A Working Semantic Model for the Integration of Occupation, Function and Health

Anil Adisesh[1], Hongchang Bao[2],
Mohammad Sadnan Al Manir[2,3] and Christopher J.O. Baker[2,3]

[1] Department of Medicine, Division of Occupational Medicine
University of Toronto, Canada
Anil.Adisesh[at]unityhealth.to
[2] Department of Computer Science
University of New Brunswick, Saint John, Canada
[3] IPSNP Computing Inc
{bakerc,hbao,sadnan.almanir}[at]unb.ca

Abstract. Occupation is an explanatory variable in health research that is used to identify the degree to which exposures to environmental hazards and working conditions are correlated with disease. Moreover disease and functional impairment can limit employment options open to patients. Despite the importance of these issues many essential data sets have yet to be integrated. In the current study we defined an integrated semantic model and populated coded patient data representing disease (ICD), functional impairment (ICF), occupation (NOC), and job attributes (NOC Career Handbook). Automated NOC coding of patient responses to "What is your job" were coded by a custom algorithm developed in previous work. To validate the utility of the model, SPARQL queries and outputs were prepared and discussed in the context of authentic physician and case worker activities.

1 Introduction

Occupation is a widely used determinant in health research representing socioeconomic status and class, as well as environmental exposures [1]. Despite this many data sets collected at point of care fail to record patients' occupations, limiting the reuse of pertinent data for applications relevant to occupational associations of disease and patient outcomes.

Research studies in the area of Occupational Health (OH) are typically targeted to specific lines of inquiry such as examining the burden of cancer attributable to occupation [2,3]. Given the challenges in accurate and timely recording of occupations in a standardized way, several studies have sought to facilitate automated coding of occupations using standardised classifications of occupations [4,5]. While successful in accelerating the recording of jobs, with up to 70% accuracy, subsequent analyses using such coded data sets remain limited. This is in part due to the need to define core objectives and integrate complex data sets.

Applications of occupation coding in OH practice include correlating: i) disease and occupation, and ii) functional impairment and fitness for work. For instance, the occupations associated with *silicosis* are a major concern. Occupational groups exposed to silica[4] include construction labourers, heavy equipment operators, plasterers and drywallers carrying out grinding, sandblasting, crushing, chipping, mixing, and plowing which are common in many industries such as mining, agriculture, and manufacturing industries [1,6]. Researchers are interested in identifying a specific disease or chronic condition associated with a type of employment. This type of investigation is an essential step towards the evolution of health policy and decisions about workplace conditions, albeit challenging because of the distributed data sources and a lack of standardization of the source data.

In this work, we have sought to design an integrated model to support investigations of these both themes leveraging standardised classifications of disease, functional impairment, patient data linked to a National Occupation Classification (NOC), using an NLP-based grounding algorithm, and job attributes. With a populated model we assess the suitability of the model for the target studies.

2 Model

In medicine, patients are primarily assessed with the goal of disease diagnosis and treatment, whereas functional impairment assessments for chronic diseases or injury occurring at home/workplace/sports are a secondary consideration for practitioners. However assessments of function are essential when determining cases of workplace compensation or health insurance claims as well as for rehabilitation guidance.

Here we describe the kinds of entities relevant to OH that exist in the application environment, along with classifications and groupings of entities, and core relations between them. Figure 1 shows an integrated model consisting of two graph models.

In consultation with domain experts we identified the core concepts and data used in OH, specific to the occupational dimensions of disease onset and assessment of functional impairment. The functional descriptors capture the essential physical abilities and aptitudes required to perform effectively in a given occupation. The following concepts were modelled: Patient, Disease, FunctionalImpairment, NOCCode, and NOCTitle in the smaller model, while JobAttribute, PhysicalActivity, Aptitude, CHNOCCode, and CHNOCTitle in the larger model. Six more concepts Vision, Hearing, LimbCoordination, ColorDiscrimination, BodyPosition, and Strength are subclassed to PhysicalActivity while nine other concepts are subclassed to Aptitude. Instances of each model can be integrated via the alignment of the concepts NOCTitle and CHNOCTitle.

In the model, each instance of a Patient represented by an identifier is *diagnosedWith* an instance of a Disease, which is represented by an ICD-10 code.

---

[4] `https://www.carexcanada.ca/profile/silica_crystalline-occupational-exposures/`

Instances of FunctionalImpairments are represented by ICF codes and they are caused by (*causes* relation) one or more instances of Disease. The job a patient *qualifiesFor* is represented by the instance of NOCCode and its corresponding title as an instance of NOCTitle, expressed by the *hasTitle* property. Instances of physical activities (PhysicalActivity) and aptitudes (Aptitude) are *partOf* of a job attribute (JobAttribute) which is *requiredBy* each job instance of CHNOC-Code from the career handbook. The title of a job is an instance of CHNOCTitle and expressed by the property *hasTitle*.

Subsequent sections describe the data represented, and how they were derived. The model is designed to support multiple queries competency detailed in Section 5.
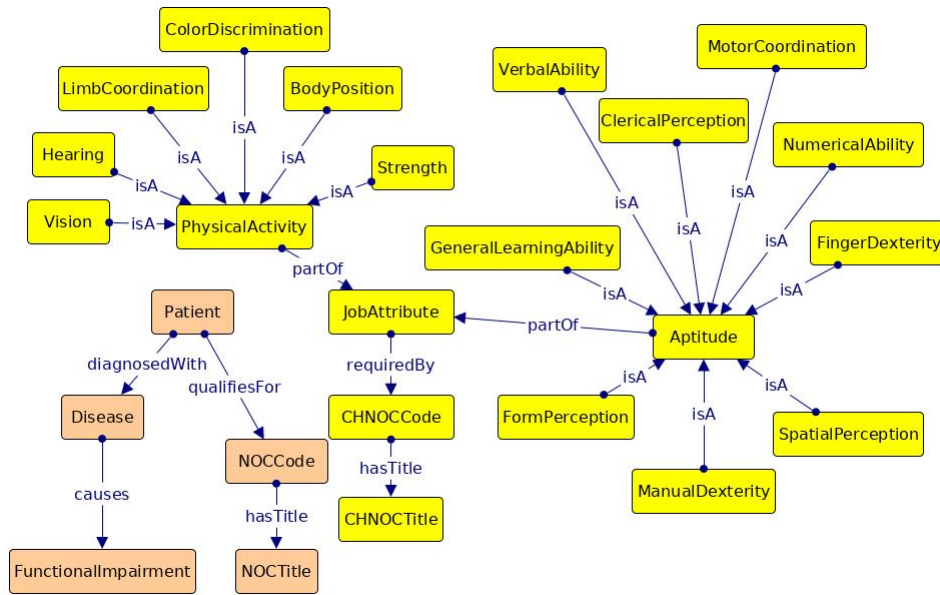

Fig. 1. Integrated data model for occupation, function and health

3 Description of Data

Patient Data Patient data was gathered as part of the Canadian Immunisation Research Network Community Acquired Pneumonia study [7] to investigate occupational associations. Pre-existing example data for patients with Diabetes mellitus was also used in this study.

Occupation Data For each patient the data set contained the fields "Current Job Title", and "Current Industry". This data originates from free text entered in response to the questions "What is your job title?" and "In which type of

industry do you work?". The dataset of 566 patients included coding to NOC 2016 and NAICS (North American Industrial Classification), added manually.

Canadian National Occupational Classification The Canadian National Occupational Classification (NOC) is the national reference on occupations in Canada providing a standard taxonomy for labour market information and employment-related program administration. NOC-2016 [6] is organized in a four level hierarchy, there are 10 broad occupational categories (first level), 46 major groups (second level), 140 minor groups (third level), and 500 unit groups (fourth level) encoding more than 30,000 occupational titles. For example, sample data for the occupation of cook is as follows: First Level: 6 Sales and service occupations, Second Level: Major Group 63 - Service supervisors and specialized service occupations, Third Level: 632 Chefs and cooks, Fourth Level: 6322 Cooks - Cooks are employed in restaurants, hotels, hospitals and other health care institutions, central food commissaries, and educational institutions.

Career Handbook Data The *Career Handbook* is the counselling component of the National Occupational Classification (NOC) [6] system. The handbook details worker characteristics and other occupation indicators and is used to help people make informed career decisions. It includes information for each occupation on the required; aptitudes, physical activities, environmental conditions, education/training, career progression and work settings.

Aptitudes[5] required for a person to learn the skills needed to perform job duties are defined numerically on a scale from 0 to 5. These include general learning ability, clerical perception, verbal ability, motor co-ordination, numerical ability, finger dexterity, spatial perception, manual dexterity and form perception.

Physical abilities[6] include vision, colour discrimination, hearing, body position, limb coordination and strength. In the case of visual performance for work there are four categories with examples; V1 - Close visual acuity (assembling micro-circuit boards), V2 - Near vision (reading and interpreting drawings and specifications), V3 - Near and far vision, (installing shingles/tiles on roofs), V4 - Total visual field (driving vehicles).

Disease The International Classification of Diseases (ICD) is maintained by the World Health Organisation. It supports the identification of health trends and statistics globally, and is the international standard for reporting diseases and health conditions. The currently used version in many jurisdictions is ICD-10 although some continue to use ICD-9.

A version of ICD-11 was released[7] on 18 June 2018 to allow Member States to prepare for implementation, including translating ICD into their national languages. The ICD-11 for *Mortality and Morbidity Statistics*[8] (Version : 04/2019)

---

[5] `http://noc.esdc.gc.ca/English/CH/AptitudesEnglish.aspx?ver=16&ch=03`

[6] `http://noc.esdc.gc.ca/English/CH/PhysicalActivities.aspx?ver=16&ch=03`

[7] `https://www.who.int/classifications/icd/en/`

[8] `https://icd.who.int/browse11/l-m/en`

allows visualisation of the coding hierarchy e.g. for code 5A10 Type 1 diabetes mellitus, the ancestors to the top are *05 Endocrine, nutritional or metabolic diseases*, *Endocrine diseases*, and *Diabetes mellitus*, in order.

Functional Impairment For functional impairment we used the International Classification of Functioning, Disability and Health (ICF) which provides a comprehensive and universally-accepted framework to describe functioning, disability and health. Specialised clinical use requires both Comprehensive and Brief ICF Core Sets e.g. there are 99 ICF categories in the Comprehensive and 33 second-level ICF categories in the Brief ICF Core set for diabetes mellitus. A sample Comprehensive ICF Core Set for Diabetes Mellitus for the component 'body functions is shown in Table 1:

ICF Code 2nd Level	ICF Code 3rd Level	ICF Category Title
b455		Exercise tolerance functions
	b4550	General physical endurance
	b4551	Aerobic capacity
	b4552	Fatiguability

Table 1. Sample of the comprehensive ICF core set for Diabetes Mellitus Categories of the component 'body functions'

4 NOC Data, Coding and Population of Semantic Model

Data imported to the model was acquired from original sources as described in Section 3. Essential to the model is the grounding of free text patient data to the NOC Classification. This is achieved using a coding algorithm [8] that iteratively performs look ups in the NOC database until all the given job titles are matched with one or more NOC codes. At each iteration free text inputs are pre-processed by splitting job titles, removing stop words, stemming, followed by spelling corrections or grammar checks. The accuracy of the algorithm for grounding to 4 digit NOC codes is 58.66 percent, based on benchmarking on manually coded data from previous studies. Instances of NOCCode and NOCTitle are populated to the model based on this algorithm.

Data from 500 Pneumonia cases was used to populate instances of Patient and Disease. Authentic ICF codes were populated as instances of functional impairments caused by the corresponding diseases. The Career Handbook data contributes instances of PhysicalActivity and Aptitude which are *partOf* one or more instances of JobAttribute and are among the sets of requirements to qualify for a job.

5 Competency Queries

In this section we focus on illustrating the suitability of the model for occupation, function and health using competency queries and offer further insights into the needs of the target community. The following questions are illustrative of the types of queries that may be of interest to a target user.

Q1. What is the job classification for patients with disease X?

A 2019 report demonstrates this type of question, a Colorado physician specializing in occupational lung disease observed an increasing number of silicosis cases in her practice. She undertook an review of electronic medical records for a one year period of patients with a silicosis diagnosis (ICD-10 code J62.8). Normal rates silicosis were two per year; however, during June 2017-December 2018, seven cases[9] of silicosis were identified, all among employees of stone fabrication companies.

The following SPARQL query represents such an investigation. It uses a ICD-10 code J62.8 for a disease X as input and produces NOC codes for the job (`?noc_code`) and the NOC title (`?noc_title`) as output. In this query, lines 1-4 show the prefixes, line 7 asserts J62.8 as the disease, *Pneumoconiosis due to other dust containing silica*, lines 8-10 assert a patient identified by `p1001` and its relation with the disease and the job s/he qualifies for, and lines 11-13 are used to find the values of the job code and the corresponding job title. The results of the query are shown in Table 2.

```
1.  PREFIX :     <http://www.example.org/ontology/patoccont#>
2.  PREFIX icd10: <https://icd.who.int/browse10/2016/en#/>
3.  PREFIX patid: <http://identifiers.org/patid/>
4.  PREFIX rdfs:  <http://www.w3.org/2000/01/rdf-schema#>
5.  SELECT ?noc_code ?noc_title
6.  WHERE {
7.    icd10ID:J62_8 a :Disease .
8.    patid:p1001 a :Patient ;
9.              :diagnosedWith icd10:J62_8 ;
10.             :qualifiesFor ?nc .
11.   ?nc :hasTitle ?nt ;
12.       rdfs:label ?noc_code .
13.   ?nt rdfs:label ?noc_title . }
```

?noc_code	?noc_title
8231	Construction trades helpers and labourers
8231	Underground production and development miners
6344	Jewellers, jewellery and watch repairers and related occupations

Table 2. Job classification (NOC code and NOC title) of patients with Pneumoconiosis due to other dust containing silica (ICD-10 Code J62.8)

---

[9] https://www.cdc.gov/mmwr/volumes/68/wr/mm6838a1.htm?s_cid=mm6838a1_x

Q2. What is the disease and the job classification for patients with acquired disability (resulting from a functional impairment)?

A functional impairment for a patient may be caused by one of more diseases. In this scenario, a case worker is interested in reviewing the job categories for a given acquired disability resulting from a functional impairment. An example of such a scenario may involve assessing *visual impairment*, the associated medical conditions and the corresponding job titles. Knowing the category of job (NOCCode) allows certain assumptions to be made about the required capabilities for the patient's job. This permits a review of the corresponding skills and attributes typically required of the job and an early determination as to whether the patient is likely to be able to return to the current job or a similar one.

The SPARQL query below is capable of answering questions in such a scenario. It uses the ICF code b2101 as an instance of :FunctionalImpairment in line 5 to represent Visual field functions (i.e. seeing functions related to the entire area that can be seen with fixation of gaze) as input. For all patients, the query returns patient identifiers (lines 6-7), job codes and job titles (lines 9, 13-15), the corresponding ICD-10 code of the disease which causes the impairment (lines 8, 10-11), and the corresponding name of the disease (line 12). The output in Table 3 lists 3 patients suffering from visual impairments.

```
1.   PREFIX icf:  <http://who.int/icf#>
2.   PREFIX sc:   <http://schema.org/>
3.   SELECT ?patient_id ?noc_code ?noc_title ?icd_code ?disease_name
4.   WHERE {
5.     icf:b2101 a :FunctionalImpairment .
6.     ?patient a :Patient ;
7.             rdfs:label ?patient_id ;
8.             :diagnosedWith ?icd ;
9.             :qualifiesFor ?nc .
10.    ?icd :causes icf:b2101 ;
11.          rdfs:label ?icd_code ;
12.          sc:name ?disease_name .
13.    ?nc :hasTitle ?nt ;
14.         rdfs:label ?noc_code .
15.    ?nc rdfs:label ?noc_title . }
```

?patient_id	?noc_code	?noc_title	?icd_code	?disease_name
1001	7511	Truck driver	E10.3	Type 1 diabetes mellitus with ophthalmic complications
1011	7442	Utility worker	H33.0	Retinal detachment with retinal break
1083	7241	Electrician	H40.2	Primary angle-closure glaucoma

Table 3. List of patients, their job classification, ICD-10 codes and names of the diseases causing functional impairment of the Visual field function (ICF code b2101)

Q3. What jobs can a patient with vision impairment likely return to?

In the same scenario as Q2, a case worker is again tasked with reviewing the 'Return to Work' options for a patient with a recently acquired disability. This time the researcher is interested to identify, in the case of the Truck Driver in Table 3, not just the category of job that the patient worked in, but to determine whether the acquired disability from the specific functional impairment (Visual field functions (ICF code b2101) caused by Type 1 diabetes mellitus with ophthalmic complications (ICD Code E10.3)), may prevent the patient from continuing in his/her job.

An explicit mapping can be established between the various job attributes such as Physical abilities (`PhysicalActivity`) and the functional impairments (`FunctionalImpairment`) described in Section 3. Based on the assessments concerning the level of vision impairment and the visual requirements for the jobs currently shown in Table 3, the mappings that could be established were: Close visual acuity-V1 maps to to Visual acuity functions, other specified (ICF code b21008), Near vision-V2 maps to Binocular acuity of near vision (ICF code b21002), Near and far vision-V3 maps to Visual acuity functions (ICF code b2100), and Total visual field-V4 maps to Visual field functions (ICF code b2101).

For 'Truck Drivers', the mapping confirms a the visual requirement of V4 is required for the job. We then explore the alternative job options for the patients where they can likely successfully transition to. The target query must return job titles with alternate vision requirements to that of the original job, but have similar `JobAttributes` of a Truck Driver, and the jobs listed must not require "Total visual field" level of vision.

The SPARQL query below filters out results from the integrated data by removing all cases of "Total visual field" which has a V score of 4, but matches ICF code b2101, and the instances of NOCCode value 7511. Lines 5-12 matches triples from the data model populated from the Career Handbook while lines 13-20 matches triples from the patient-centric populated data model.

```
1. PREFIX vision: <http://identifier.com/vision/>
2. SELECT distinct ?noc ?noc_title ?icf_code ?ch_noc
3.                 ?ch_noctitle ?v_score  ?att_name
4. WHERE {
5.    vision:v1001 a :Vision ;
6.                 sc:ratingValue ?v_score ;
7.                 rdfs:label ?att_name ;
8.                 :partOf ?ja .
9.    ?ja :requiredBy ?chnoc .
10.   ?chnoc rdfs:label ?ch_noc ;
11.          :hasTitle ?chnt .
12.   ?chnt rdfs:label ?ch_noctitle .
13.   ?patient :diagnosedWith ?icd ;
14.           :qualifiesFor ?nc .
15.   ?nc rdfs:label ?noc ;
16.       :hasTitle ?nt .
```

```
17.    ?nt rdfs:label ?noc_title .
18.    ?icd :causes ?icf .
19.    ?icf a :FunctionalImpairment ;
20.        rdfs:label ?icf_code .
21.    FILTER(?v_score != 4 && ?icf_code="b2101" && ?noc="7511") }
```

In Q3 we seek to accommodate the need to cross reference impairment with given job attributes in an attempt to list jobs where a patient might return-to-work, successfully whether it be the same job or another. The results in Table 4 show the NOC code and impairment code for a patient (who is a Truck Driver), as well as the career handbook code and career handbook label of the jobs that the patient might return to with the given impairment together with the corresponding vision score in the job attributes, and the vision attribute label. Specifically, physicians and case workers like to query directly for impairments that prevent a patient from returning to their the most recent job, and compare these side by side with the physical attributes of a given job.

?noc	?noc_title	?icf_code	?ch_noc	?ch_noctitle	?v_score	?att_name
7511	Truck driver	b2101	5212.1	Conservation and restoration technicians	1	Close visual acuity
7511	Truck driver	b2101	7384.1	Gunsmiths	1	Close visual acuity
7511	Truck driver	b2101	7281.0	Bricklayers	3	Near and far vision
7511	Truck driver	b2101	2225.7	Lawn care specialists	3	Near and far vision
7511	Truck driver	b2101	9414.1	Concrete products forming and finishing workers	2	Near vision
7511	Truck driver	b2101	5212.7	Picture framers	2	Near vision

Table 4. List of jobs (column ?ch_noctitle) a Truck Driver can transition to with different visual requirements other than Total visual field (V-4)

6 Discussion

In the current study we have assessed the core objectives in occupational medicine and reviewed the target data that needs to be integrated to support these goals, namely to explore occupation and disease associations, and to cross-reference the NOC Career Handbook. With knowledge of a patient's occupation, in the form of a job title, integrated with information from disease diagnosis, it is possible to indicate likely functional impairments and consequent difficulties in work performance.

The preliminary model we developed appears to be fit for the initial purposes and can support our sample queries. Instantiation of the model depended on additional algorithmic computations for the coding of job coding (NOC Code) and our most complex query had to rely on expert curated mappings between ICF and the Career Handbook for details of attributes required in the given job.

To our knowledge this is the first study integrating data specific to occupational medicine using semantic technologies. The utility of the model is not limited to OH, and it may find application in other areas such as human resources or for government agencies to assist with accommodation of disability or ensuring appropriate allocation of social benefits.

In addition to the Canadian NOC there are similar coding schemes in different countries and jurisdictions and established crosswalks between classifications. The descriptors for job attributes will also be similar across countries, moreover the disease and function classifications are international being maintained by the World Health Organisation. Therefore the model can be applied in any location context using the same or different occupation descriptors.

References

1. Leslie A. MacDonald, Alex Cohen, Sherry Baron, and Cecil M. Burchfiel. Occupation as Socioeconomic Status or Environmental Exposure? A Survey of Practice Among Population-based Cardiovascular Studies in the United States. *American Journal of Epidemiology*, 169(12):1411–1421, 05 2009.
2. France Labrèche, Joanne Kim, Chaojie Song, Manisha Pahwa, Calvin B. Ge, Victoria H. Arrandale, Christopher B. McLeod, Cheryl E. Peters, Jérôme Lavoué, Hugh W. Davies, Anne-Marie Nicol, and Paul A. Demers. The current burden of cancer attributable to occupational exposures in canada. *Preventive Medicine*, 122:128 – 139, 2019. Burden of Cancer in Canada.
3. Mark P. Purdue, Sally J. Hutchings, Lesley Rushton, and Debra T. Silverman. The proportion of cancer attributable to occupational exposures. *Annals of Epidemiology*, 25(3):188 – 192, 2015. Causes of Cancer.
4. Daniel E Russ, Kwan-Yuet Ho, Joanne S Colt, Karla R Armenti, Dalsu Baris, Wong-Ho Chow, Faith Davis, Alison Johnson, Mark P Purdue, Margaret R Karagas, Kendra Schwartz, Molly Schwenn, Debra T Silverman, Calvin A Johnson, and Melissa C Friesen. Computer-based coding of free-text job descriptions to efficiently identify occupations in epidemiological studies. *Occupational and Environmental Medicine*, 73(6):417–424, 2016.
5. Igor Burstyn, Anton Slutsky, Derrick G. Lee, Alison B. Singer, Yuan An, and Yvonne L. Michael. Beyond Crosswalks: Reliability of Exposure Assessment Following Automated Coding of Free-Text Job Descriptions for Occupational Epidemiology. *Annals of Work Exposures and Health*, 58(4):482–492, 02 2014.
6. Employment and Social Development Canada and Statistics Canada. National occupational classification 2016. `http://noc.esdc.gc.ca/English/noc/welcome.aspx?ver=16`, 2016. [Online; accessed September-2019].
7. Canadian Immunisation Research Network. Serious outcomes surveillance (sos) network. `http://cirnetwork.ca/network/serious-outcomes/`, 2019. [Online; accessed September 09, 2019].
8. Bao, Gary and Baker, Christopher J.O. and Adisesh, Anil. Development of the ASOC (Automated Semantic Occupation Coding) Algorithm. *JMIR Preprints 27/09/2019:16422*, 2019. DOI: 10.2196/preprints.16422.