The Occupation Ontology (OccO): Building a Bridge between Global Occupational Standards

John Beverley[1,*], Sam Smith[2,*], Matthew A Diller[3,*], William D. Duncan[3], Jie Zheng[4], John W. Judkins[4], William R. Hogan[3], Robin McGill[4], Damion Dooley[6] and Yongqun He[2]

1 University at Buffalo, Buffalo, USA.
2 University of Michigan Medical School, Ann Arbor, MI, USA.
3 University of Florida, FL, USA.
4 University of Pennsylvania, Philadelphia, PA, USA.
5 Alabama Commission on Higher Education, AL, USA.
6 Simon Fraser University, Vancouver, BC, Canada.

** Contributed equally to the work.*

Abstract

The Occupation Ontology (OccO) is a community-based ontology for occupations, which extends from the upper-level Basic Formal Ontology (BFO) in accordance with Open Biological and Biomedical (OBO) Foundry principles. In this article, we report on updates to core OccO definitions after expanding representational coverage to include codes from US SOC/O*NET and ESCO, as well as information from Wikidata. We also report on OccO working group collaborations with the ESCO team in the interest of integrating ESCO terms into OccO, and the Alabama ontology project team in the interest of representing competencies and credentials. We close by highlighting our strategy for integrating Wikidata and occupation standards using OccO as a *lingua franca*. We outline various updates and challenges throughout and encourage further participation from the community as we pursue this effort.

Keywords

Occupation, OccO, ISCO, ESCO, US SOC, O*NET, OBO Foundry, Basic Formal Ontology

1. Introduction

The Occupation Ontology (OccO) [1] is designed to facilitate ontological representations of existing occupation standards, such as the US Bureau of Labor Statistics Standard Occupational Classification (US SOC) [2], the International Standard Classification of Occupations (ISCO) [3], the UK National Statistics Standard Occupational Classification (UK SOC) [4], and the European Skills, Competences, Qualifications and Occupations (ESCO) of the European Union 2010 [5]. Initial development of OccO was reported in JOWO2022 [6]. This version of OccO focused on US SOC - and its companion Occupation Information Network (O*NET) database - to provide a proof of concept for ontology modeling. Details of OccO were subsequently presented at the 2022 International Conference of Biological and Biomedical Ontologies (ICBO 2022) [7], which resulted in an expansion of the OccO working group to include representatives working on ontology projects in nearby domains from the University of Florida, University at Buffalo, and University of Pennsylvania. For example, the Ontology of Medically Related Social Entities (OMRSE) [8] - an Open Biological and Biomedical Ontology Foundry [9, 10] ontology - has a need for occupation coverage, and representatives have accordingly worked to identify overlap, opportunities for reuse, and co-development between OMRSE and OccO. Moreover, the OccO development team has curated relationships with representatives from ESCO and has opened lines of communication with an occupation ontology development group associated with the Alabama Committee on Credentialing and Career Pathways (ACCCP), a committee of the Alabama Workforce Council tasked with setting standards for in-demand occupations across the state and identifying credentials of value that can prepare individuals to gain employment within those occupations [11].

Proceedings International Workshop on Ontologies for Services and Society, July 17–20, 2023, Sherbrooke, Canada EMAIL: johnbeve@buffalo.edu (A. 1); smsmith508@aol.com; yongqunh@med.umich.edu (A. 2); mcgill.r.e@gmail.com. ORCID: 0000-0002-1118-1738 (A. 1); 0000-0001-5193-942X (A. 2)

We intend to leverage OccO as a common framework for distinct coding standards. US SOC, UK SOC, ISCO, and ESCO standards exhibit significant overlap in coverage, though they differ with respect to representations of occupation: metadata attributes, skills and abilities, as well as descriptions of associated tasks. Traditionally, mappings – or *crosswalks* – between occupation standards have been created to promote interoperability. Crosswalks become, however, outdated following occupation standard updates, often require significant manual curation, and susceptible to the "order n-squared" problem [12]. To illustrate, suppose crosswalks exist between US SOC and ESCO. Adding UK SOC requires four new crosswalks, two mapping UK SOC to ESCO and vice versa, and two mapping to US SOC and vice versa. This is unsustainable growth; as occupation standards are updated, new versions require n^2 new crosswalks, where n is the number of occupation standards. In the presence of a common ontology language, such as that provided by OccO, the order n-squared problem dissolves. For any n distinct coding systems, maintaining new additional systems requires one bridge into OccO.

Recently, NLP strategies have been deployed to address some of these limitations. Arguably, it should not matter how many crosswalks are needed if they can be automatically generated and maintained. For example, a 2022 EU Commission technical report [5] described the deployment of a BERT-based NLP strategy to develop a crosswalk between ~3000 ESCO and ~1000 O*NET occupations. Pairs of occupation titles were ranked as exact, broad, related, or inexact based on similarity scores between occupation labels and textual descriptions. This strategy, however, resulted in 7385 matches, which then had to be manually inspected for accuracy. After inspection, ~500 were deemed exact, ~200 narrow, and ~2000 broad matches, while ~600 had no match. Such a result suggests current NLP strategies may result in a need for more, rather than less, manual inspection. Again, it seems wise to avoid the need for crosswalks, or at least minimize the number needed.

These observations have demonstrated that there is a clear, pressing, need for constructing a comprehensive occupation ontology that harmonizes existing standard language taxonomies, represents major occupations, and features such as skills, abilities, credentials, and competence, and that facilitates bridging across occupation coding standards. Accordingly, OccO is designed as a semantically rich *lingua franca* into which occupation standards and occupation data can be mapped to promote interoperability and avoid the n-squared problem. To that end, we report on ontological refinements of key OccO terminological content, as well as our successes and outstanding challenges associated with incorporating other occupation terminologies within its purview. More specifically, we detail subtleties that emerged in key OccO definitions when placed in conversation with additional occupation standards, provide an updated design pattern for core OccO classes and relations, expand on our previous treatment of skills and abilities, and report our use of Wikidata to identify potential coverage issues for ESCO and US SOC and enrich OccO. Lastly, we outline challenges encountered, and invite researchers interested in collaboration to aid in efforts to build a bridge across international occupation standards.

2. Refinement of the Core OccO Design Pattern

OccO is designed as an extension of the Basic Formal Ontology (BFO), which is a top-level ontology covering general classes such as *material entity, quality, process, function*, and *role* [13] and providing general architecture for approximately 500 ontology projects, such as those found in the OBO Foundry, the Industrial Ontology Foundry (IOF) [14], and the Common Core Ontologies suite [15]. BFO is, moreover, the only ISO/IEC 21838 approved top-level ontology standard [16]. Given its wide use, standardization, and well-developed architecture, BFO is a natural starting point for exploring ontological representations of occupations. Moreover, because of its wide use and longevity, many of ontology terms that OccO will ultimately adopt regarding specific skills, abilities, and occupations, exist in BFO-conformant ontologies. Starting with BFO thus allows us to cut down on future ontology engineering work. BFO is designed to characterize the most general classes relevant to scientific investigation [17] and divides reality into disjoint categories of *continuant* and *occurrent*. Instances of *continuant* exist entirely at any time at which they exist and lack temporal parts. Instances of *occurrent*, in contrast, have temporal parts and so are often stretched along a temporal axis. BFO adopts several further subclasses of *continuant* and *occurrent*. Both sides are needed for our work. Two subclasses of *continuant* are of particular interest: *disposition* and *role*.

Instances of *disposition* are *realizable entities* - entities that if realized, are realized in processes; they are said to be "internally grounded", a metaphor meant to track the fact that were the instance to cease to exist, then its bearer would be physically changed [18]. For example, if a piece of sodium chloride is soluble at some time, then not soluble at another, there must be some change to its physical makeup. In this sense, instances of *disposition* are not optional for bearers. Realizations of instances of *disposition* occur, moreover, owing to the material constitution of the bearer – and its associated qualities - and the fact that the bearer is in some environment the material bearer is not always in. The realization of salt's solubility occurs owing to the lattice structure and bonding forces of the salt when placed in unsaturated water. The class *role* is a disjoint sibling class of disposition, and so a *realizable entity*, but one with characteristics sharply distinguishing it from its sibling. In contrast to *disposition*, instances of *role* are optional in the sense that bearers may gain or lose them without physical change. A student, for example, who graduates from a university no longer bears the *role* of student at that institution. That does not, however, imply any physical change in the student. More generally, whether an entity bears a *role* depends largely on what happens external to the entity.

2.1 Hallmarks of "Occupation"

Natural language and relevant occupation standards suggest "occupation" is often used in either a *disposition* or *role* sense. On the one hand, "occupation" may be used synonymously with "job". For example, when asked to describe one's occupation, it is not uncommon to hear responses such as "cashier" or "physician" or "professor". As another example, the ISCO defines an occupation as "[T]he kind of work performed in a job" and 'job' as "[A] set of tasks and duties performed, or meant to be performed, by one person, including for an employer or in self-employment" [19]. A job is most naturally understood as a type of *role* in BFO, as individuals may gain or lose jobs without necessarily undergoing material change.

On the other hand, "occupation" may be used to describe a set of capabilities one has independent of whether they also have a job. For example, a Python developer in between jobs may be described as having a developer occupation, despite not being employed. Indeed, Classification Principle 2 of US SOC states that "Occupations are classified based…in some cases, on the skills, education, and/or training needed…" to perform a job [2]. Additionally, the US Bureau of Labor and Statistics - which employs US SOC to model occupation data – relies on categories such as "unemployed persons by occupation and sex" [20]. This sense of "occupation" appears best understood as a type of *disposition* in BFO, as the skills and abilities borne by an agent are grounded in the material basis of that bearer, such as mental competencies or physical acumen. Moreover, occupations understood as a type of *disposition* are clearly related occupations understood as a type of *role*. A skilled, unemployed, jobseeker has desirable abilities to hiring agencies. Skills are often why agents are given job.

We maintain that the use of "occupation" should contrast with the processes associated with occupations. An employed Python developer is authorized to participate in code review for an organization; an unemployed Python developer has a skillset manifested in repository updates, open-source contributions, etc. This position contrasts with, for example, the American Occupational Therapy Association which defines occupation as "Various kinds of life activities in which individuals, groups, or populations engage, including activities of daily living, instrumental activities of daily living, rest and sleep, education, work, play, leisure, and social participation" [21]. Besides such a definition being so broad as to include nearly all human activity, applying this definition to occupation standard data would conflate jobs, occupations, and activities associated with either.

There are identifiable commonalities across the above uses of "occupation". Occupations are inherently transactional. Employers seek talent to fill job and unemployed individuals having occupations often seek employment. This should not suggest that anyone hired for an occupation is thereby compensated for work performed. Employees *should* be compensated, but employers do not always behave as they should; an unpaid employee who deserves compensation may count as holding an occupation in the job title sense, nevertheless. Similarly, individuals holding occupations may not perform corresponding duties or perhaps may perform them poorly. We do not say, however, that a poor performer does not hold the relevant occupation.

This leads to points of ontological complexity worth belaboring: competency levels and relevant intentions. Consider Chris, a cashier at a local grocery store who – despite lacking training or having previously expressed interest in doing so – decides he has a computer programmer occupation and intends to develop the skills needed to obtain a programming job. It is implausible to count Chris as having a computer programmer occupation simply because of his stated declaration; he needs to do more to be counted as having such an occupation. Though we perhaps cannot draw a firm line between when one has sufficient relevant skill to count as having an occupation and when one does not, we can point to paradigmatic cases where individuals clearly have an occupation and where not. In this example, Chris lacks relevant skills. That said, Chris's intention to develop relevant skills and seek employment, seems an important component to having an occupation. Despite being a skilled cashier, after Chris quits his cashier job and decides to no longer pursue cashier jobs, he should not be counted as having a cashier occupation. Worth noting, the ISCO definition of "occupation" supports the importance of intentions to having an occupation, stating that occupations are associated with activities performed and activities intended to be performed [3]. In short, to count as having an occupation one must bear an intention to develop relevant skills, seek employment with the relevant titles, and perform relevant work; appropriate intentions with appropriate competencies are necessary. In contrast, appropriate intentions are not necessary to hold a job, as a job may be given to someone with no intention to work or borne by someone who loses any intention to continue working. Observations in hand, we summarize our discussion by identifying the following hallmarks of occupations:

1. Occupations are ultimately transactional.
2. When understood as jobs, occupations require occupations holders be empowered to perform tasks associated with that occupation.
3. When understood as abilities or skills, occupations holders must be capable of and intend to perform tasks associated with that occupation.

The preceding hallmarks were used to refine OccO definitions and construct additional terminological content needed to pursue our goals of bridging occupation standards.

2.2 OccO Design Pattern

Figure 1 illustrates the core OccO design pattern, which has been re-engineered to reflect the preceding hallmarks of occupation. We import the class *human* from the NCBITaxon [22] as a bridge between the BFO class *material entity* and those who hold occupations. An *occupation holder* is someone bearing either an occupation role or an occupation disposition. This class captures the bearers of occupations, such as a pharmacist, welder, ontologist, etc. The substance of this description of *occupation holder* is found in the definitions of *occupation role*, *occupation disposition*, *skill*, and *ability*.

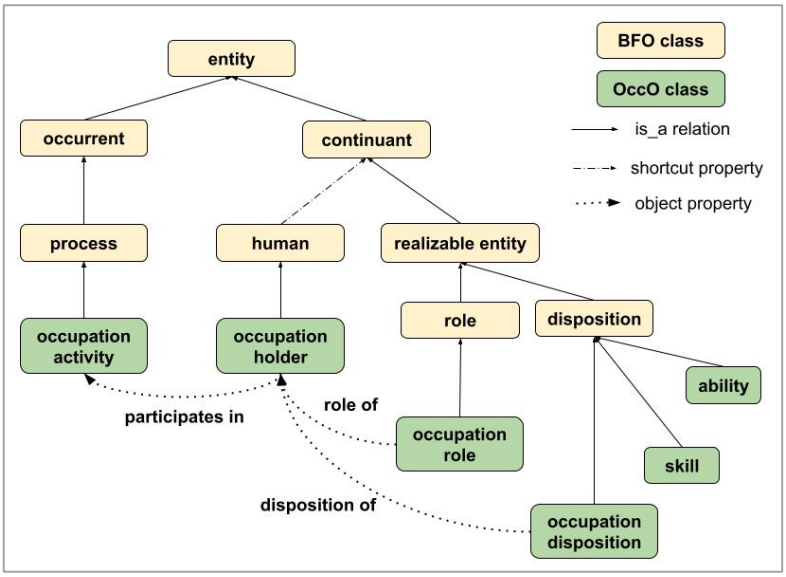

Figure 1: Key OccO Classes Extending from BFO

The class *occupation role* is meant to capture the "job" sense of occupation. Bearers of such roles provide labor or services in exchange for compensation, and the existence of such a role requires an authority capable of empowering the bearer to perform relevant tasks for compensation. Individuals are often hired to perform tasks based on perceived ability and skill to do so. Accordingly, OccO introduces *ability* and *skill*, which fall under *disposition*. Regarding the former, *ability* represents dispositions - naturally acquired during an individual's development – to perform tasks, e.g., walking, jumping, eating. As such, *ability* is a rather broad class. Even so, abilities relevant to an occupation are bound by the scope of duties associated with that occupation. A surgeon may have the ability to jump, but a hospital employer would likely not find that as interesting as, say, being able to communicate with patients. Regarding the latter, *skill* represents dispositions emerging from training that involve a high-level of proficiency and which can be exercised reliably. The hospital employer would likely find a potential surgeon's dispositions to successfully complete open-heart surgery, placing a stent, etc. to be of significant interest. Training, reliability, and proficiency distinguish *skill* from *ability*, though both are needed to characterize *occupation disposition*.

We treat *skill* and *ability* as siblings, which may seem odd since intuitively skills are simply abilities that have been honed through training. However, classifying *skill* as a subclass of *ability* would entail that any instance of *skill* counted as an instance of *ability*, and it seems mistaken to say that in every case once one cultivates a skill, they thereby gain an ability. First, developing a skill involves many abilities; there is a many-one relationship between abilities and skills. Second, developing a skill should not entail that one thereby gains a disposition of the sort naturally acquired during an individual's development. Third, treating *skill* as a subclass of *ability* runs the risk of suggesting that for every skill there is some ability that was trained into that skill, which is false. One does not become an expert programmer by training up a native programming ability. Consequently, we do not classify *skill* as a subclass of *ability*.

The class *occupation disposition* is defined partly in terms of some *ability* relevant to an *occupation role* and partly some *skill* aimed at performing tasks associated with that role reliably and proficiently. However, *skill* and *ability* are insufficient alone to capture the hallmarks of this sense of occupation. One may have skills and abilities relevant to a job but have no intention to work in such a position. Thus, to adequately characterize *occupation disposition* we must appeal to the intentions of bearers to pursue work of a sort relevant to an occupation [23]. Discussion of the precise nature of intentions is notably complex and beyond the scope of our work [24]. For our purposes, in general we understand intention as the directedness - or aboutness towards - of some cognitive process [25], and we understand intention as the directedness of a bearer's mental representation towards pursuing work relevant to an occupation. In short, *occupation disposition* is defined in terms of relevant abilities, skills, and intentions of a bearer. Lastly, occupation holders participate in instances of *occupation activity*, which are - roughly - processes in which such participants engage in pursuits associated with associated job titles or intended career paths. Definitions of key OccO terms are found in **Table 1**.

Table 1. Key OccO Definitions

OccO Term	Definition
occupation holder	A human bearing an occupation role or occupation disposition.
occupation role	A role borne by a human that, if realized, is realized when the bearer provides labor or services in exchange for compensation as specified by some deontic declaration.
occupation disposition	A disposition that, if realized, is realized when the bearer intends to, and does, exercise abilities and skills in pursuit of obtaining or maintaining an occupation role.

ability	A disposition that inheres in a human, exists in virtue of natural biological development, and is such that if realized, is realized in the performance of one or more tasks.
skill	A disposition that inheres in a human, exists in virtue of training, and is such that if realized, is realized in the performance of one or more tasks with reliable high proficiency.
occupation activity	A process in which an occupation holder participates that realizes either an occupation disposition or occupation role.

3. OccO as a Hub for ESCO and O*NET Spokes

The Alpha version of OccO mapped the US SOC/O*NET occupation categories and detailed occupation labels, but our aim is to integrate other occupation categories under the umbrella of OccO. For example, the ESCO taxonomy overlaps significantly in terminology with US SOC, thereby facilitating mapping to OccO. Even so, integrating ESCO with OccO representations of US SOC/O*NET is challenging. ESCO shares the same 10 Major Groups with ISCO, represented with a single digit code. In contrast, US SOC includes 23 Major Groups. While both ESCO and US SOC/O*NET use a four-layer hierarchy of intermediate classes, in each case "dotted suffix" codes are used to extend the ISCO and US SOC/O*NET codes. An example is "barista" where the ESCO code is 5132.1.1 using the four-digit ISCO code and the corresponding O*NET code is 35-3023.01, using the SOC seven-position code. It is important to preserve the existing group categories because many reports are summarized by major or minor group.

As discussed in the introduction, there have been recent attempts to employ NLP strategies to automate crosswalks between occupation standards. One result of this growing interest has been "Dynamic Ontology Matching Challenge" [26], a community call to action to automate – to the extent possible, while acknowledging there will likely need to be significant manual labor involving in such strategies. The OccO development team has had several meetings with ESCO team members Gianluca Bortoletto and Jan Luts – authors of the NLP technical report outlined in the introduction - over how OccO might be employed to address this challenge. To that end, we propose a schema for connecting codes from distinct occupation standards along the following lines, using "barista" as an example:

```
OccO:barista OccO:has_ESCO_code "5132.1.1"^^xsd:string ;
OccO:has_O*NET_code "35-3023.01"^^xsd:string .
```

In words, for a given category common between these occupation standards, we introduce a relevant occupation code datatype property with the corresponding literal value. In this way, the respective codes are linked in OccO by – in this case – the OccO class *barista*.

Exact matches are straightforward, but do not address all mapping challenges. In many cases, a term in ESCO is a broad or narrow match to an O*NET term; in other cases, the relationship between occupations in these respective standards is unclear. Our current strategy for capturing exact matches in OccO is illustrated above with *barista*. For close and inexact matches, we employ the following strategy:

```
OccO:Example1 OccO:has_ESCO_code "###"^^xsd:string ;
              OccO:has_O*NET_inexact_match OccO:Example2 .
OccO:Example2 OccO:has_O*NET_code "'$$$"^^xsd:string ;
              OccO:has_ESCO_inexact_match OccO:Example1 .
```

In other words, the term XYZ which bears the ESCO code ### also bears an inexact match relationship to OccO:IRI123, which bears the O*NET code $$$. Additionally, because ESCO and O*NET provide descriptions of relevant abilities and skills for these occupations, OccO can facilitate semantic representation of the similarities and differences between them. Worth stressing is that the preceding proposal is tentative, as we are considering adopting the SSSOM mapping standard [27].

4. OccO Integration with the Alabama Talent Triad

A program related to OccO is being conducted by the Alabama Committee on Credentialing and Career Pathways (ACCCP), with the leadership from the Governor's Office of Education and Workforce Transformation (GOEWT). This group has developed an informal Alabama occupation ontology as part of its Talent Triad initiative, which aims to connect individuals, employers, and credential providers using a common skills-based framework. As its name suggests, the Talent Triad has three components: Skills-Based Job Descriptions, Learning and Employment Records, and Skills Defined Credentials [28].

Figure 2: Alabama Talent Triad Design

The Talent Triad pilot phase [29] will be accessible through a software application that relies on so-called "Skills DNA" to support algorithmic matches between individuals seeking employment and employers. The Skills DNA model relies on an informal ontology of skills and/or competencies [30]. Alabama's ontology took as its starting point the Building Blocks Model developed by the Competency Model Clearinghouse within the Employment and Training Administration of the U.S. Department of Labor, which organizes competencies in a tiered structure with increasing specificity, culminating with occupation-specific competencies. Through work with experts including Corporation for a Skilled Workforce [31], Alabama's ontology has made several enhancements, such as grouping skills/competencies into standard "functions" that can be contextualized to specific occupations. Approximately 350 high-demand occupations from US SOC/O*NET are targeted.

In addition to matching individuals with job opportunities based on skills they currently possess, the Talent Triad will suggest additional education and training that an individual can pursue to increase skill level or acquire new skills. This feature relies on credential information being organized in a standardized way around skills and/or competencies. Alabama, along with over 20 other states, has partnered with non-profit organization Credential Engine [32, 33] to represent credential information using the Credential Transparency Description Language (CTDL) [31], written in OWL. CTDL allows for the representation of key descriptors for credentials, including the organization offering the credential, its requirements, and alignment with specific occupations. In addition to serving as the clearinghouse for the development of CTDL, Credential Engine maintains an extensive database of hundreds of thousands of credentials known as the Credential Registry, and Alabama will continue to publish credential information from the Talent Triad to the registry maintained by Credential Engine.

A key element of the Alabama program is the articulation and development of skills and abilities associated with the target occupations, described as credentials and competencies. While under development, the OccO working group is considering the following representations of such phenomena:

- Following [34], *credential* can be understood as a *document* that reflects an attestation to the holder's competence with respect to a given task, at or above a certain threshold. Accordingly, credentials are issued by some authority and are numerous.
- Related, *competence* can be understood in terms of measurements of the degree to which the combination of skill and/or ability has been developed. Accordingly, competence deals with evaluations of an individual's ability to perform a job.

By focusing on skills, abilities, credentials, and competence, developers of ACCCP aim to unite employers, educators, and members of the workforce with a common ontological language. In this respect, the aims of ACCCP align with the aims of OccO, as we too seek to integrate and enrich occupation data using standardized language. Future development of OccO will seek to align not only our aims, but also our respective understandings of these phenomena.

5. Wikimedia Use Case

As a concrete illustration of how we understand OccO's utility, we turn to Wikimedia resources, such as Wikipedia, Wikidata, and DBpedia. Wikipedia contains significant quantities of occupation data as well as references to occupation holders. DBpedia is a linked open data project aimed at extracting structured content from Wikipedia to be placed in a triple store for semantic querying and analysis. Wikidata is a vast semantic web resource used to generate linked open metadata as a supplement to Wikipedia pages. Wikipedia is used primarily as a crowd-sourced publicly editable information storehouse. DBpedia, on the other hand, is used more often by researchers, and is updated infrequently. Wikidata may also be edited by users directly. Consequently, Wikidata needs curation, hence our focus on this Wikimedia product for our OccO use case.

For example, we retrieved a total of 4,617 people from Wikidata who died from COVID-19 as of January 2022. One might be interested in investigating the classes of occupations represented among this population, perhaps to explore occupation risks associated with the disease. Such analyses would, however, require robust semantic structure, which Wikidata does not exhibit. Of the 4,617 with both date of death and one or more occupations, the result is 4,007. **Figure 3** shows the number of deaths per month for people in Wikidata through January 2022.

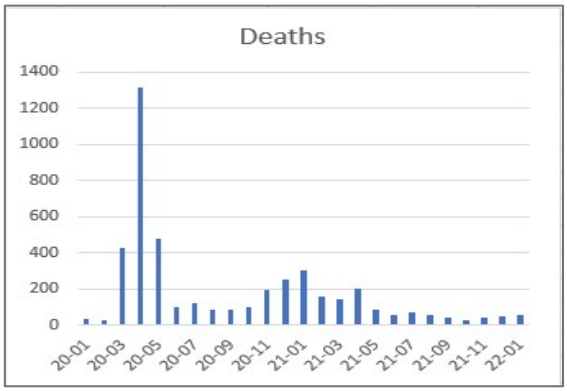

Figure 3: Deaths from COVID-19 according to Wikidata.

Data about individuals in Wikidata exhibits a triple structure, where resources are connected via properties. Two important identifiers and labels are listed in **Table 2**.

Table 2

Wikidata Resources Relevant to Occupation

Identifier	Label	# of Occurrences
Q12737077	occupation	~3000
P106	has occupation	~17000

Of the occupation relation occurrences in Wikidata, only a quarter of these match a resource classed as a Q12737077 occupation or descendant. Moreover, perusal of occupations in Wikidata revealed that only 772 of its occupation labels are associated with ISCO-2008 occupation classes, while 397 are associated with ESCO occupation identifiers, 600 with ESCO skill identifiers, 112 with US SOC identifiers, and 85 with O*NET codes. Each Wikidata label is associated with a specific number of occupation labels in the Wikidata taxonomy, as illustrated in **Table 3**. Notably, the listing of ISCO occupations in Wikidata is complete, following a comprehensive update of ISCO resources, such as breaking up terms such as “mathematician and actuaries” into individual occupations.

Table 3

Wikidata Occupations Associated with Occupation Standards

Wikidata Label	Wikidata Code ID	# of Labels Matching Occupation Code
ISCO-08 occupation class	P8283	772
ESCO (V1) Occupation ID	P4652	397
ESCO (V1) Skill ID	P4644	600
SOC ID	P919	112
*O*NET*	P8734	85

Wikidata lacks ontological rigor despite exhibiting triple structures. This makes integration with Wikidata challenging, if pursued in a systematic fashion. For example, in Wikidata, the class

"occupation entity" (Q12737077) is an instance of two parents: "concept" (from which all its entities descend) and "second-order class" (a term that we have not yet found a formal ontology using). The entity "profession" (Q28640) is an instance of "Wikidata meta class" and "second-order class". Therefore, querying for all instances of "occupation" will miss the many occupations that are instances of "profession." More specific examples are also provided in our previous OSS2022 paper [6].

Connecting Wikidata to other occupation standards will be made easier by using OccO. Our strategy – illustrated in **Figure 4** – is to represent distinct occupation terminologies within the umbrella of OccO. It is our belief that such an approach will lead to the achievement of data integration across occupation standards. Roughly speaking, the strategy is to use OccO as a *lingua franca.* This allows for the enrichment of these standards with ontological structure; such enrichment facilitates the use of a query language like SPARQL to run quality control over datasets ontology, avoid scope creep, all while maintaining a common language into which each standard can be translated to promote interoperability.

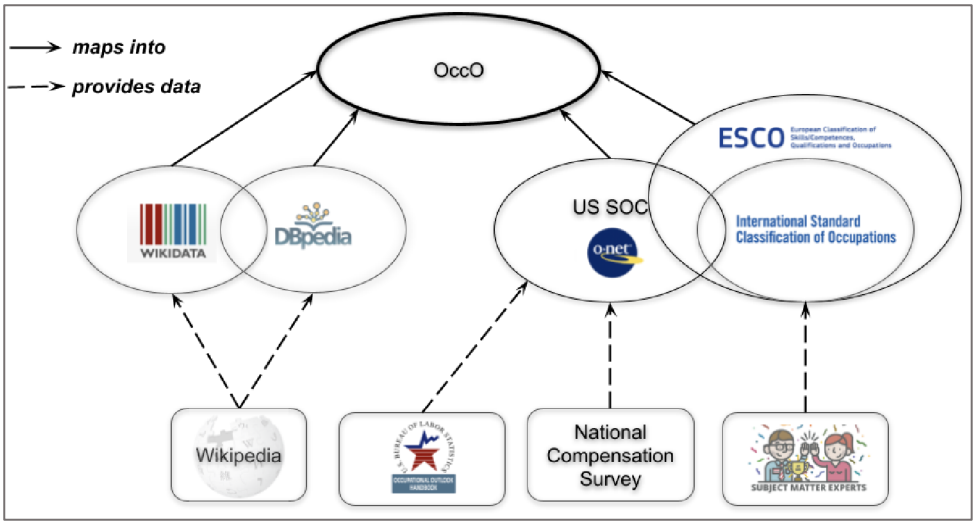

Figure 4: Proposed OccO Integration across Standards and Datasets

A major goal going forward will be to implement the above strategy while attempting to apply some formal structure to domains. By doing so, we will then be empowered to analyze individuals with different occupations or occupation groups, and their associations with different scenarios such as death due to infectious diseases like COVID-19.

6. Discussion

We have described several contributions in this article. First, we formed an OccO working group consisting of researchers from a wide variety of academic institutions. Second, we achieved consensus on the definitions of occupation-related terms within the BFO/OBO framework. Third, we introduced and developed our framework in discussion with representatives of similar efforts in the ESCO and Alabama GOEWT, in the interest of benefitting from subject-matter expertise and promoting stakeholder investment. Fourth, we investigated the use of the OccO ontology in studying Wikidata resources of various disintegrated occupation types and individuals.

Next steps include incorporating ISCO and ESCO occupation titles into OccO, after which work may begin on harmonizing ESCO skills with corresponding O*NET skills and abilities, as each set of skills will be represented in OccO and associated with occupation classes. Additionally, our work will focus on achieving consensus on ontology content needed to represent specific skills, abilities, credentials, and competencies. Given how important skills and abilities are to US SOC/O*NET and ESCO, and how important credentials and competencies are to many relevant stakeholders working in this space, incorporating this content into OccO is of crucial importance.

Our collaboration between the OccO project and Alabama project will bring many benefits and applications. OccO will bring these skill-ability-educational properties into some degree of alignment

or uniformity. It is possible to combine O*NET skills and abilities properties along with Credential Engine properties to harmonize these important occupational properties. It would be a worthwhile exercise to see to what degree credentials that are now widely diverse could be organized into a more orderly and common structure, using CTDL and OccO.

The usage of the three Wikimedia resources allows us to access thousands of persons with specific occupation types, which may be standardized and integrated using our ontological strategy. However, our study found that the classification of occupations in Wikidata is uncurated and confusing. Therefore, we propose to use OccO for standardizing and integrating these resources, which will significantly improve our study of occupation related instance data and allow us to address various occupation related questions. One further issue worth noting that stems from working with Wikidata is that because this data source is ultimately crowd-sourced, it may contain false or misleading information concerning occupations. We envision that stakeholder engagement, subject-matter expertise, and careful representation of Wikidata occupation content in OccO, will provide a crucible in which to uncover inconsistencies, identify misinformation, and correct it through careful analysis. In this respect, the OccO workflow will act as a line of defense against such misinformation.

Many future directions are possible for our OccO working group research. The ISCO is updated every twenty years, with the next revision in 2028. We believe the next version would benefit from OccO development. We also expect that the OccO ontology effort will be a part of the Behavioral and Social Sciences Ontology (BSSO) Foundry, which has been formed to address the extensive array of social science-related information. The OccO workgroup is evaluating whether to join forces with this activity for our next stage development, in its current informal status, but the hope is that OccO activity can be elevated to a more formal, endorsed, and funded activity by one or more occupation standard organizations as this step was taken.

7. Acknowledgements

We would like to thank Gianluca Bortoletto, Jan Luts, and the ESCO team managed by the European Commission; our infrequent meetings were informative and invigorating. We would also like to thank our correspondents in Alabama, Nick Moore, head of the Governor's Education and Workforce Transformation project.

8. References

[1] "Occupation Ontology (OccO)." GitHub, Accessed 6 June 2023, https://github.com/Occupation-Ontology/OccO.

[2] "Standard Occupational Classification." U.S. Bureau of Labor Statistics, Accessed 6 June 2023, https://www.bls.gov/soc/.

[3] "About the ILO: History." International Labor Organization, Accessed 6 June 2023, https://www.ilo.org/global/about-the-ilo/history/lang--en/index.htm.

[4] "Standard Occupational Classification." UK National Statistics, Accessed 6 June 2023, https://www.ons.gov.uk/methodology/classificationsandstandards/standardoccupationalclassificationsoc.

[5] "What is ESCO?" The European Skills, Competences, Qualifications, and Occupations of the European Union, Accessed 6 June 2023, https://esco.ec.europa.eu/en/about-esco/what-esco.

[6] Smith, S., Dooley, D., and He, Yongqun. "Foundational Development of an Occupation Ontology." Proceedings of The Eighth Joint Ontology Workshops, 2022, https://ceur-ws.org/Vol3249/paper6-OSS.pdf.

[7] Smith, S., et al. "Toward an Occupation Ontology, OccO." Proceedings of the International Conference on Biomedical Ontologies, University of Michigan, MI, USA, 2022.

[8] Hicks, A., et al. "The Ontology of Medically Related Social Entities: Recent Developments." J Biomed Semant, vol. 7, 47, 2016, https://doi.org/10.1186/s13326-016-0087-8.

[9] Smith, B., et al. "The OBO Foundry: Coordinated Evolution of Ontologies to Support Biomedical Data Integration." Nat Biotechnol, vol. 25, pp. 1251-5, Nov 2007.

[10] Jackson, R., et al. "OBO Foundry in 2021: Operationalizing Open Data Principles to Evaluate Ontologies." Database (Oxford), vol. 2021, Oct 26 2021.
[11] "Internship Opportunities." Alabama Governor's Office, Accessed 6 June 2023, https://governor.alabama.gov/assets/2019/09/Governor-IveyGOEWT-Office-Internship.pdf.
[12] Menzel, Christopher. "Knowledge Representation, the World Wide Web, and the Evolution of Logic." Synthese, vol. 182, no. 2, 2011, pp. 269-295.
[13] Otte, J. Neil, et al. "BFO: Basic Formal Ontology." Applied ontology, vol. 17, no. 1, 2022, pp. 17-43.
[14] Kulvatunyou, B., et al. "The Industrial Ontologies Foundry (IOF) Core Ontology." Formal Ontologies Meet Industry (FOMI), Tarbes, FR, 2022, https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=935068.
[15] Rudnicki, Ron. "Modeling Information with the Common Core Ontologies." 2016, GitHub, Accessed 6 June 2023, https://github.com/CommonCoreOntology.
[16] "ISO/IEC 21838-2:2021." International Organization for Standardization, Accessed 6 June 2023, https://www.iso.org/standard/74572.html.
[17] Arp, R., Smith, B., and Spear, A. D. Building Ontologies with Basic Formal Ontology. MIT Press: Cambridge, MA, USA, 2015.
[18] Arp, R., Smith, B. "Function, Role, and Disposition in Basic Formal Ontology." Nat Prec, 2008, https://doi.org/10.1038/npre.2008.1941.1.
[19] "Concepts and definitions." International Labour Organization, Accessed 6 June 2023, https://ilostat.ilo.org/resources/concepts-and-definitions/classificationoccupation/#:~:text=and%20ISCO%2D88.
[20] "Labor Force Statistics from the Current Population Survey." U.S. Bureau of Labor Statistics, Accessed 6 June 2023, www.bls.gov/cps/cpsaat25.htm.
[21] American Occupational Therapy Association. "Occupational Therapy Practice Framework: Domain and Process (3rd ed.)." American Journal of Occupational Therapy, 68(Suppl. 1), S1–S48, 2014, https://www.aota.org/~/media/Corporate/Files/Advocacy/Federal/coding/OT-Practice-Framework-Table-1-Occupations.pdf.
[22] "Taxonomy Browser." National Center for Biotechnology Information, Accessed 6 June 2023, https://www.ncbi.nlm.nih.gov/taxonomy.
[23] Hastings, J., et al. "Representing Mental Functioning: Ontologies for Mental Health and Disease." Proceedings of the Third International Conference on Biomedical Ontology, 2012.
[24] Toyoshima, F., Barton, A., Grenier, O. "Foundations for an Ontology of Belief, Desire and Intention." 11th International Conference on Formal Ontology in Information Systems (FOIS 2020), Sep 2021, Bolzano, Italy, pp.140-154.
[25] Kasmier, D., Limbaugh, D., Smith, B. "Foundation for a Realist Ontology of Cognitive Processes." Proceedings of the International Conference on Biomedical Ontology, University at Buffalo, NY, 2019.
[26] de Boer, M. H. T., et al. "Dynamic Ontology Matching Challenge." TNO - Netherlands Organisation for Applied Scientific Research; Universiteit Leiden, Accessed 6 June 2023, https://www.example.com/dynamic-ontology-matching-challenge.
[27] Matentzoglu, N., et al. "A Simple Standard for Sharing Ontological Mappings (SSSOM)." Database, vol. 2022, 2022, baac035, https://doi.org/10.1093/database/baac035.
[28] "EBSCO Education Solutions and Services." EBSCO Information Services, 2023, Accessed 6 June 2023, https://solutions.ebscoed.com/.
[29] "Building Blocks Model." CareerOneStop, U.S. Department of Labor, Employment and Training Administration, Accessed 6 June 2023, https://www.careeronestop.org/CompetencyModel/CompetencyModels/building-blocks-model.aspx.
[30] "Competencies & Credentials." National Network of Business and Industry Associations, Corporation for a Skilled Workforce, Accessed 6 June 2023, https://skilledwork.org/what-we-do/competenciescredentials/.

[31] "Federal, State, and Regional." Credential Engine, Accessed 6 June 2023, https://credentialengine.org/credential-enginepartners/federal-state-and-regional/.
[32] "Credential Transparency Description Language (CTDL) Handbook." Credential Engine, Accessed 6 June 2023, https://credreg.net/ctdl/handbook#credential.
[33] "Alabama." Credential Engine, Accessed 6 June 2023, https://credentialengine.org/partners/alabama/.
[34] Smith, B., Loddo, O. G., Lorini, G. "On Credentials." Journal of Social Ontology, 6(1):47-67, 2020.