algorithm is applied as follows.

1) $\mathrm{C}[\text{"card"}, \text{"cord"}] = \mathrm{C}[\text{"car"}, \text{"cor"}] + 1 = 2 + 1 = \mathbf{3}$
2) $\mathrm{C}[\text{"car"}, \text{"cor"}] = \mathrm{C}[\text{"ca"}, \text{"co"}] + 1 = 1 + 1 = \mathbf{2}$
3) $\mathrm{C}[\text{"ca"}, \text{"co"}] = \max(\mathrm{C}[\text{"c"}, \text{"co"}], \mathrm{C}[\text{"ca"}, \text{"c"}]) = \max(1,1) = \mathbf{1}$
4) $\mathrm{C}[\text{"c"}, \text{"co"}] = \max(\mathrm{C}[\emptyset, \text{"co"}], \mathrm{C}[\text{"c"}, \text{"c"}]) = \max(0,1) = \mathbf{1}$
5) $\mathrm{C}[\text{"ca"}, \text{"c"}] = \max(\mathrm{C}[\text{"c"}, \text{"c"}], \mathrm{C}[\text{"ca"}, \emptyset]) = \max(1,0) = \mathbf{1}$
6) $\mathrm{C}[\emptyset, \text{"co"}] = \mathbf{0}$
7) $\mathrm{C}[\text{"c"}, \text{"c"}] = \mathrm{C}[\emptyset, \emptyset] + 1 = \mathbf{1}$
8) $\mathrm{C}[\text{"ca"}, \emptyset] = \mathbf{0}$

$\therefore\ the\ length\ of\ the\ LCS\ between\ X\ and\ Y = 3$

Represented as an array, the solution to the LCS algorithm is shown in the table below. Any index that contains only a numerical value is the final result of the max () function and indicates no match between the strings, while any index that shows a calculation indicates a match between the strings. Therefore, it is clear that these two strings match at the characters c, r and d. Finally, the value of index C [4, 4] is the length of the LCS (length 3).

Table 8: The solution to the LCS algorithm for "card" and "cord".

	Ø	**c**	**a**	**r**	**d**
Ø	0	0	0	0	0
c	0	1 + C[0,0] = 1	1	1	1
o	0	1	1	1	1
r	0	1	1	1+ C[2,2] = 2	2
d	0	1	1	2	1 + C[3,3] = 3

The LCS algorithm is implemented in this research to analyze the standardized SNOMED CT IDs of patient CP executions. All SNOMED CT IDs for a single patient's CP execution are concatenated to a single chain of IDs. This chain is an amalgamation of all of the interventions that take place in an execution of the CP for a patient. Instead of comparing two strings for the LCS on a character by character basis, two intervention chains are compared on an

ID by ID basis. One of these chains will always be the "ideal" execution of the CP, with no variance or other impediments. The other chain will be the actual execution of the CP, which may include sources of variance, repetition or other impediments not expected under normal execution. Given an ideal execution chain,

W = ("123-456", "123-101", "123-111", "123-765")

And an actual execution chain,

Z = ("123-456", "000-111", "123-111", "123-755")

The solution to the LCS algorithm, represented in array form is given in the table below. It is clear from these results that the LCS is ("123-456", "123-111") and its length is 2. It is also important to note that in this process, the entire hyphenated code must match. It is not sufficient for only the first or second component of the code to match the ideal execution.

Table 9: The solution to the LCS algorithm for 2 SNOMED CT ID chains.

	∅	**123-456**	**123-101**	**123-111**	**123-765**
∅	0	0	0	0	0
123-456	0	1 + C[0,0] = 1	1	1	1
000-111	0	1	1	1	1
123-111	0	1	1	1+ C[2,2] = 2	2
123-755	0	1	1	2	2

The pseudocode for the LCS algorithm, adapted to analyze SNOMED CT IDs, is shown below.

FUNCTION LongestCommonSubsequence

Input: *patientID*

OPEN clinical pathway execution output file
GET the *actual output row* that CONTAINS the inputted *patientID*
OPEN clinical pathway ideal execution output file
GET the *ideal execution output row*
CALL createExecutionSequence for *actual output row*
CALL createExecutionSequence for *ideal output row*
CALCULATE the number of events in the actual execution
CALCULATE the number of events in the ideal execution
DISPLAY the number of events in the actual execution
DISPLAY the number of events in the ideal execution
CALL calculateLongestCommonSubsequence
DISPLAY the longest common subsequence between these two sequences
DISPLAY the length of the longest common subsequence
CLOSE clinical pathway execution output file
CLOSE clinical pathway ideal execution output file

```
ENDFUNCTION LongestCommonSubsequence

FUNCTION createExecutionSequence
Input: executionRow, count, isActualExecution
    FOR EACH element IN executionRow
        IF element = a snomed ct id
            DISPLAY element
            IF isActualExecution is TRUE
                ADD element to LIST of actual executions
            ENDIF
            ELSE
                ADD element to LIST of ideal executions
        ENDIF
        INCREMENT count
    ENDFOR
    RETURN count
ENDFUNCTION createExecutionSequence

FUNCTION calculateLongestCommonSubsequence
Input: sequence A, sequence B
    CALCULATE length of A
    CALCULATE length of B
    CREATE 2D array C of size [length of A +1, length of B +1]
    FOR EACH row of C
        FOR EACH column of C
            IF current column index is 0 or current row index is 0
                C [current row index, current column index] = zero
            ENDIF
            ELSEIF A at current row index-1 = B at current row index-1
                C [current row index, current column index] =
                C [current row index-1, current column index-1] + 1
            ELSE
                C [current row index, current column index] =
                max(C [current row index-1, current column index],
                C [current row index, current column index-1])
        ENDFOR
    ENDFOR
    RETURN C [length of A +1, length of B +1]
ENDFUNCTION
```

In practice the LCS algorithm can be applied to the output file of Figure 51. This output file has been edited to display 3 interventions and their SNOMED CT IDs. In its complete form, the output file will contain the following data: Patient ID, Intervention Name, Intervention ID, Intervention Start Time, Intervention End Time, Event Outcome, Outcome ID and Observation Value. An example row of this output is provided in Figure 50.

1	Screening for Dysphagia	422504002-431765005	12-14-2019 19:30:31	12-14-2019 19:55:31	Successful	385669000	Pass

Figure 50: The output for intervention Screening for Dysphagia.

Screening_for_Dysphagia	422504002-431765005	Discussion_about_changes_in_lifestyle	422504002-223488008	Provision_of_educational_material	422504002-445283009
Screening_for_Dysphagia	422504002-431765005	Discussion_about_changes_in_lifestyle	422504002-223488008	Provision_of_educational_material	422504002-445283009
Screening_for_Dysphagia	422504002-431765005	Discussion_about_changes_in_lifestyle	422504002-223488008	Provision_of_assistive_equipment	422504002-228610003
Screening_for_Dysphagia	422504002-431765005	Assessment_of_substance_withdrawal	422504002-711008001	Provision_of_educational_material	422504002-445283009
Screening_for_Dysphagia	422504002-431765005	Discussion_about_changes_in_lifestyle	422504002-223488008	Provision_of_educational_material	422504002-445283009
Assessment_of_substance_withdrawal	422504002-711008001	Provision_of_educational_material	422504002-445283009	Evaluation_procedure	422504002-386053000
Screening_for_Dysphagia	422504002-431765005	Discussion_about_changes_in_lifestyle	422504002-223488008	Provision_of_educational_material	422504002-445283009
Screening_for_Dysphagia	422504002-431765005	Provision_of_assistive_equipment	422504002-228610003	Provision_of_educational_material	422504002-445283009
Screening_for_Dysphagia	422504002-431765005	Assessment_of_substance_withdrawal	422504002-711008001	Evaluation_procedure	422504002-386053000
Discussion_about_changes_in_lifestyle	422504002-223488008				

Figure 51: The output file for the Stroke CP.

The original output file documented the Assessment CP Events of Day 1 of the stroke CP ontology (totaling eight interventions) for ten patients. These output chains where analyzed using the LCS algorithm and the following results were obtained. Of special interest are the percentages of events that occur during actual execution compared to ideal execution (Compliancy Score). From these results, it becomes obvious that not every patient execution includes all ideal CP events. This alone gives an indication if the CP is being followed correctly. In this output file, compliance results vary from 12.5% to 100%. Another area of interest is the percent of actual CP events that are part of LCS (Ordering Score). This score indicates if the events actually occurring are happening in order (in respect to the ideal execution). These metrics highlight the strength of the LCS algorithm in CP auditing. By applying the LCS algorithm to the standardized codes of CP execution output files, an overall sense of CP compliancy can be gained. These analytical techniques would not be possible without a standardized CP ontology.

Table 10: LCS Algorithm output.

Patient	Number of Events	LCS Length	Compliancy Score (%)	Ordering Score (%)
1	8	8	100	100
2	8	8	100	100
3	6	5	75	83.3
4	5	4	62.5	80
5	3	3	37.5	100
6	3	2	37.5	66.7
7	6	6	75	100
8	4	3	50	75
9	3	3	37.5	100

10	1	1	12.5	100

The LCS algorithm results can also indicate the most common types of CP progression, based on the frequency of certain LCS lengths. Figure 52 presents the most common LCS lengths for the previous group of ten patients. Since the maximum length of the LCS can only be 8 in this scenario, that is the maximum horizontal axis value. The most popular LCS length by patient count is 3, followed by 8. Presentation of the LCS in this format can lead to focused analytical and auditing questions such as why certain LCS lengths are most popular. These results are the first step in grouping patients by LCS length and determining potential common characteristics between patients in these groups.

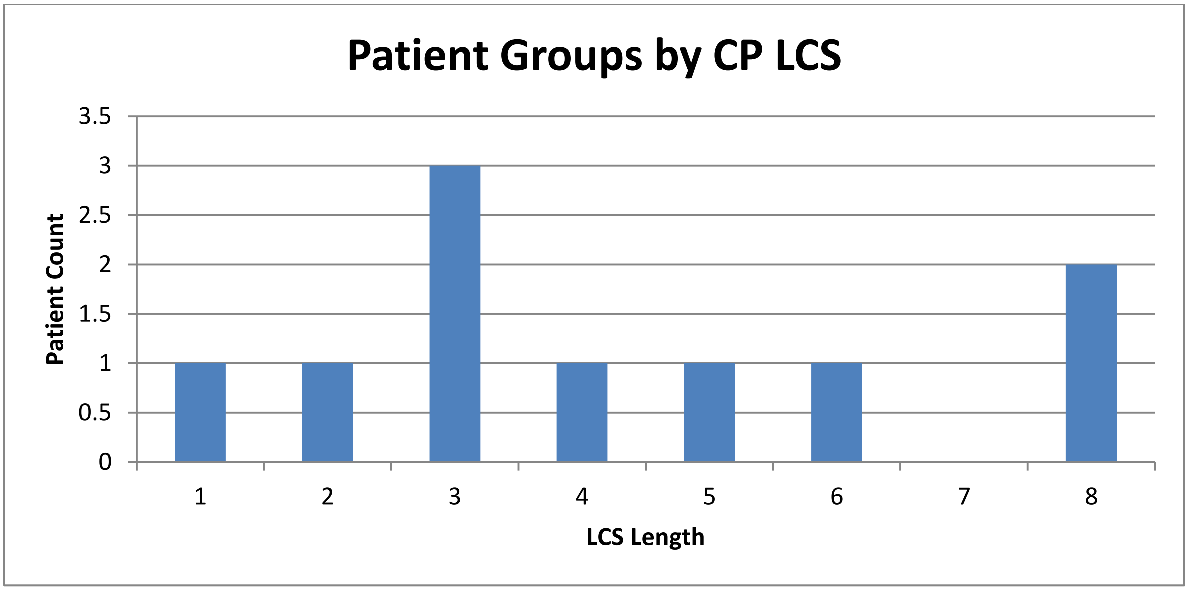

Figure 52: Patient Groups by CP Longest Common Subsequence.

4.2.2 Spatial Analysis – Longest Common Subsequence in 3D Space

The concept of the Longest Common Subsequence can also be applied in higher dimensionality. In the case of this research, it is applied in 3 dimensional space. The execution of a clinical pathway will occur over a span of several days. It will also include events of several different categories. As seen in the stroke Clinical Pathway ontology. These events/interventions may also be repeated over the time span of the clinical pathway. Under these conditions, the execution of a clinical pathway can be expressed as a cube, with each cubic 'unit' being a specific intervention, belonging to a certain category, happening on a specific day. This modelling approach is strengthened by the current ontological structure given its foundations in the Meta, Time and SNOMED ontologies. As an example, category can be represented on the x-axis, intervention can be represented on the y-axis and time can be represented on the z axis. Specific cubic units can then be represented through Cartesian coordinates. For example, the intervention Screening for Dysphagia, which is of the Assessment category, occurring on Day 1 may be represented by the coordinate (0, 0, 0). However, the intervention Screening for Dysphagia occurring on Day 2 would be represented by the coordinate (0, 0, 1).

Each dimension could then be explored using the LCS algorithm. In the y direction, the order of events can be compared, for a specific x or z plane. This can be comparing all of the Events occurring for a specific day, or comparing the events in a specific category across all days. For example, if the expected events for Day 1 are the events documented in the table below, the subsequence "A-B-C-D-E-F-G-H-I" (when presented linearly) might be expected. This subsequence can then be compared to the actual sequence of events for a specific patient. Where in one dimension, the application of the LCS is limited to the temporal order of the events within the CP; in higher dimensionality, the LCS can be applied across categories, interventions and days.

Table 11: A subsection of a CP over a certain day.

Category X	Category Y	Category Z
Event C	Event F	Event I
Event B	Event E	Event H
Event A	Event D	Event G

A second analytical technique that can be used on Clinical Pathway cubes is statistical shape analysis. This analysis would focus on the differences or 'deformations' of each cube compared to an ideal cube of the expected Clinical Pathway Execution. Therefore, unlike the LCS algorithm, the actual contents of each dimension do not necessarily have to be known, only the overall resulting cube shape. One mathematical basis for this analysis would be to take a similar approach to Large Deformation Diffeomorphic Metric Mapping (LDDMM). Other statistical shape analysis techniques that could be used include Principal Component Analysis (PCA) and Point Distribution Model (PDM).

While LDDMM has its basis in medical imaging, there is potential to apply its algorithms on diffeomorphic differences to CP cubes. Similarly, PDM has its basis in computer vision and medical imaging [39] but it may also be applied to CP cube analysis. Of note is the concept of landmark points for CP cubes. These points could represent special significance on the CP cube.

For example, a landmark point could denote a crucial activity or potential exit point of a CP. A potential representation of a CP cube with landmark points is shown in Figure 53 below.

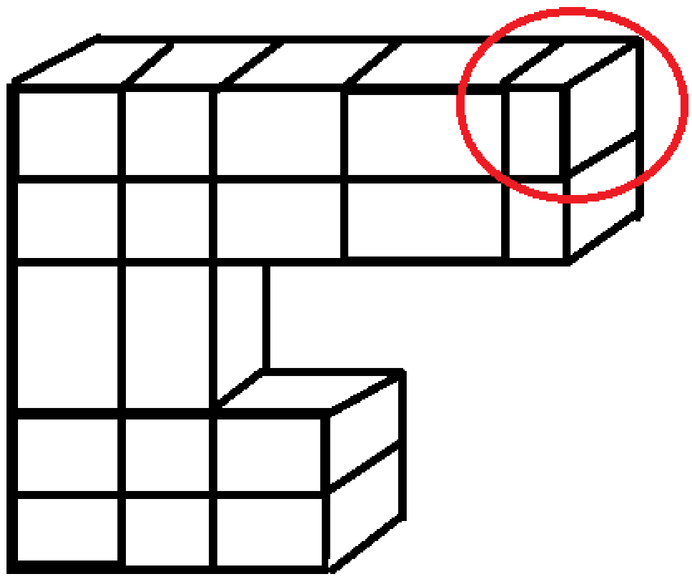

Figure 53: A potential CP cube with landmark point circled in red.

4.3 Discussion

Using the CPMS and the stroke CP ontology, this research is able to generate patient-tailored CP results. The creation of patient knowledge graphs from a default knowledge graph was not found in other literature and can be considered another contribution of this research. Retaining an original version of the stroke CP knowledge graph maintains the integrity of the knowledge base for future use and reference. This also contributes to the objective of sematic interoperability for this research. Distributed medical teams will be able to share a consistent ontology without any concerns about data loss or overwriting.

Data analytics were also performed on these patient-tailored CP results. Using the LCS algorithm, trends can be found in patient CP execution that may otherwise not be obvious. Common event paths, compliancy, and event ordering can all be discovered using the LCS algorithm. In a simulation of patient outcomes, it was found that the most common LCS length was 3. Therefore, further investigations can be performed to determine why the most common length is 3. These targeted investigations would not be possible without the use of the LCS algorithm. Finally, the LCS algorithm was considered in 3D space. Further investigation, as well as simulation, should be performed in this area.

In future work, additional conversations should be had with domain experts to determine other useful metrics that can be determined from the LCS algorithm. The value of compliancy score and ordering score should also be discussed with domain experts. Guided investigation of these experimental results is crucial to the acceptance of this modelling approach in the medical field.

5 Conclusion and Future Work

In conclusion, a generic CP ontology was successfully designed and developed, incorporating generalized processes, medical knowledge and temporal constraints. This ontology was extended to create ischemic stroke CP ontology, based on the paper ischemic stroke CP of Ottawa Hospital. A prototype system was then developed, which allowed the processing of the ischemic stroke CP and the generation of standardized output files during simulations of patient CP execution. This system is also capable of aiding in the future standardization of paper-based CP. The output of this system, in the form of patient specific knowledge graphs, was then analyzed using the LCS algorithm to reveal deeper trends in CP execution. These trends include compliancy to the expected CP progression and most common ordering of events.

Additions were made to the Meta CP ontology, Time ontology and Medical Knowledge ontology in order to combine them into a consistent, non-ambiguous and generic CP ontology. Through both top-down and bottom-up investigation based modelling techniques, ontology was developed for a stroke CP at Ottawa Hospital. It should be noted that this ontology can be expanded to a higher level of detail based on a combination of user need, expert recommendation and quality of the paper-based CP. However, this is easily performed because the design structure of the ontology allows for extensibility. OWL entities can be added to and deleted from the ontology without significant refactoring. This is possible due to the compartmentalization of the 3 base ontologies and the SWRL statement rule base.

The prototype CPMS system was used to simulate the progression of 10 patients through the stroke CP. The LCS algorithm was then successfully applied to the patient-tailored results of CP execution to identify the longest common subsequence between patient activities and expected activities. It was found that an LCS length of 3 was the most common. This is significantly lower than the activity length of ideal execution and therefore, raises guided questions about the CP. These guided questions, derived from analytical technique, are more useful than unguided investigation of raw output. In future work, additional conversations should be had with domain experts to determine other useful metrics that can be determined from the LCS algorithm. Furthermore, it should be a future aim of this research to present the prototype CPMS to stroke professionals for input and feedback.

Contributions of this research include expanding the detail and semantic meaning of the variance class; the development of interrelationship between event path and variance; the integration of admission, discharge and LOS within the time ontology; and the encapsulation of medical knowledge using the SNOMED CT standard. By expressing medical knowledge in a semantically rich hierarchy of terms, each individual term is given a greater significance and weight. A medical term no longer exists as just a stand-alone term but as a concept having interrelationships with other medical concepts. An overarching contribution of this research is the combination of 3 previously stand-alone ontologies into a larger CP ontology. This CP ontology is able to express any concept within the domain of Clinical Pathway as a certain entity, occurring at a certain time, referencing certain medical knowledge.

Some additional topics that can be explored for this research include the expansion of temporal reasoning and temporal-based SWRL statements, clustering of patient's based on common LCS characteristics, recommendation of potential CP interventions and 3D modelling of patient CP journeys. Furthermore, improvements can be made to the CPMS to provide additional functionality to users.

Specialized SWRL statements can be crafted in collaboration with domain experts to infer specific temporal knowledge for certain CP. This goes beyond generic statements to determine intervals of events, overdue interventions and LOS. It may involve querying disease specific interventions for internal temporal knowledge. For example, during toileting procedures it is only captured that planned voiding should occur every two hours. What is not currently captured is the internal timing of the toileting process. This involves transfer, voiding and activities such as washing hands. It could be valuable to capture the length of washing hands for additional data analytics purposes.

Additional research can also be performed in the area of the data analytics and the LCS algorithm. Using the already present data on LCS length, compliancy score and ordering score, patients can be grouped based on these results. A clustered patient data set based on LCS can then be compared to the same patient data set clustered on patient information such as age, weight or sex. Potential insights could then be made based on the similarity of clusters. Furthermore, similarity metrics could be gathered on patients through the use of information gain and similarity matrices. Finally, additional research can be performed in the area of spatial analysis of CP execution cubes. This would involve deeper analysis and application of LDDMM and PDM. The long-term objective of such research would be to generate 3D shapes corresponding to a patient's unique execution of a CP. These shapes would differ in both size and shape based on which events were performed during the execution of a CP. Entire subsections may be added or removed from a 3D model depending on if an activity was removed from, adhered to or added to the expected execution of the CP. Finally, each patient specific 3D model can then be compared to the expected or ideal 3D model of the CP. This will give a general representation of patient compliancy.

Improvements that can be made to the CPMS include graphical representation of a CP ontology, pathway recommendation, export of standardized terminology (SNOMED CT terms) and download/upload of other CP ontology. Providing users an overarching view of the CP ontology would be valuable for medical professionals who may want additional context or to see available pathways. Recommendation capabilities may be achieved in collaboration with other data analytics techniques. Whether through AI techniques or simpler classification, the CPMS could be developed to review historical pathways taken by patients. It may then compare the current patient to the historical patients and determine if there is a pathway available that will increase the probability of positive outcomes. Exportation of standardized terms to a text document for example would also be valuable to medical professionals who would like clear documentation of the list of standardized terms. Finally, allowing the upload and download of CP ontology would be logical considering that ontologies are typically hosted on the web. While security is a serious consideration when hosting sensitive patient data, simply uploading or downloading the generic CP ontology should be viable. This movement towards web-based ontology models would also be a step towards e-healthcare and distributed medicine.

6 References

1) National Health Expenditure Database, Canadian Institute for Health Information
2) Pearson S., Kleefield S., Soukop J., Cook E., Lee T., Critical pathways intervention to reduce length of hospital stay, 2001
3) Weingarten, S., Riedinger, M., Sandhu, M., Bowers, C., Ellrodt, A., Nunn, C. Greengold,. Can practice guidelines safely reduce hospital length of stay? Results from a multicenter interventional study. The American Journal of Medicine,105 (1), 33-40. 1998. doi:10.1016/s0002-9343(98)00129-6
4) Campbell, H., Hotchkiss, R., Bradshaw, Porteous, M. Integrated care pathways. BMJ 316, 133-137. 1998
5) Musen, M.A. The Protégé project: A look back and a look forward. AI Matters. Association of Computing Machinery Specific Interest Group in Artificial Intelligence, 1 (4), June 2015. DOI: 10.1145/2557001.25757003
6) Daniyal, A., Abidi S. R., Abidi S. S. R. NICHE Research Group, Faculty of Computer Science, Dalhousie University, Canada. Computerizing Clinical Pathways: Ontology-Based Modeling and Execution. 2009
7) Hu, Z., Li, J., Yu, H., Zhang, X., Suzuki, M., & Araki, K. Modeling of Clinical Pathways based on Ontology. 2009 IEEE International Symposium on IT in Medicine & Education. 2009. doi:10.1109/itime.2009.5236262
8) J.F. Allen; G. Ferguson. Actions and events in interval temporal logic In: Spatial and Temporal Reasoning. O. Stock, ed., Kluwer, Dordrecht, Netherlands, pp. 205-245. 1997. URL: http://dx.doi.org/10.1007/978-0-585-28322-7_7
9) Uschold, M. Demystifying OWL for the Enterprise. 2018. Morgan & Claypool Publishers.
10) Allen, J. F. Maintaining Knowledge about Temporal Intervals. Communications of the ACM, vol. 26, no. 11, 1 Jan. 1983, pp. 832–843., doi:10.1145/182.358434.
11) Canadian MIS Database, Canadian Institute for Health Information
12) Canadian MIS Database and Discharge Abstract Database, Canadian Institute for Health Information
13) Canadian MIS Database, Discharge Abstract Database and National Ambulatory Care Reporting System, Canadian Institute for Health Information

14) Fikes, R. E., and Nilsson, N. J. STRIPS: A new approach to the application of theorem proving to problem solving. Artificial Intelligence 2. 1971. 189-205.

15) Bruce, B. C. A model for temporal references and its application in a question answering program. Artificial Intelligence 3. 1972. 1-25.

16) McCarthy, J., and Hayes, P. Some philosophical problems from the standpoint of artificial intelligence. Machine Intelligence 4, Edinburgh U. Press, 1969.

17) National Health Expenditure Trends 1975 to 2019, Canadian Institute for Health Information

18) Hu, Z., Li, J., Yu, H., Zhang, X., Suzuki, M., & Araki, K. Modeling of Clinical Pathways based on Ontology. 2009 IEEE International Symposium on IT in Medicine & Education. 2009. doi:10.1109/itime.2009.5236262

19) K. Zander, Nursing case management: strategic management of cost and quality outcomes, J. Nurs. Adm. 18. 1988. 23–30.

20) Ye, Y., Jiang, Z., Diao, X., Yang, D., Du, G. An ontology-based hierarchical semantic modeling approach to clinical pathway workflows. 2009.

21) OWL. *OWL - Semantic Web Standards*, www.w3.org/OWL/. Accessed 15 Jan 2020

22) Klyne, G., Carroll, J., McBride, B. RDF 1.1 Concepts and Abstract Syntax. 2014. https://www.w3.org/TR/rdf11-concepts/. Accessed 15 Jan 2020

23) SNOMED International, 5 Step Briefing, http://www.snomed.org/snomed-ct/five-step-briefing. Accessed 15 Jan 2020

24) Hobbs, J., Pan F. Time Ontology in OWL. 2017. https://www.w3.org/TR/owl-time/. Accessed 15 Jan 2020

25) Chang, Phei Lang, et al. Effects Of Implementation Of 18 Clinical Pathways On Costs And Quality Of Care Among Patients Undergoing Urological Surgery. The Journal of Urology, 1999, pp. 1858–1862., doi:10.1097/00005392-199906000-00038.

26) Wakamiya, Shunji, and Kazunobu Yamauchi. A New Approach to Systematization of the Management of Paper-Based Clinical Pathways. Computer Methods and Programs in Biomedicine, vol. 82, no. 2, 2006, pp. 169–176., doi:10.1016/j.cmpb.2006.03.001.

27) Afandi, Radhi et al. Ontology Development in Patients Information System for Stroke Rehabilitation CEUR Workshop Proceedings vol. 2137, 2017

28) Jensen, Mark et al. The Neurological Disease Ontology Journal of Biomedical Semantics 2013

29) Ruttenberg, A. et al. OGMS Ontology. GitHub - OBOFoundry/Purl.obolibrary.org: A New System for Managing OBO PURLs, purl.obolibrary.org/obo/ogms.owl. Accessed 15 Jan 2020
30) Ye, Yan, et al. An Ontology-Based Hierarchical Semantic Modeling Approach to Clinical Pathway Workflows. Computers in Biology and Medicine, vol. 39, no. 8, 2009, pp. 722., doi:10.1016/j.compbiomed.2009.05.005.
31) Ye, Yan, et al. An Ontology-Based Hierarchical Semantic Modeling Approach to Clinical Pathway Workflows. Computers in Biology and Medicine, vol. 39, no. 8, 2009, pp. 724., doi:10.1016/j.compbiomed.2009.05.005.
32) Owlcs. "Owlcs/Owlapi." GitHub, 2019. github.com/owlcs/owlapi/. Accessed 15 Jan 2020
33) Fan, Wu, et al. Constructing Clinical Pathway Ontology to Incorporate Patient State, Intervention and Time. 2011 4th International Conference on Biomedical Engineering and Informatics (BMEI), 2011, pp. 1, doi:10.1109/bmei.2011.6098597.
34) Fan, Wu, et al. Constructing Clinical Pathway Ontology to Incorporate Patient State, Intervention and Time. 2011 4th International Conference on Biomedical Engineering and Informatics (BMEI), 2011, pp. 2, doi:10.1109/bmei.2011.6098597.
35) Fan, Wu, et al. Constructing Clinical Pathway Ontology to Incorporate Patient State, Intervention and Time. 2011 4th International Conference on Biomedical Engineering and Informatics (BMEI), 2011, pp. 3, doi:10.1109/bmei.2011.6098597.
36) Liu, Junjian, et al. An Ontology-Based Real-Time Monitoring Approach to Clinical Pathway. 2014 7th International Conference on Biomedical Engineering and Informatics, 2014, p. 756, doi:10.1109/bmei.2014.7002874.
37) Liu, Junjian, et al. An Ontology-Based Real-Time Monitoring Approach to Clinical Pathway. 2014 7th International Conference on Biomedical Engineering and Informatics, 2014, p. 757, doi:10.1109/bmei.2014.7002874.
38) Liu, Junjian, et al. An Ontology-Based Real-Time Monitoring Approach to Clinical Pathway. 2014 7th International Conference on Biomedical Engineering and Informatics, 2014, p. 759, doi:10.1109/bmei.2014.7002874.
39) T. F. Cootes. Statistical models of appearance for computer vision. 2004
40) Hu, Zhen, et al. Ontology-Based Clinical Pathways with Semantic Rules. Journal of Medical Systems, vol. 36, no. 4, 2011, pp. 2203–2212., doi:10.1007/s10916-011-9687-0.

41) Hu, Zhen, et al. Ontology-Based Clinical Pathways with Semantic Rules. Journal of Medical Systems, vol. 36, no. 4, 2011, pp. 2204, doi:10.1007/s10916-011-9687-0.

42) Alexandrou, Dimitrios A., et al. A Holistic Environment for the Design and Execution of Self-Adaptive Clinical Pathways. 2009 9th International Conference on Information Technology and Applications in Biomedicine, 2009, doi:10.1109/itab.2009.5394358.

43) Abidi, Samina Raza, and Syed Sibte Raza Abidi. An Ontological Modeling Approach to Align Institution-Specific Clinical Pathways: Towards Inter-Institution Care Standardization. 2012 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS), 2012, doi:10.1109/cbms.2012.6266392.

44) Omaish, M., Abidi, S., Abidi S. S. R., Ontology-based Computerization of Acute Coronary Syndrome Clinical Guideline for Decision Support in the Emergency Department 2012 European Federation for Medical Informatics, doi: 10.3233/978-1-61499-101-4-437