Expanding knowledge: Linking a theoretical hospital care environment Ontology on the Semantic Web

Abstract. Studies suggest that health clinicians have little input into ontology design depicting their environment. Ontologies designed by non-clinicians often display a limited knowledge of the more traditional sense of ontology, the study of our 'reality'. Linking Of Data (LOD) technology may provide 'tools of choice' for future informatics clinicians to translate their environments so they are meaningful to computer science and other disciplines. An ontology reflecting a theoretical hospital care environment is constructed and linked using LOD standards. The health care environment contains concepts such as nurses, doctors and medications. Links of commonality between concepts add to and expand their meaning forming a computer-readable map tracing lines of care. It was found concepts reflecting a patient's care in the theoretical environment could be linked to vocabularies and descriptions residing in BioPortal, MeSH and DbPedia repositories. Links could be traversed and resources retrieved which added to and expanded knowledge of a concept. Time versus distance to retrieve resources from the ontology was measured. It was found DpPedia had a relatively 'flat' response time compared to BioPortal.

Keywords. Ontology, Clinical, Environment, Semantic

Introduction

There has been a lack of input by health practitioners translating their clinical environment to a form understandable to non-clinicians and computers. It has been suggested that Linking Of Data (LOD) tools such as ontologies and the resource Description Framework (RDF) are not in common use outside the computer science domain [1]. As a consequence, a common limitation computer science encounters when designing clinical ontologies is a lack of specialised knowledge required to define these specialised environments. For example, a 'smart hospital' in computer science literature is often depicted by an ontology assisting in the placement of embedded technology [2] or electronic health documentation [3] in the clinical environment. Modelling medical pathways and defining terminology for hospital scenarios requires clinical expertise [4].

To describe their world, clinicians working within the clinical informatics discipline require tools to design ontological representations. This study contributes to understanding of the clinical environment by applying LOD tools and standards underpinning the Semantic Web to describe an environment. The study asks "can a clinician build a simple theoretical model of his/her environment and link concepts therein to the Semantic Web?" An ontological representation of the hospital care environment built by informatics clinicians may assist stakeholders to:

- Achieve greater understanding the care environment.
- Expand concepts in the ontology by linking to more detailed knowledge and specifications.

The aim of the study is to use LOD tools to construct ontology which describes a theoretical hospital nurse care environment and link it to vocabularies on the Semantic Web. A vocabulary in the Semantic Web context is an ontology which contains agreed upon definitions of terms that are used to represent (and expand) data [8].

The study applies Gruber [5]'s artificial intelligence definition of ontology where ontology is a specification and description of 'real world' concepts and relationships, particularly reflecting some 'domain of interest'. Ontology's primary purpose in the computer science context is to enable knowledge sharing and data reuse by linking similar concepts [6]. Sir Tim Berners-Lee [7] described a World Wide Web of interconnected ontologies which may be joined together by common concepts to form a 'Semantic Web' (SW).

1. Theoretical underpinnings of ontology's structure

1.1. Donabedian's Framework

The study's ontology is ordered through Donabedian [8]'s overarching Structure Process Outcome (SPO) paradigm in Figure 1. Donabedian [8]'s framework for measuring the quality of care through a structure-process-outcome relationship underline many health quality studies which explore the nursing clinical environment through quality indicators [9] [10] [11].

Figure 1. Donabedian's Framework.

1.2. Nursing Role Effectiveness Model

Building on Donabedian [8]'s SPO, Irvine, Sidani, and McGillis-Hall [12] developed the Nursing Role Effectiveness Model (NREM) which underpins this study. The NREM proposes that the relationship between structure, process and outcome quality indicators can be enhanced by dividing the process domain into three 'nursing roles' namely:

- The independent role, these are processes in which a nurse acts autonomously.
- The dependent role encompasses actions requiring a physician's order and the
- Interdependent role covers tasks in which the nurse liaises with allied health.

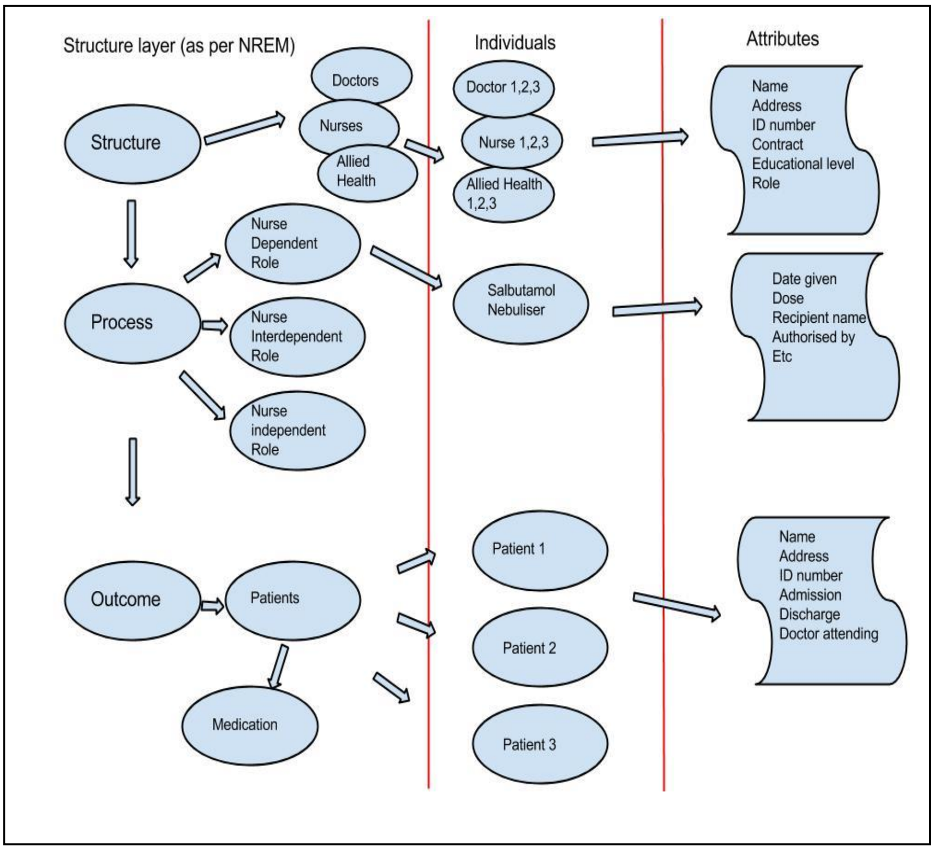
2. Ontology Construction and linking Methodology

2.1. Scope of the environment

It is impossible to reflect every nuance and relationship in a clinical environment. The scope was limited to 4 patients 4 doctors and 3 nurses. Each individual had attributes including name, role, designation and education level. Doctors were responsible for individual patients and nurses cared for patients through NREM roles. Patients were limited to one medication and one diagnosis. The clinical environment describes multi-disciplinary teams interacting for the purpose of achieving desired patient outcomes as described by Doran, et al. [13].

The environment was divided into three sections as can be seen in Figure 2's conceptual depiction. The left most pane orders the structure of the ontology through the NREM framework. The framework in turn orders the abstraction of individuals representing doctors, nurses, and patients. Not shown in this simplified diagram are abstractions of medications, diagnosis and interventions. All data and object attributes describing individuals are depicted in the right pane.

Figure 2. Simplified overview of the clinical ontology.



2.2. Ontology Construction using triples to facilitate linking of data

The ontology was constructed using the Protégé [15] development platform which has the capacity to code various 'flavours ' of the World Wide Web Ontology (OWL) language [16]. The study used 'OWL-DL' as it is the most descriptive.

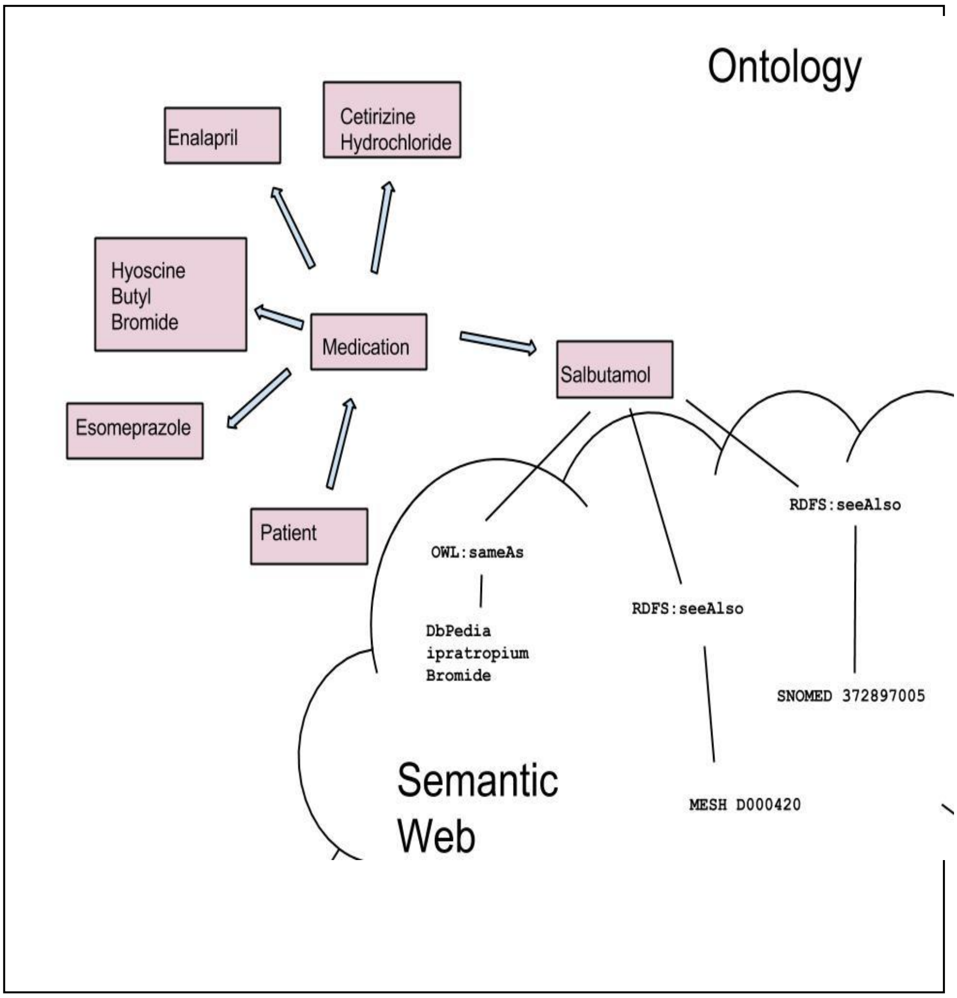
Owl facilitates linking of data on the Semantic Web by grouping three web addresses, Uniform Resource Indicators (URIs), together to form a ‘triple’. Each triple represents some concept and/or relationship. Triples can be compared to simple sentences which contain subject, predicate and object which either describe a resource or the relationship between resources. For example, ‘Bob knows Alice’ is a triple which can be represented as three URIs which link to various locations on the Web to further expand the concepts of Bob and Alice and their relationship.

2.3. Linking three concepts to vocabularies

One role of ontology on the Semantic Web is to expand knowledge by linking to other knowledge. Patient’s medications and diagnosis within the ontology are linked to external vocabularies which provide standardised descriptions of the concepts. A vocabulary in the Semantic Web context contains agreed upon definitions of terms that are used to represent data. Vocabulary links make data self-descriptive and enable Linked Data applications to understand and integrate data across the Semantic Web [14].

Three concepts in the ontology were linked to external vocabularies to expand their knowledge. Salbutamol, a bronchodilator and two diagnosis ‘insufficient gas exchange’ and ‘gout’ belonging to patients were linked to the Systematised Nomenclature of Medicine, SNOMED-CT [17] in the BioPortal [18] repository, DbPedia [19] and Medical Subject Heading (MeSH). All three vocabularies describe the concepts in some way and provided its respective identification code. This information now becomes part of our ontology’s knowledge about these concepts and enables further exploration along links beyond these vocabularies. Figure 3 is a conceptual depiction of Salbutamol ‘s connection from the ontology to three vocabularies on the Semantic web.

Figure 3: Graphic representation of Salbutamol in the ontology and external links to vocabularies.


Ontology
Enalapril
Cetirizine Hydrochloride
Hyoscine Butyl Bromide
Medication
Salbutamol
Esomeprazole
Patient
OWL:sameAs
DbPedia ipratropium Bromide
RDFS:seeAlso
RDFS:seeAlso
SNOMED 372897005
MESH D000420
Semantic Web

2.4. Exploration further afield

Sir Tim Berners-Lee's fourth rule of linked data states that links should lead people to discover even more things [7]. A serendipitous exploration along links radiating from the theoretical nurse care environment ontology to various vocabularies residing in repositories was conducted. Links that connect the three original vocabularies to even more vocabularies describing the same concept are explored. Time and distance to retrieve codes representing the diagnosis or medication are measured from all discovered vocabularies.

2.5. Discovering and measuring links to vocabularies

A standard code corresponding to a concept in the ontology is retrieved by following links to distant vocabularies. Distance and time taken to retrieve a code describing some resource from a distant vocabulary was charted by measuring time in seconds to retrieve it and distance from the starting point in our ontology to the vocabulary. This is not a strict performance measurement but more 'proof of concept'. Distance to a vocabulary from the originating ontology was counted by the number of 'hops' (vocabularies) a link passes through to get to its destination. Time in seconds was measured by stop watch as there is no application-based method available. Time is measured starting from when the URI is activated ending when the code describing a resource appeared.

The study used A graphite [20] based RDF browser [21] hosted by the University of Southampton UK to follow links to vocabularies using Sir Tim Berners-Lee's 'Follow Your Nose' retrieval guidelines [7]. These guidelines simply state that knowledge can be retrieved by dereferencing a URI, on arrival; the URI should provide links to more knowledge and so on. The study did not to use Protocol and RDF Query Language (SPARQL) (W3C, 2013) endpoint on the repositories to retrieve data . A SPARQL query relies on its triple pattern matching the triple pattern in the ontology to retrieve meaningful results. In reality, the user has to know patterns peculiar for each ontology prior to data retrieval. It was reasoned that a-priory knowledge of data structure defeats the purpose of serendipitous discovery.

3. Results

3.1. Concepts used as starting points in the ontology

Concepts describing two diagnoses and one medication belonging to two different theoretical patients were linked from the ontology to SNOMED-CT [17] in the BioPortal [18] repository, DbPedia [19] and Medical Subject Heading (MeSH) vocabularies. These provided expanded descriptions of the concepts and acted as 'starting points' for further discovery. Patient 'Gag Halfrunt' previously described medication, the bronchodilator Salbutamol, and his diagnosis of 'insufficient gas exchange' provides a base from which links could radiate out to distant vocabularies. Likewise, Patient 'Gogrilla Mincefriend' diagnosis of Gout provides a starting point.

3.2. Discovered vocabularies from the Salbutamol concept in order

The following tables show discovered vocabularies starting from their respective concept in the ontology, including the first three 'hard-wired' ones.

1) DbPedia
2) SNOMED
3) Medical Subject Headings (MESH)
4) US Veterans Health Administration National Drug File
5) US national Drug File reference Terminology
6) Logical Observation Identifier Names and Codes ontology
7) Chemical Entities of Biological Interest Ontology
8) US National Cancer institute Thesaurus
9) RxNorm Ontology
10) Physician data Query ontology
11) Galen Ontology

Table 1 shows each vocabulary's discovery number against the number of hops from the original ontology's concept of Salbutamol, time in seconds for retrieval and code retrieved from each vocabulary.

Table 1: Discovered vocabulary, number of 'hops' from Salbutamol starting point and time in seconds to retrieve the displayed code

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1	1	1	2	2	2	2	2	2	3	3
4.06	8.47	8.32	7.6	11.92	7.41	9.69	17.07	8.91	21.76	21.28

224 727 6	3728 9700 5	D0004 20	401879 6	C25170	MTHU 013596	CHE BI:25 49	C215	453	CDR00 000413 39	Salbuta molSul phate

3.3. Discovered vocabularies from the concept of 'Insufficient gas exchange' diagnosis in order

1) SNOMED
2) International Classification for Nursing Practice (ICNP) Ontology
3) Logical Observation Identifier Names and Codes Ontology

Table 2: Discovered vocabulary, number of 'hops' from 'Insufficient gas exchange' starting point and time in seconds to retrieve the displayed code

1	**2**	**3**
1	2	2
10.88	2.71	3.84
70944005	10001177	MTHU013234

3.4. Discovered vocabularies from the concept of 'Gout' diagnosis in order

1) SNOMED
2) International Classification of Diseases, Version 10 - Clinical Modification
3) US National Drug File – Reference Terminology
4) Coding Symbols for a Thesaurus of Adverse Reaction Terms

5) US National Cancer Institute Thesaurus

6) International Classification of Primary Care

7) Human Phenotype Ontology

8) Bone Dysplasia Ontology

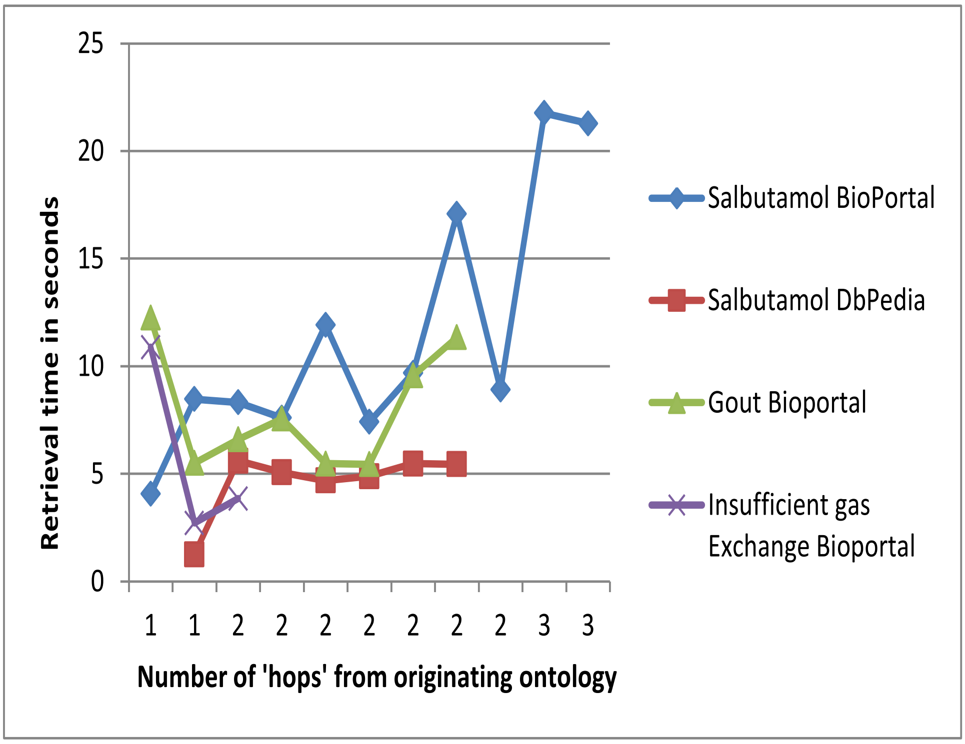
Table 3: Discovered vocabulary, number of ‘hops’ from ‘Gout’ starting point and time in seconds to retrieve the displayed code

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	2	2	2	2	2	2	2
12.25	5.50	6.60	7.56	5.47	5.44	9.56	11.34
90560007	M10	N0000001382	METGOUT	C34650	T92	HP:0001997	HP:0001368

3.5. Time versus distance per ontology repository and search

The study used Bioportal and DbPedia ontology repositories to conduct distance and time measurements to vocabularies. Strictly speaking DpPedia is not a repository but a representation of WikiPedia on the Semantic Web. Figure 4 shows a graph of hops versus response time for each search ordered as per repository.

Figure 4: *Graph of time vs distance in 'Hops' ordered by repository*



4. Conclusion

An OWL ontology was constructed in Protégé development platform by a nurse clinician which reflected a simple theoretical clinical environment. The environment was based on the Nursing Role Effectiveness Model (NREM) and contained concepts of people, medication, and diagnosis. Concepts were linked together according to theoretical 'real world' relationships, care relationships were determined by process roles described in the NREM.

Concepts representing things in a clinical nurse care environment were selected to be linked to concepts within external vocabularies. Linking of data was used to expand knowledge of a concept and link concepts to international standards represented as codes and descriptions in the vocabulary. Proof of concept was tested by a serendipitous 'follow your nose' search strategy measuring the time it took to retrieve codes representing the common concept from each vocabulary as it was discovered in two repositories, BioPortal and DbPedia.

References

[1] E. Little, "Current trends and Challenges in Ontological Research," presented at the A talk to the Ontological Research and its Applications to the Biomedical Domain Conference, New York, USA, 2009.

[2] S. Anand and A. Verma, "Development of Ontology for Smart Hospital and Implementation using UML and RDF," *International Journal of Computer Science Issues(IJCSI),* vol. 7, 2010.

[3] S. T. Rosenbloom, R. A. Miller, K. B. Johnson, P. L. Elkin, and S. H. Brown, "Interface terminologies facilitating direct entry of clinical data into electronic health record systems," *Journal of the American Medical Informatics Association,* vol. 13, pp. 277-288, 2006.

[4] M. Becker, C. Heine, R. Herrler, and K.-H. Krempels, "OntHoS—an Ontology for Hospital Scenarios," in *Applications of Software Agent Technology in the Health Care Domain*, ed: Springer, 2003, pp. 87-103.

[5] T. Gruber, "Every Ontology is a Treaty," *SIGSEMIS Bulletin: Interview for Semantic Web and Information Systems SIG of the Association for Information Systems,* vol. 1, 2004.

[6] L. Feigenbaum, I. Herman, T. Hongsermeier, E. Neumann, and S. Stephens. (2007, 19 March 2013). *The Semantic Web In Action*. Available: http://thefigtrees.net/lee/sw/sciam/semantic-web-in-action#citation

[7] T. Berners-Lee. (2006, 19July). *Notes on Linked Data*. Available: http://www.w3.org/DesignIssues/LinkedData.html

[8] A. Donabedian, "The quality of care: How can it be assessed?," *Journal of the American Medical Association,* vol. 260, pp. 1743-1748, 1988.

[9] O. Health. (2010, 11/5/2011). *project hobic report*. Available: http://www.health.gov.on.ca/english/providers/project/hobic/archive/report.pdf

[10] L. Aiken, S. Clarke, D. Sloane, J. Sochalski, and J. Silber, "Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction.," *Journal of the American Medical Association,* vol. 288, pp. 1987–1993, 2002.

[11] M. Chiarella and J. K. Roydhouse, "Hospital churn and casemix instability: implications for planning and educating the nursing workforce. ," *Australian Health Review,* vol. 35, pp. 95-98, 2011.

[12] D. Irvine, S. Sidani, and L. McGillis-Hall, "Linking outcomes to nurses' roles in health care," *Nursing Economics,* vol. 16, pp. pp58-64, 1998.

[13] D. Doran, B. Mildon, and S. Clarke, "Towards a national report card in nursing: A knowledge synthesis," *Nursing Leadership,* vol. 24, pp. 38-57, 2011.

[14] T. Heath and C. Bizer, "Linked data: Evolving the web into a global data space," *Synthesis lectures on the semantic web: theory and technology,* vol. 1, pp. 1-136, 2011.

[15] Protege. (2011, 12/6/2011). *Protege ontology editor*. Available: http://protege.stanford.edu/

[16] W3C. (2013, 4 october). *Semantic web standards*. Available: http://www.w3.org/standards/semanticweb/

[17] SNOMED-CT. (2013, 10 October). *Systematised Nomenclature of Medicine*. Available: http://bioportal.bioontology.org/ontologies/SNOMEDCT?p=classes

[18] Bioportal. (2013, 7 October). *Bioontology development platform*. Available: http://bioportal.bioontology.org/

[19] DBpedia. (2013, 24 October). *The distributed encyclopedia*. Available: http://wiki.dbpedia.org/About

[20] C. Gutteridge. (2013, 16 December). *Graphite PHP Linked Data Library*. Available: http://graphite.ecs.soton.ac.uk/

[21] C. Gutteridge. (2013, 18 December). *Quick and Dirty RDF browser*. Available: http://graphite.ecs.soton.ac.uk/browser/