



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

The 2nd International Workshop on Pervasive and Ambient Applications, Systems and Technologies for Healthcare (PASTH 2012)

HAIKU: A Semantic Framework for Surveillance of Healthcare-Associated Infections

Arash Shaban-Nejad^{a,*}, Alexandre Riazanov^b, Katia M. Charland^a, Gregory W. Rose^c, Christopher J.O. Baker^b, Robyn Tamblyn^a, Alan J. Forster^c, David L. Buckeridge^a

^a*McGill Clinical & Health Informatics, Department of Epidemiology and Biostatistics, McGill University, 1140 Pine Avenue West, Montreal, Quebec H3A 1A3, Canada*

^b*Department of Computer Science, University of New Brunswick, Saint John, NB Canada*

^c*Department of Medicine, University of Ottawa, Ottawa, ON Canada*

Abstract

Healthcare-Associated Infections (HAI) impose a substantial health and financial burden. Surveillance for HAI is essential to develop and evaluate prevention and control efforts. The traditional approaches to HAI surveillance are often limited in scope and efficiency by the need to manually obtain and integrate data from disparate paper charts and information systems. The considerable effort required for discovery and integration of relevant data from multiple sources limits the current effectiveness of HAI surveillance. Knowledge-based systems can address this problem of contextualizing data to support integration and reasoning. In order to facilitate knowledge-based decision making in this area, availability of a reference vocabulary is crucial. The existing terminologies in this domain still suffer from inconsistencies and confusion in different medical/clinical practices, and there is a need for their further improvement and clarification. To develop a common understanding of the infection control domain and to achieve data interoperability in the area of hospital-acquired infections, we present the HAI Ontology (HAIO) to improve knowledge processing in pervasive healthcare environments, as part of the HAIKU (Hospital Acquired Infections – Knowledge in Use) system. The HAIKU framework assists physicians and infection control practitioners by providing recommendations regarding case detection, risk stratification and identification of diagnostic factors.

© 2011 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of [name organizer].
Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Knowledge processing in pervasive healthcare environments; Semantic Web; Recommender Systems; Biomedical Ontologies; Hospital acquired infection, Web Ontology Language (OWL)

* Corresponding author. Tel.: +1-514-934-1934 ext. 32986; fax: +1-514-843-1551.
E-mail address: arash.shaban-nejad@mcgill.ca

1. Introduction

Healthcare-Associated Infections (HAIs) are a common complication of patient care, and are caused by exposure to various bacteria, fungi, or viruses in settings such as hospitals, long-term care facilities, rehabilitation centers, and community clinics [1]. Some of the most common HAI's are bloodstream infections, urinary tract infections (UTIs), surgical site infections (SSIs), pneumonia, tuberculosis, *Clostridium difficile* infections, methicillin resistant *Staphylococcus aureus* infections (MRSA), severe acute respiratory syndrome, gastroenteritis, pertussis, and ventilator associated pneumonia (VAP). HAIs impose tremendous direct (e.g. hospitalization, utilities, devices, medications, lab tests, etc.), indirect (e.g. mortality, lost/wages, etc.), and intangible (e.g. anxiety, disability, job loss, pain, etc.) costs [2]. In the United States, HAIs are estimated to result in "direct annual costs of \$35.7 - \$45.0 billion for hospitals and healthcare facilities with combined medical costs of \$20,549 - \$25,903 per infection occurrence" [2]. Many of the HAIs and related costs can be effectively prevented through proper surveillance and control plans. In most hospitals in Canada, infection control practitioners rely upon routine microbiology test results to flag patients with a possible HAI. Once a patient is flagged, staffs gather the patient's data from multiple sources to determine whether a HAI has occurred [3]. Automating this data gathering and integration is currently very challenging because the data in each of the source systems lack the contextual information necessary for automated integration and reasoning. Standardization and consistent integration of disparate biomedical data, however, would enable automated data integration and reasoning. This capacity could improve the decision-making process by allowing infection control staff to querying over several heterogeneous knowledge sources. In this context, a knowledge-based approach to surveillance of Healthcare-Associated Infections could advance data and knowledge integration, and thereby decrease the incidence of HAIs and facilitate effective resource allocation to improve patient care and enhance overall health systems performance.

The increasing success in the application of knowledge modeling to support automated data integration and reasoning has spurred an explosion in biomedical ontology development and a trend towards institutionalizing ontologies and standardizing representation language within information systems serving the biomedical sciences. Ontologies capture the domain knowledge by defining concepts, relationships, individuals, rules and axioms. Our research mainly focuses on the four most common HAIs (bloodstream infection, UTI's, SSIs, and pneumonia), often referred to as the "Big Four", due to the high morbidity and mortality rates.

HAIs (more specifically, nosocomial infections) can be defined as one type of "hospital adverse event" as described in the Ontology of adverse event (OAE) [4]. It is also defined in the Infectious Disease Ontology (IDO) [5] as an infection "resulting from a transmission process that unfolds in a hospital". So as a first step, we have defined a set of properties to distinguish between HAIs and Community Acquired Infections (CAIs). In our early work [6], we demonstrated the potential of the HAIKU framework (Figure 1) in the domain of clinical intelligence for case detection, risk/causative factor identification/evaluation, and diagnostic factor identification/evaluation with a focus on surgical site infections (SSIs). We have also explored [7] the possibility of using SADI [8] Semantic Web services for semantic querying of clinical data and reported preliminary progress on prototyping a semantic querying infrastructure for the surveillance of, and research on hospital-acquired infections. The rest of this paper proceeds as follows. In Section 2 we will describe our conceptual modelling method and ontology development approach. We represent the usability of our ontological framework in the domain in Section 3 through a set of semantic queries. The paper concludes a discussion of the implications of our study and possible future work directions.

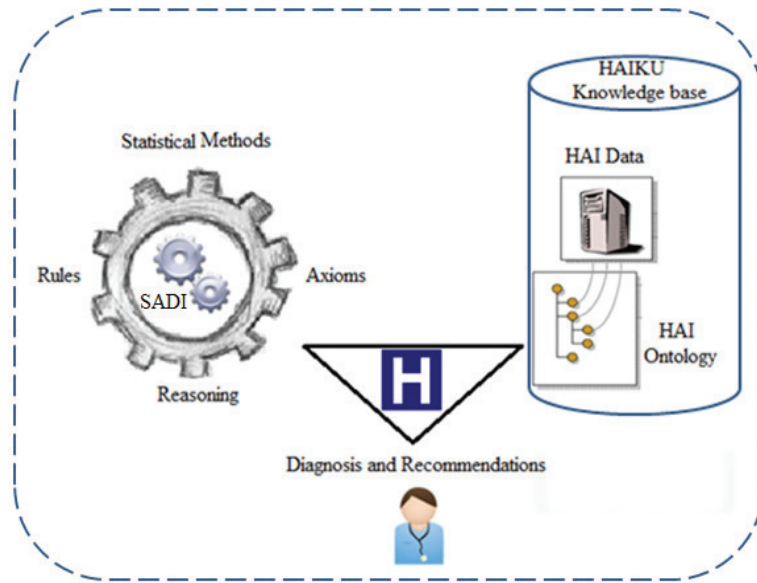


Fig. 1. An abstract representation of the HAIKU framework, which assist physicians and infection control practitioners by providing recommendations used for case detection, risk stratification and diagnostic factor identification. The SADI semantic web services can populate the relevant HAI Ontology predicates based on the semantic mapping between the data warehouse schema and the ontology. Also the results of statistical inference from epidemiological and clinical data can be combined with the logical inference to enable inductive reasoning about formal knowledge in the ontological structure.

2. The HAI Ontology Design

The HAIKU knowledge base has been developed using the W3C standard Web Ontology Language (OWL 2.0)¹, which enables us to define complex concepts and the properties in the domain in a highly expressive manner. The HAI ontology represents general knowledge about HAI and should be reusable in multiple health care contexts. To adapt and customize the ontology for a specific use, however, concepts in the ontology must be mapped to data elements within information systems in a specific hospital. In this stage of our research project, we therefore aligned and mapped the concepts in the ontology to terms used in the local data warehouse (Figure 2) in the Ottawa Hospital (TOH). We used a subset of the data warehouse containing data for 715 cardiac surgery patients who had 6,132 encounters, received 12,275 diagnoses, and underwent 6,029 procedures [9] at the University of Ottawa Heart Institute between 2004-2007.

2.1. The Conceptual Model

Our conceptual model has been created based on the existing knowledge sources, textual resources and the relevant literature including the data provided by the National Nosocomial Infections Surveillance (NNIS) System [10] (from Jan 1992 - June 2004), the 2012 CDC/NHSN Surveillance definition of Healthcare-Associated Infections [11], 2007 CDC Guideline for Isolation Precaution [12], and several patient data from TOH. Also we have used several information resources (e.g., the databases containing

¹ <http://www.w3.org/TR/owl-overview/>

information on hospital morbidity and discharge abstracts), existing bio-ontologies (e.g., SNOMED CT², ICD-9 [13], HL7 (<http://www.hl7.org/>), FMA [14], CheBI [15], Infectious Disease Ontology (IDO)).

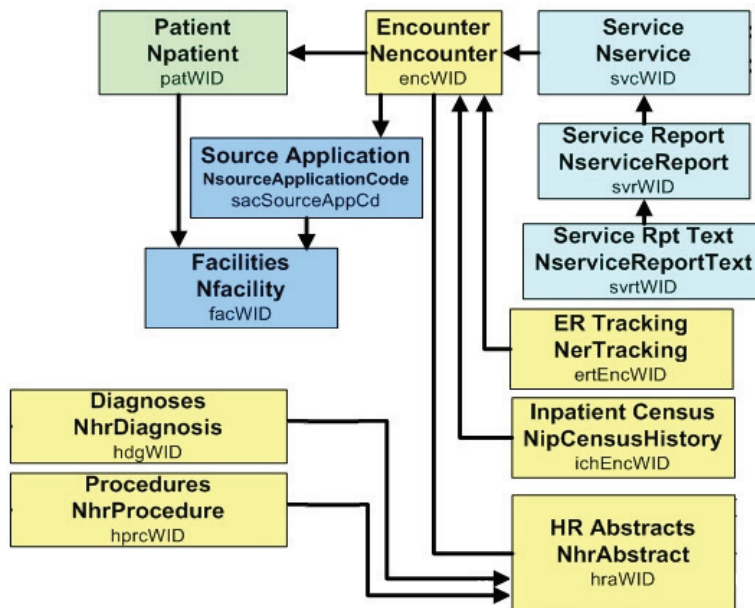


Fig. 2. A partial view of the major entities in the Ottawa Hospital Data Warehouse data model. The three major components are Patient (represents patient level information), Encounter (captures encounter information), and Service (contains generic information regarding services received by patients during encounters). The services represented in the diagram may include radiology, pharmacy, transcription, laboratory services, etc.

2.2. SemanticScience Integrated Ontology (SIO) and the Extra Simple Time Ontology (ESTO)

In order to facilitate a broad semantic interoperability between the HAI Ontology and other related standards and ontologies we are aligning our ontology with SIO³, which offers an integrated upper level ontology (types, relations) for consistent knowledge representation across physical, processual and informational entities [16]. To support temporal knowledge management we have developed the Extra Simple Time Ontology (ESTO)⁴ based on Allen's temporal logic [17], which supports defining time instants, proper time intervals, and infinite time intervals. As an example, the following query can control if a specific time interval is subsumed by another interval.

`?DiagnosisTime esto:during ?HospitalizationPeriod`

For more information on ESTO we refer the readers to our recent paper [7].

2.3. Incorporating Statistical Inference

In existing surveillance approaches, the incidence of HAIs has been very difficult to document using statistical models alone, because statistics have been based upon different definitions. We combine the

² <http://www.ihtsdo.org/snomed-ct/>

³ SemanticScience Integrated Ontology (SIO)

⁴ <http://unbsj.biordf.net/ontologies/extra-simple-time-ontology.owl>

results of statistical inference from epidemiological and clinical data with formal knowledge to enables inductive reasoning about formal knowledge in the ontological structure. We also use these statistical data for enriching the background knowledge used for defining relationships (causal or associative), rules and other axioms within the HAIKU conceptual model. For example, admitted patients have different levels of susceptibility to acquiring an infection following exposure to a given infectious agent. Statistical models can be constructed to assist in the case detection process by estimating the probability that a patient is a suspect, probable or confirmed case, based on various criteria such as identification factors (e.g. readmission to hospital for therapy of SSI), risk factors (e.g. old age, or prolonged operative time) (Figure 3) and clinical findings (i.e. lab order, lab result, imaging, procedures, drug order, etc.) [9].

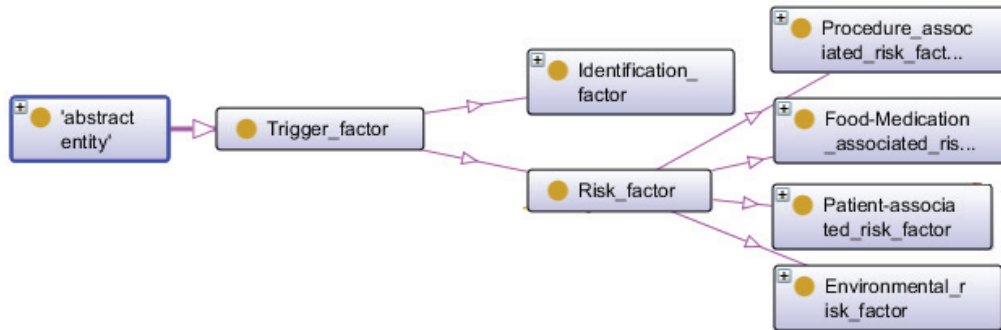


Fig. 3. The general overview of the Trigger factors taxonomy (visualized using OntoGraph⁵). The identification factors, which are usually identified after the development of HAI, can be used to identify confirmed cases, and risk factors can be used to infer the status of patients as a suspect or probable case. When infection control programs use risk factors, or identification factors, to trigger surveillance activities, they are referred to as *trigger factors* [9].

3. Evaluation and Semantic Querying Using SADI Semantic Web services

We evaluate the ontology using OWL reasoners to check for consistency, satisfiability, expected or unexpected inferred relationships, and subsumption. To evaluate case detection methods, we rely on the results from the TOH's ongoing chart review process for HAI. We also assess the ontology based on its ability to meet the initial design requirements by defining different queries (simple or conjunctive) over the defined axioms. We use SADI services [7, 8] that populate the relevant HAI Ontology predicates based on the semantic mapping between the data warehouse schema and the ontology, using relevant data stored in the data warehouse. For querying purpose we use SPARQL⁶ Query Language for RDF⁷. A sample case detection query can be modeled as follows:

Query: Which patients diagnosed with SSI had older age as a risk factor?

```

PREFIX haio: <http://unbsj.biordf.net/haiku/HAI.owl#>
PREFIX haiso: <http://unbsj.biordf.net/haiku/haiku-sadi-service-ontology.owl#>
PREFIX esto: <http://unbsj.biordf.net/ontologies/extra-simple-time-ontology.owl#>
PREFIX sio: <http://semanticscience.org/resource/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT DISTINCT ?Patient
FROM <http://unbsj.biordf.net/haiku/old_age_as_risk_factor.rdf>
  
```

⁵ <http://protegewiki.stanford.edu/wiki/OntoGraf>

⁶ <http://www.w3.org/TR/rdf-sparql-query/>

⁷ Resource Description Framework (RDF): www.w3.org/RDF/

```

WHERE
{
  # Enumerate patients with haio:Old_age_as_risk_factor
  haio:Old_age_as_risk_factor haio:inverse-what_risk_factor
  ?HavingRiskFactorSit .
  ?HavingRiskFactorSit haio:who_has_risk_factor ?Patient .
  ?HavingRiskFactorSit haio:situation_has_time ?HavingRiskFactorSitTime .

  # Enumerate their diseases:
  ?Diagnosis haio:is_performed_for ?Patient .
  ?Incident haio:identified_through ?Diagnosis .
  ?Diagnosis haio:situation_has_time ?DiagnosisTime .

  # Temporal check:
  ?HavingRiskFactorSitTime esto:contains ?DiagnosisTime .

  # Check if the disease is an SSI:
  ?Incident haio:disease_has_type haio:SSI .
}

```

Questions of this type allow surveillance practitioners to estimate the importance of particular risk factors.

4. Conclusions

HAIs place a major burden on patients and the healthcare system. Surveillance is critical for preventing HAI, but current surveillance efforts are limited by the need to manually integrate data, which lack uniform semantics and are scattered across disparate information systems. Knowledge modeling has been used successfully in many domains to add context to data and enable automated data integration and reasoning, especially in the form of query answering. Initial applications of knowledge modeling to surveillance have shown considerable promise, but they have been limited in the scope and the extent to which they have been used to drive knowledge-based surveillance systems within hospitals. In this paper we have described how our effort in knowledge modeling enabled us to address these limitations. Our proposed knowledge-based framework provides evidence on the ability of semantic technologies to support HAI surveillance. As future work we consider enriching the ontological structure to address more complex queries. We will also continue to explore the potential benefit of incorporating the results of the statistical inference into our logical framework. In particular, we intend to apply these enhancements to improving the accuracy of case detection.

Acknowledgements

This project is funded by the Canadian Institutes of Health Research (CIHR), the Natural Sciences & Engineering Council of Canada (NSERC), and Canadian Network for the Advancement of Research, Industry and Education (CANARIE).

References

- [1] Centers for Disease Control and Prevention: <http://www.cdc.gov/hai/>

- [2] Douglas Scott II R. *The Direct Medical Costs Of Healthcare-Associated Infections In U.S. Hospitals And The Benefits Of Prevention*. Division of Healthcare Quality Promotion National Center for Preparedness, Detection, and Control of Infectious Diseases Coordinating Center for Infectious Diseases Centers for Disease Control and Prevention. Mar. 2009.
- [3] Zoutman DE, Ford BD, Bryce E, et al. The state of infection surveillance and control in Canadian acute care hospitals. *Am J Infect Control*. 2003 Aug; **31**(5):266-72; discussion 72–3.
- [4] He Y, Xiang Z, Sarntivijai S, Toldo L, Ceusters W. *AEO: a realism-based biomedical ontology for the representation of adverse events*. In Proc. of the Adverse Event Representation Workshop, International Conference on Biomedical Ontologies (ICBO), University at Buffalo, NY, July 26-30, 2011.
- [5] Cowell L G, Smith B. *Infectious Disease Ontology*. Chapter 19 in Sintchenko V., *Infectious Disease Informatics*, 2010; pp. 373–395.
- [6] Shaban-Nejad A, Rose G, Okhmatovskaia A, Riazanov A, Baker C, Tamblyn R, Forster A, Buckeridge D. *Knowledge-based Surveillance for Preventing Postoperative Surgical Site Infection*. In Proc. of XXIII International Conference of the European Federation for Medical Informatics (MIE 2011), *Studies in Health Technology and Informatics Volume 169*, 2011, pp. 145–149.
- [7] Riazanov A, Klein A, Shaban-Nejad A, Rose G, Forster A, Baker C, Buckeridge D. (2011). *Towards Clinical Intelligence with SADI Semantic Web Services: a Case Study with Hospital Acquired Infections Data*. In Proc. of the 4th international conference in Semantic Web Application and Tools for Life Science (SWAT4LS-2011), ACM Press, Dec 9, 2011, London, UK, pp. 106–113.
- [8] Wilkinson M, McCarthy L, Vandervalk B, Withers D, Kawas E, Samadian S. SADI, SHARE, and the in silico scientific method. *BMC Bioinformatics*, 2010; **11**(Suppl 12):S7.
- [9] Rose GW. Use of an Electronic Data Warehouse to Enhance Cardiac Surgical Site Infection Surveillance at a Large Canadian Centre. MSc Thesis, University of Ottawa, 2009.
- [10] National Nosocomial Infections Surveillance System. National Nosocomial Infections Surveillance (NNIS) System Report, data summary from January 1992 through June 2004, issued October 2004. *Am J Infect Control*. 2004 Dec; **32**(8):470–85.
- [11] CDC/NHSN Surveillance Definition of Healthcare-Associated Infection and Criteria for Specific Types of Infections in the Acute Care Setting: www.cdc.gov/nhsn/PDFs/pscManual/17pscNosInfDef_current.pdf
- [12] Siegel JD, Rhinehart E, Jackson M, Chiarello L. 2007 Guideline for Isolation Precautions: Preventing Transmission of Infectious Agents in Health Care Settings. *Am J Infect Control*. 2007; **35**(10 Suppl 2):S65–164.
- [13] Schraffenberger LA. *Basic ICD-9-CM Coding*. Ahima Press; 1st edition (August 2, 2010).
- [14] Rosse C, Mejino JVL. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform*. 2003; **36**:478–500.
- [15] de Matos P, Alcántara R, Dekker A, Ennis M, Hastings J, Haug K, Spiteri I, Turner S, Steinbeck C. Chemical entities of biological interest: an update. *Nucleic Acids Res*. 2010 Jan; **38**(Database issue):D249–54.
- [16] The SemanticScience Integrated Ontology (SIO): code.google.com/p/semanticscience/wiki/SIO
- [17] Allen JF. Time and Time Again: The Many Ways to Represent Time. *International Journal of Intelligent Systems*. 1991; **6**(4).