Foundational Development of an Occupation Ontology

Sam Smith [1], Damion Dooley [2] and Yongqun He[1]

[1] *University of Michigan Medical School, Ann Arbor, MI, USA.*
[2] *Simon Fraser University, Vancouver, BC, Canada.*

Abstract
The classification of occupations, key to policy development regarding the economy, education, research, organizational infrastructure, as well as health and safety analysis, has received considerable attention and development over the last half century. However, international and national standardized taxonomies, while very well executed, are dissimilar, dated and could benefit from a new organization employing ontological methods. This paper builds the case for the development of an Occupation Ontology, showing an approach that employs successful methods of the Open Biological and Biomedical (OBO) Foundry. The importance of occupations and their characterization is described, and four popular taxonomies available in English are presented and compared. We selected and transformed the US Bureau of Labor Statistics Standard Occupational Classification into a prototype OBO compliant ontology, designated OccO, to illustrate the presentation of occupational information as an ontology. OccO is then compared to the representation of occupations in the fastest growing semantic web knowledge base, Wikidata, where occupations are organized neither in compliance with any standard taxonomy nor with a consistent upper level ontological framework. We then present the benefits of developing an ontology-driven occupational framework. This development is at the alpha stage, with the hope that the concept of an occupation ontology will gain the support of research organizations who could cooperate to formulate and demonstrate the utility of a well-designed and widely used ontology for occupations.

1. Introduction

Occupations occupy a great deal of each person's life. The central role of occupation, how one earns his or her livelihood, was recognized at the end of World War I with the formation of the League of Nations International Labor Organization (ILO, now a UN agency) [1]. The ILO was motivated by "the belief that universal and lasting peace can be accomplished only if it is based on social justice." To facilitate the analysis of work practices around the world, standardized occupation taxonomies were developed, with the ILO International Standard Classification of Occupations (ISCO), adopted in 1957 as ISCO-58, becoming the official international reference. ISCO was updated as ISCO-68, ISCO-88, with the current version ISCO-08.

However, individual nations needed references that were tailored to local needs. Today four English language taxonomies are in use that constitute the input data for this project. In addition to ISCO, these are: (1) US Bureau of Labor Statistics Standard Occupational Classification (US SOC), first released in 1958 [2]; (2) The UK National Statistics Standard Occupational Classification (UK SOC) 2020, first issued in 1990 [3]; (3) The European Skills, Competences, Qualifications and Occupations (ESCO) of

the European Union, first released in 2010 [4]. These taxonomies reflect decades of development and adoption, but they are dissimilar, and all are based on a digital coding system embedded in the occupation codes that is restrictive. To address these deficiencies, we propose a novel organization of occupational information as an applied ontology suitable for computer reasoning.

2. Methods
2.1. Data collection

Our data sources are the four English language occupation sources, including ILO ISCO, US SOC, UK SOC, and EU ESCO. These taxonomies provide the range of occupations encompassed in this alpha development of OccO: their occupaton labels and categories constitute the information around which OccO is built. Additionally, there is a wealth of information on skills, abilities, and educational requirements within these taxonomies, which would be difficult to obtain from other sources. Also, we have reviewed the classifications of "field of work" (or industry) with reference to the US NAICS system and the international ISIC system, choosing to focus on ISIC for "field of work" descriptors. The Environment Ontology (ENVO) [5], an existing OBO Foundry compliant reference, is also employed. English language taxonomies for Canada and Australia also exist, but they are not included in this alpha development.

2.2. OccO ontology development

The Protege OWL-editor [6] was used for editing. Ontobee [7] was used as our primary ontology search program. We aligned our work with OBO Foundry principles such as openness and collaboration [8]. OccO is aligned with the Basic Formal Ontology (BFO) [9] upper layer ontology. BFO was chosen for this development, since it is used quite successfully in over 150 active OBO Foundry ontologies [10], facilitating integration of these ontologies into OccO. The OccO GitHub website is: https://github.com/Occupation-Ontology/OccO. The OccO source code, which uses the open license CC BY 4.0, is also available on the website.

2.3. Use case testing

While thousands will reference the standardized occupational taxonomies, millions will obtain occupational information from the Internet and Semantic Web resources such as Wikipedia and its related database Wikidata. Our use case will be to compare and contrast the treatment of occupations in Wikidata, a widely used instance of the Semantic Web, and demonstrate that an OBO aligned ontological organization of OccO is superior. The case is made that, if a widely supported occupational ontology could be developed, it would make sense to ensure that Semantic Web applications should comply with this ontology.

3. Results

Initially we formulated various occupations under the BFO:role, which was first presented in our poster presentation in the BioCuration 2018 conference [11]. Later we realized that the existing occupation systems, including ILO ISCO, US SOC, UK SOC, and EU ESCO, inherently represent occupations as occupation holders that have specific abilities and skills. These occupation systems define terms such as "dentist" and "professor" instead of "dentist role" or "professor role". After thorough consideration, the OccO occupation ontology focuses on classifying various types of occupation holders and their associated characteristics including skills and abilities as specified in the existing occupation systems. This does not diminish the use of BFO:role, it simply guarantees that we have an easily referenceable identifier for each "prototype" human holding an occupational role. The

resulting ontology builds from an existing standardized taxonomy that links to BFO entities and relationships. Its application will be illustrated with a specific example, contrasted with US SOC and Wikidata.

3.1. Basic occupation definition in OccO

There are different definitions of "occupation." Each of the taxonomies defines occupation within their framework, but to build an ontology from scratch, would it not be better to begin with the common understanding of the term? The Merriam-Webster dictionary defines it as "an activity in which one engages". Wikipedia, reflecting a consensus among contributors (not necessarily specialists) defines it as a synonym of job (or employment), as "one's role in society, often a regular activity performed for payment". The meanings of "occupation" have also been discussed in many journal articles [12-15]. For example, Royeen [15] has surveyed approximately 20 articles on the various types of 'occupation' definition, such as occupation as a process or means, as the outcome (i.e., state or condition) of the process of engaging in occupation, or as people going back to work.

In the initial scope of OccO, our reference sources are from the field of economics, so we propose that occupation within the scope of OccO be considered as a means of livelihood, excluding non-economic activities such as hobbies, non-professional sports and volunteer activities, but including military as well as non-military occupations. Extending the scope of OccO is an objective once the initial implementation is completed. This would allow inclusion of non-Western occupations, multi-lingual occupations, and non-livelihood avocations such as hobbies and volunteer activities.

Regardless of definition, there are two uses of occupation: Ms. Jones 'is a' pharmacist, versus "the occupation of pharmacist requires the skill of reading comprehension." In OccO, we classify Ms. Jones as a pharmacist (which is an 'occupation-holder'), who has taken the "pharmacist role" (a term internally there but not represented in OccO), which is a specific 'occupation role'. In order to be a pharmacist, Ms. Jones requires the skill of reading comprehension, for example. OccO is grounded by focusing on the human being who 'is' the occupation holder, rather than on occupation roles. (OccO does not provide relations for describing current accreditation of professionals).

Table 1:
Classes by Level in each taxonomy.

	ILO ISCO	US SOC	UK SOC	EU ESCO
Major	10	23	9	10
2nd	43	98	26	43
3rd	130	459	104	130
Detail	436	1016	412	3008

Each of the four reference taxonomies is organized into a hierarchy of four or more levels, the top levels in each case called Major Group. The breakdown of these groups is as follows (Table 1, ESCO has the same organization as ISCO but at the detailed level a dotted suffix notation allows a given ISCO code to be broken into several more detailed occupations).

Except for the UK SOC, military occupations are defined as their own major group. The detail level occupational terms still represent classes of occupations, with tens of thousands of individual occupation labels associated with these detail level terms. In addition to these layers of occupations, the occupation sources include two attributes "skills" and "abilities". Within OccO, these attributes are treated as classes and can be linked to specific occupation holders by object properties. The occupation holder is linked to the requisite skills and abilities by the object properties 'has skill' and 'has ability.' The source taxonomies also include attributes, such as "field of work", "knowledge," "context," and

"tasks.", but these are not covered in OccO. The set of occupation terms and categories described by these taxonomies constitute the initial nucleus of scope for OccO. Further development can analyze the tens of thousands of occupation terms that appear as alternate terms in the taxonomies, and the 17,000 terms for occupations associated with people contained in Wikidata.

3.2.   OccO is developed using the US SOC as the initial model

Several factors affected the selection of a taxonomy for the initial formulation of OccO, with integration of other taxonomic systems possible in the future. The ISCO faces a major challenge of getting agreement by virtually all nations, including developing nations where non-Western occupations must be incorporated. The three country-affiliated sources have greater flexibility and are well-developed, impressive government systems with substantially improved online access and download capabilities. Of these, the US SOC was selected for several reasons:

- Both the UK SOC and ESCO comply with the limited top level grouping of occupations that is artificially restricted to ten groups, simply because major groups are identified by a single digit. The US SOC has greater granularity with 23 major groups. These are not often ontologically pure categories but are reused as a starting point in development.
- The O*Net system has recently been developed as an enhancement to the US SOC, with ease of access to the data and extensive online search capabilities, enabling easier conversion and curation into OccO. (O*Net provides online querying for individuals interested in career choices as well as for general information about occupations.)
- Additionally, the O*Net system has extensive definition and organization of occupational attributes, for skills, abilities, knowledge, tasks and other characteristics of workers and roles [16]. The treatment of skills is also part of UK SOC and ESCO, and is especially prominent with the newest release of ESCO in March 2022. Harmonizing the way these occupational attributes are organized ontologically would be a worthy endeavor for later development.

3.3.   US SOC treatment of pharmacist

The treatment of occupations in the various sources is illustrated by the occupation of pharmacist (which appears in each taxonomy). The US SOC places it within the following 4-level hierarchy:

Major Group 29, Healthcare Practitioners and Technical Occupations.
    Minor Group 29-1 Healthcare Diagnosing or Treating Practitioners
        Broad Occupation 29-1050 Pharmacists
            Detailed Occupation 29-1051 Pharmacists.

In US SOC, occupations are plural, representing the individuals within that occupation, and capitalized. In OccO each occupation holder is a class representing one individual holding that occupation. It is labeled in the singular form and, in compliance with OBO, lower case. In many cases a subclass in US SOC will have the same term as its parent, but as this is not allowed in an OBO formulation, the phrase ("broad" or "minor", depending on occupation level) is added to the parent term in OccO to make it unique. In this regard, we consider the parent term (e.g., "veterinarian (broad)" is defined as a broad level that includes a narrow level term with the same name). This is not a recommended pattern in OBO Foundry, where the label of a class should hint at least at the difference between itself and its parent or siblings, and where other siblings are encouraged – in other words an ontology with a parent and child class, named the same, and with no siblings and no definition or axiom

differentia (which could guide improvement in labels) creates semantic ambiguity; generally ontologies forsake parent classes just to preserve some arbitrary source taxonomy notion of depth. This will have to be solved in a production ready version of OccO.

OccO incorporates the treatment of skills and abilities based on O*Net's model. Although the O*Net system has a metric of importance and level of each skill and ability for each occupation, these metrics are not achieved by global consensus, and are awkward to represent in an OWL ontology, and so are not a target of ontology representation, but are more fitting as nation specific database content.

3.4.   Semantic Web (Wikidata) treatment of pharmacist

Semantic Web resources such as Wikipedia are de facto reference for millions of people worldwide, but less well-known are the semantic graph databases, Wikidata.org [17] and DBPedia.org [18], which surface content into Wikipedia. Our use case focuses on Wikidata, which, as a Semantic Web resource, supports powerful query capabilities, allowing such searches as "List all of the people who have died of COVID-19 and their respective occupations." There are over 6,000. In Wikidata all items are assigned a numeric identifier with a Q prefix, for example George Washington has an identifier Q23. The entity "occupation" appears as Q12737077, and "pharmacist" is Q105186. Wikidata has over 9000 relations between items or between an item and a datum. Relation identifiers have a P prefix, for example the "occupation" property is identified as P106. Wikidata organizes an occupation hierarchically by two properties, P279 "subclass of," and P31 "instance of" [19], but this distinction is neither enforced nor used consistently, leading to a confusing graph in many cases where for example, a pharmacist has multiple inverse subclass-of and instance-of parents.

3.5.   OccO high level hierarchy

OccO uses BFO as its upper level ontology. BFO is an upper level framework with a small number of entities, it contains a few entities under which OccO will associate occupational entities. Select terms from existing BFO aligned "reference" ontologies are also reused. For instance, "human", defined by the NCBITaxon ontology as a synonym of 'Homo sapiens,' positioned under the BFO class hierarchy "material entity," "independent continuant," "continuant," and finally "entity." In OccO we assert that **occupation holder** 'is a' human, which bestows upon occupation holder any of the necessary and sufficient relationships that more abstract classes such as "human" hold.

Likewise the other terms of OccO are "plugged into" existing BFO entities. "Skill" and "ability," as well as "occupation role" fall under the "realizable entity" of BFO, which is an important property.

Realizable entities are described in [20] as:
Functions, roles, dispositions and capabilities are realizable entities in BFO. A realizable entity is defined as a specifically dependent continuant that has an independent continuant entity as its bearer, and whose instances can be realized (manifested, actualized, executed) in associated processes in which the bearer participates.

Typically an instance of a realizable entity is realized throughout the course of its existence. However it may exhibit periods of dormancy, when it exists by inhering in its bearer but is not manifested — as, for example, in the case of diseases which are marked by periods of dormancy, or by many occupational roles, which are not realized when the bearer is asleep.

In OccO, skill and ability are BFO dispositions, and occupation role is a BFO role, as shown in Figure 1. Merrell et al. have proposed that within BFO capabilities are a special type of disposition that can be evaluated on the basis of how well they are realized [21]. Therefore, we consider both skill and ability are subclasses of capability (Figure 1).
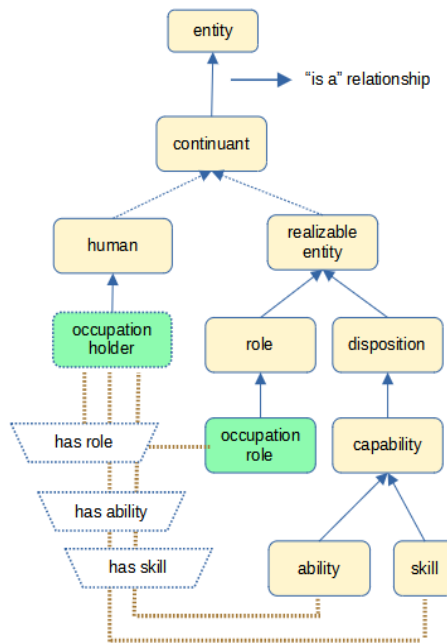
Figure 1 OccO entities within BFO framework

3.6.   OccO treatment of pharmacist

In the initial OccO up for discussion, the major, minor, broad occupation groupings of US SOC are preserved as parent classes, with pharmacist appearing four levels below the root term occupation holder.

To avoid duplication, we can purposely ignore the first 'has role' part in OccO. As a result, OccO defines occupation as:

'occupation holder': =def. A human who has a role in society that is realized in an occupation process or an activity as a livelihood (i.e., "means of support or subsistence"). To fulfill such an occupation, the occupation holder is required to have necessary capabilities including skill(s) and abilities.
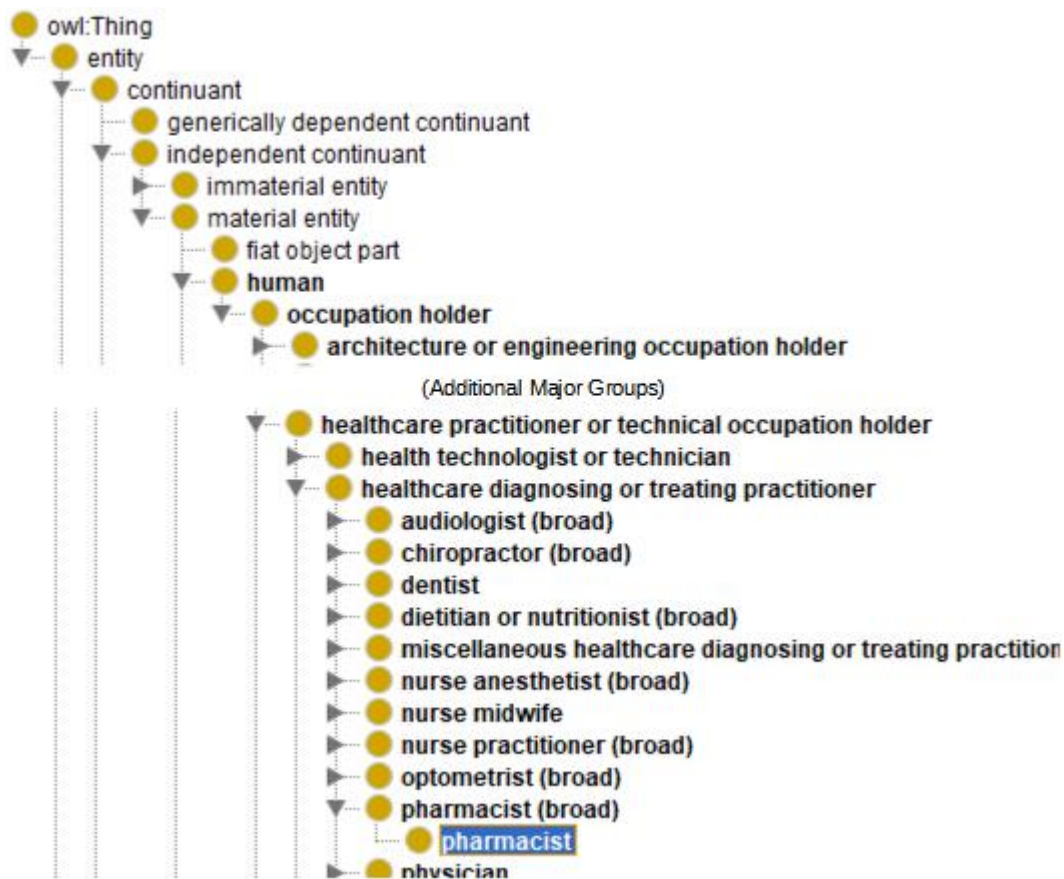
Figure 2 OccO representation of the occupation holder pharmacist

Incorporation of Skills: A defining characteristic of an occupation is the set of capabilities one must have to hold that occupation. Each of the source taxonomy systems includes the treatment of skills, but there are significant differences. The ESCO system, encompassing over 13,000 skills (as of March 2022), accentuates the skills dimension. O*Net defines skills as "cross-functional skills are developed capabilities that facilitate performance of activities that occur across jobs." There are 35 skills in four categories: Complex problem-solving, resource management, social and technical skills. Abilities are "enduring attributes of the individual that influence performance," of which there are 53 in four categories: cognitive, physical, psychomotor and sensory. Skills and abilities within OccO need to be defined ontologically, but prior to undertaking this, there needs to be agreement among the occupation taxonomy organizations. One problem, for instance, in O*Net is that a given occupation does not have a list of skills needed, but all skills are shown with a metric for level of skill and another metric for importance of skill. ESCO is actively enhancing their extensive skills characterizations, but a different metric is used as well as distinct skill labels.

In order to integratively represent these together, we propose that the "pharmacist" here is an occupation holder having the role of 'pharmacist role". When a person assumes this role, that person is a pharmacist. In a conference of pharmacists and dentists, if pharmacists eat in one hall and dentists in another, these are two sets of humans. Therefore, ideally, we would have the following definition:

'pharmacist' ('is a' person):

- 'has role' some ('pharmacist role' and realized_in some 'pharmacist occupation activity')
  (*note*: the term 'pharmacist role' and the axiom may be hidden and not represented in OccO to reduce duplications.)

- 'has skill' some 'active listening' (and some other specific skills)
- 'has ability' some 'oral comprehension' (and some other specific abilities)
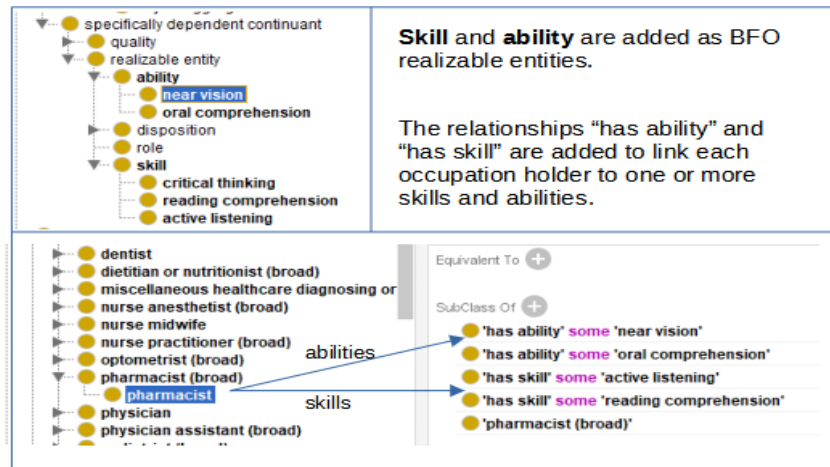

Figure 3 OccO representation of skill, ability and properties "has skill" and "has ability"

3.7. Use case: pharmacist in OccO vs Wikidata

For a use case to contrast OccO with Wikidata, the treatment of pharmacist will be used as illustrative. Wikidata describes "pharmacist" as a "healthcare professional who practices in pharmacy," similar to OccO in associating this occupation with the occupation holder. Merriam-Webster defines a pharmacist as a person: "a health-care professional licensed to engage in pharmacy with duties including dispensing prescription drugs, …".
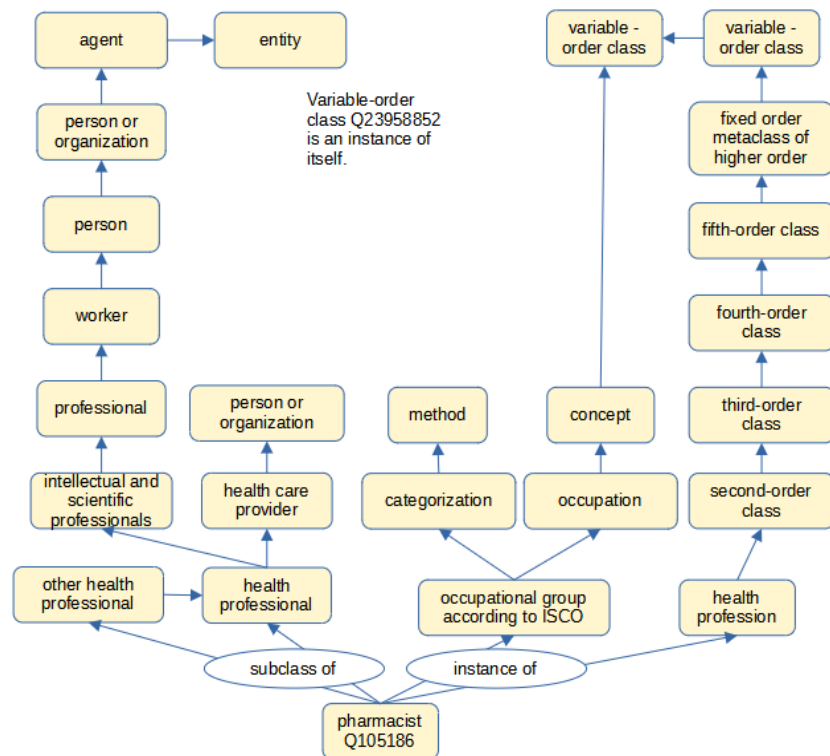

Figure 4: Wikidata representation of pharmacist.

As shown in Figure 4, Wikidata characterizes pharmacist as the "subclass of" two entities, and an "instance of" two other entities, one of which is an instance-of occupation (Q12737077), which

Wikidata defines as a "label applied to a person based on an activity they participate in". This means a pharmacist is ultimately an instance of a label. In total, two dozen entities appear as parents of pharmacist, and several branches do not extend to the subclass-of root term "entity" or instance-of root term "variable order class" (the latter being of no ontological use for OccO.) The hierarchical structures based on "instance of" and "subclass of" are so incomplete and inconsistent that they do not facilitate automatic harmonization with OccO. The correspondence would need to be based on the occupation labels – but in Wikidata, although 17,000 occupations are identified, only 17% of the ISCO labels are found among Wikidata occupations.

In contrast to Wikidata, OccO characterizes the occupation holder as a person, and the occupation role, skills and abilities as object properties possessed by (or that "inhere in") the person.

3.8.   Characterizing by object properties

Object properties in addition to skill and ability can be added to the characterization of occupations. These include field of work, environment (land, sea, air, subterranean, etc.), materials or devices, tools employed. Each possible property needs to be assessed for its utility. Field of work is a key property, and one that would be incorporated in an expanded OccO. An existing taxonomy of industries provides a good initial model, the International Standard Industrial Classification of All Economic Activities (ISIC). A Field of Work Ontology (possibly FOWO) could be created from the ISIC, and linked into OccO. For our pharmacist example, the utility of Field of Work can be shown by distinguishing the occupation of Ms. Jones as community pharmacist, from Mr. Brown, the prison pharmacist. The former would correspond to ISIC code 45611, "Pharmacies and Drug Retailers", and the latter to 92214, "Correctional Institutions".

There is an existing BFO based Environment Ontology that can be linked to OccO, and it already has an entry for "prison." An object property for environment would be helpful - in some cases it could be assigned to the field of work and others to the occupation holder, if the field of work can occur in diverse environments. Another major dimension of occupations is the services they represent or fulfill. A service is a provision of one or more processes that achieve some objective. Integrating this service dimension would help delineate and define the occupation in question.

4. Conclusion: OccO is feasible and beneficial

This presentation describes an initial foray into the development of an Occupation Ontology (OccO), with the anticipation that a more extensive and broadly supported activity may build on this initial work and generate an ontology suitable for widespread use. The use of OBO Foundry principles is recommended. The restriction of scope to the areas encompassed by popular occupation taxonomies, focusing on vocations, permits the wealth of object property information that already exists to be employed. Employing the BFO upper level ontological framework would illustrate a new implementation of the first internationally standardized ontology.

Our selected example taxonomy, the US SOC, allowed the major groupings to be aligned with field of work. However, with field of work handled as an object property, the use of the ten ISCO skill-level related major groups may be preferred. The EU ESCO implementation of ISCO has significant enhancements, and it is suggested as a better base to start from, followed by incorporation of additional elements from the other standard taxonomies.

Careful selection and use of object properties would significantly enhance OccO, suggesting that the ISIC be used to capture field of work, and the existing BFO Environment Ontology ENVO be used for the environment property. Skills, abilities, and areas of knowledge are represented among the standard taxonomies, but harmonization among these references is needed and would be a beneficial accomplishment. Wikidata or other Semantic Web resources could employ OccO results to better categorize occupations; these resources are too important not to have a solid ontological structure.

Once the occupation ontology framework is in place, it could be the model for additional spheres of human activity: Avocations, Family Roles, Social Roles could all be implemented within this framework. Having such an integrated representation of human activity may have widespread benefits in coordinating world activities in medicine, social policy, education, employment and social services. Our pharmacist Mr. Brown, could be a prison pharmacist, a model railroader, hold social roles as a deacon, veteran, and member of the National Guard, all within the scope of an expanded OccO. We hope that the advantages of an occupation ontology as presented here will gain the interest and support of one or more institutions that could benefit from having such a resource.

5. Acknowledgements

6. References

[1]   *United Nations International Labor Organization (ILO)*. Available: https://www.ilo.org/global/about-the-ilo/history/lang--en/index.htm

[2]   *US Bureau of Labor Statistics Standard Occupational Classification (US SOC)*. Available: https://www.bls.gov/soc/

[3]   *UK National Statistics Stand Occupational Classification (UK SOC)*. Available: https://www.ons.gov.uk/methodology/classificationsandstandards/standardoccupationalclassificationsoc

[4]   *The European Skills, Competences, Qualifications and Occupations (ESCO) of the European Union*. Available: https://esco.ec.europa.eu/en/about-esco/what-esco

[5]   P. L. Buttigieg, N. Morrison, B. Smith, C. J. Mungall, S. E. Lewis, and E. Consortium, "The environment ontology: contextualising biological and biomedical entities," *J Biomed Semantics,* vol. 4, p. 43, 2013.

[6]   M. A. Musen, "The Protégé project: A look back and a look forward. AI Matters. ," *Association of Computing Machinery Specific Interest Group in Artificial Intelligence,* vol. 1, p. DOI: 10.1145/2557001.25757003., 2015.

[7]   E. Ong, Z. Xiang, B. Zhao, Y. Liu, Y. Lin, J. Zheng*, et al.*, "Ontobee: A linked ontology data server to support ontology term dereferencing, linkage, query and integration," *Nucleic Acids Res,* vol. 45, pp. D347-D352, Jan 04 2017.

[8]   B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters*, et al.*, "The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration," *Nat Biotechnol,* vol. 25, pp. 1251-5, Nov 2007.

[9]   R. Arp, B. Smith, and A. D. Spear, *Building Ontologies with Basic Formal Ontology*. MIT Press: Cambridge, MA, USA, 2015.

[10]   R. Jackson, N. Matentzoglu, J. A. Overton, R. Vita, J. P. Balhoff, P. L. Buttigieg*, et al.*, "OBO Foundry in 2021: operationalizing open data principles to evaluate ontologies," *Database (Oxford),* vol. 2021, Oct 26 2021.

[11]   Y. He, S. Smith, B. Smith, J. Zheng, Y.-F. Li, and E. Ong, "WikiO: A BFO-based ontology to support Wikipedia data integration and analysis," in *The 11th International Biocuration Conference, April 8-11, 2018*, Shanghai, China, 2018.

[12]   P. Price and S. Miner, "Occupation emerges in the process of therapy," *Am J Occup Ther,* vol. 61, pp. 441-50, Jul-Aug 2007.

[13]   D. Pierce, "Untangling occupation and activity," *Am J Occup Ther,* vol. 55, pp. 138-46, Mar-Apr 2001.

[14]   D. L. Nelson, "Therapeutic occupation: a definition," *Am J Occup Ther,* vol. 50, pp. 775-82, Nov-Dec 1996.

[15]   C. B. Royeen, "Occupation reconsidered," *Occupational therapy international,* vol. 9, pp. 111-120, 2002.

[16]   *The O*Net Content Model*. Available: https://www.onetcenter.org/content.html

[17]   D. Vrandečić and M. Krötzsch, "Wikidata: a free collaborative knowledgebase," *Communications of the ACM,* vol. 57, pp. 78-85, 2014.

[18]   S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, "Dbpedia: A nucleus for a web of open data," in *The semantic web*, ed: Springer, 2007, pp. 722-735.

[19]   *WikiData Help: Basic membership properties*. Available: https://www.wikidata.org/wiki/Help:Basic_membership_properties

[20]   R. Arp and B. Smith, "Function, role, and disposition in basic formal ontology," *Nature Precedings,* pp. 1-1, 2008.

[21]   E. C. Merrell, D. G. Limbaugh, P. M. Koch, and B. Smith, "Capabilities," *Philpapers,* pp. 1-26. https://philpapers.org/archive/MERC-14.pdf, 2022.