Integrating a Survey Ontology into an Upper Level Ontology
Using the Data Collection Ontology (DCO) as the Basis for a Survey Ontology

Joel Cummings and Deborah Stacey
School of Computer Science, University of Guelph, Guelph, Ontario, Canada

Keywords: Mid Level Ontology, Data Collection Ontology (DCO), BFO, Data Collection, OBO Foundry, Foundational Ontology, Upper Level Ontology, Domain Ontology.

Abstract: Capturing data in a step-by-step manner is generally completed using surveys that maintain some flow between questions to capture data from a large number of respondents in a consistent manner. In other words capturing data using surveys is a form of data collection that imposes a specific process to collect data. In this paper we present the benefit of utilizing the mid-level Data Collection Ontology (DCO) to construct a survey ontology that is domain independent and compare to an existing Survey Ontology (Fox M.S., 2016) implementation.

1 INTRODUCTION

Capturing data in a step-by-step manner is generally completed using surveys that maintain some flow between questions to capture data from a large number of respondents in a consistent manner. In other words capturing data using surveys is a form of data collection that imposes a specific process that is used to capture data. This process enforces a particular flow for the collection of data and provides a specific format for each step that the captured data must meet in order to be consistent. In the context of an ontology this is an interesting problem in that ontologies are capable of modelling such a solution but there are subtle complexities we want to examine in developing a domain agnostic survey ontology. For which we define domain agnostic as not containing any terms that would be considered as specific to a particular domain or domains. The notion of an ontology to model a survey is not a unique idea or development (Fox M.S., 2016) however, such designs face the issue of being applicable to any domain or survey purpose. These ontologies are typically created at the domain level and as a result include domain specific terms that does not allow for reuse. Reuse is one of the most important hallmarks in ontology design (Gruber, 1995) and thus it is arguably more important for surveys since questions and response formats can be reused between surveys and in the analysis of data.

Recently an ontology designed for surveys at a domain agnostic level was developed (Fox M.S., 2016) which seeks to create a design that capture any survey regardless of domain. This ontology is a custom design and itself does not reuse any existing ontologies (Fox M.S., 2016). This ontology will serve as the baseline in our study of developing domain agnostic, and therefore, reusable survey ontologies. Our question or problem will be will does utilizing a high level ontology design improve upon this existing Survey Ontology?

In this paper we present the benefit of utilizing the mid-level Data Collection Ontology (DCO) to construct a Survey Ontology that is domain independent. We start by discussing the DCO as a base ontology, then introduce our concept of a survey, and finally we introduce the Survey Ontology as our comparison point presenting our experiment and results.

2 BACKGROUND

The Data Collection Ontology (DCO) (Cummings and Stacey, 2017) is based off of the Basic Formal Ontology (BFO). The BFO was chosen for its simplicity and relative popularity as well as its focus on being domain agnostic (Cummings and Stacey, 2017). The DCO serves to define data collection terminology and relations within the BFO hierarchy to establish a starting point for ontology developers that seek to collect data of some kind. The DCO extends key classes such as processes and continuants to provide necessary classes that one would use in data collection. In addition it includes a philosophy for data collection to make use of First Order Logic and reasoning to es-

Cummings J. and Stacey D.
Integrating a Survey Ontology into an Upper Level Ontology - Using the Data Collection Ontology (DCO) as the Basis for a Survey Ontology.
DOI: 10.5220/0006582703100319
In *Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (KEOD 2017)*, pages 310-319
ISBN: 978-989-758-272-1

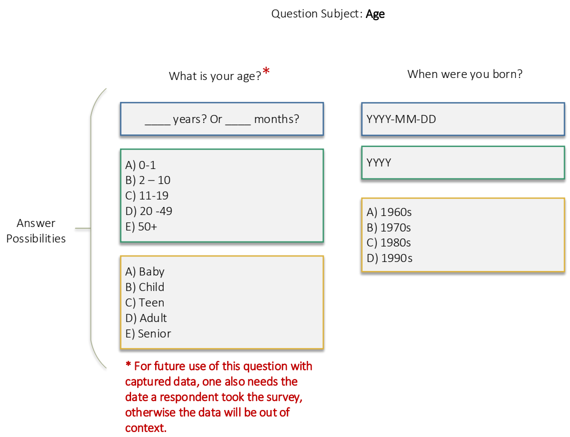

Figure 1: An example of a single survey question.

tablish relationships and categories for captured instances. It does this using best guesses or anecdotal data to determine if the expected world view is seen with the collected data. In the DCO these anecdotal data or guesses are referred to as *classifiers* that impose equivalence relations for expected values or ranges on your captured instances such that they are grouped by the reasoner. In addition assigned to captured instances are *has expected type* relations that map to the expected type or category based on the anecdotal evidence is compared with the reasoner assigned classifier. When the *has expected type* and the *classifier* are out of sync an error exists with the data or a world view issue exists. This affords the ability to detect issues early on during data collection. In addition to its philosophy on reasoning the DCO is an ontology that is focused on maintaining a small size and not including domain specific classes or terminology that may conflict with domain ontologies that utilize it. These properties are what make the DCO more suitable as a base for data collection when compared to the BFO alone which focuses on the organization of terms at a very high level into their place in space time which can be overwhelming when constructing an ontology for surveys.

In the context of our problem we must introduce a survey representation in the form of an ontology and then discuss the Survey Ontology (Fox M.S., 2016) in depth as that will be the basis for our experiment which is to construct a domain agnostic design of the Survey Ontology using the DCO. The next sections will introduce the background of the Survey Ontology problem and our solution.

2.1 Examining the Survey

Breaking a survey down as shown in Figure 1 illustrates the components of a survey, each of which can be reused across surveys. From a high level, surveys are a grouping of components that are executed in a particular order which can be linked or taken from other surveys. Additionally, the linkings between these components, allow for traceability across questions to determine the origin, the uses cases, and the popularity of each the components. Furthermore, through the use of branching and conditional links one can produce the flow of a survey that will result in the capturing of respondent results as ontological instances. It is this structure that we seek to model when creating a survey ontology.

2.1.1 The Survey Ontology

The goal of the Survey Ontology is to develop a standard for a survey structure's representation and allow for communication over the Semantic Web (Fox M.S., 2016). The Survey Ontology has the goal to enable reuse of survey data for other purposes in future works (Fox M.S., 2016). This purpose of the Survey Ontology overlaps well with our problem and design intentions of the DCO but closer to the domain level since it is focused on survey which we consider as a form of data collection but not data collection as a whole. Finally, the Survey Ontology is recent but still under construction making it a reasonable point of comparison as it is in similar infancy to that of the DCO.

The Survey Ontology is focused on creating a generic and reusable framework for producing surveys and capturing survey answers. The work is based off of early work that used XML to create a general format for questions, answers, and survey logic (Fox M.S., 2016). The main issue noted is that one often wants to compare questions as well as responses to questions in surveys. More generally, analysis is often missed out in existing implementations (Fox M.S., 2016). The Survey Ontology focuses on creating a representation for the Semantic Web that allows for questions, answers, and responses to be linked and reused.

The Survey Ontology contains a number of classes starting with **Survey**. The Survey class defines a recursive structure of **Survey_Part** instances that allow surveys to be sectioned and broken down into smaller chunks. Next it defines **Question** which in turn defines the basic format including text, an identifier, and a sequence. Question is then subclassed into several types based on usecase. **Survey_Response** and **Survey_Answer** are similar to Question in that they define required components that focus on establishing links between user responses, the Question they are related to, as well as start times and end times so surveyors can track when each component was used (Fox M.S., 2016). The last key component is the **Sub-**

ject which serves to link questions and responses to a general subject which serves two purposes. Firstly, it allows for more semantic understanding of what a Question or Response seeks to capture and secondly it allows the use of more general relations since a relation does not need to focus on describing the Subject it collects data on (Fox M.S., 2016). An example would be *surveytaker hasHadHeartAttack yes* versus *surveytaker hasAilment HeartAttack* and assigning the subject Heart Attack to the question being asked i.e. *surveyquestion hasSubject HeartAttack*.

The structure of the ontology makes it clear through its restrictions and relations that Answers, Questions, Response Formats, and Surveys are linked so that one can examine each component individually while maintaining a connection to relevant components so one can move between related classes. In addition the Survey Ontology imposes some default values (Fox M.S., 2016) and ranges for question subtypes to enable faster construction. It also provides the ability to produce some standard survey types (Fox M.S., 2016).

A specific example of a survey question can be seen in Figure 1 where we present questions for the subject of age. We note that there are multiple ways of asking the question of age and depending on the way we ask the information we require to analyze and make inferences changes. The first question we pose is *What is your age?* where we note that during analysis one would also need to know when a respondent answered the question in order for current age to be determined. The second question we pose is *When were you born?* where we can calculate age based on the current date. However, both questions have the same subject of age and that is one important component of the Survey Ontology. The second area where an ontology is important is with the different formats that respondents can provide their responses in. In other words, they are guided by the answer format of which we provided 3 examples for the age question. For comparative and analytical purposes the format that is used is important as not all responses can be converted depending on their format. For example if age is provided in ranges it cannot be converted into specific integer values, however, this conversion can be done the opposite way. A survey ontology provides benefit through linking formats to responses, questions, and subjects so that one is able to make decisions about how to analyze their results.

This generic structure of capturing survey data and survey components is an example of where the DCO could be used to speed up ontology development and provide structure and components which is the motivation for demonstrating this design through integrating the Survey Ontology into the DCO and secondly creating a DCO variant of the Survey Ontology with the same goal in mind.

3 EXPERIMENT

Through using DCO as the base for our problem we can view our experiment as the application of the DCO in a real world problem to build a domain agnostic survey ontology. To do so we will establish a comparison using the Survey Ontology (Fox M.S., 2016) as the baseline of a good ontology and we hypothesize that through using an upper level ontology, in this case the DCO we will achieve a better result through our evaluation methods, see Section 3.1.

Our experiment involves the working version of the Survey Ontology (Fox M.S., 2016) (Fox, 2016) in reconstructing the ontology as well as taking the premise of the Survey Ontology and constructing a survey ontology from the DCO point of view. We acknowledge that the Survey ontology is not complete but is recent (Fox, 2016) and provides a good example for a use case of the DCO. We will take into consideration its recency and incompleteness when doing our full evaluation.

The three ontologies to be compared will be defined below with names which will be used throughout the remaining chapters. The rest of the chapter will be dedicated to the complete design and terms of each ontology. Due to the size of the ontologies they are not presented here but can be downloaded from: http://dx.doi.org/10.5683/SP/PFHFUQ

- **Survey Ontology** - This is the original unaltered Survey Ontology as this writing (Fox, 2016). This is main point of comparison as it establishes the premise, design, and use cases of a Survey Ontology.
- **Integrated Survey Ontology** - This the first of the DCO based Survey Ontologies which integrates all the terms into the DCO reusing equivalent DCO classes and relations where applicable. The Integrated Survey Ontology seeks to establish that it is possible to model any data collection ontology within the DCO.
- **DCO Survey Ontology** - This is the DCO designed and developed Survey Ontology, it does not reuse all classes and terms, instead it takes the premise of what a Survey Ontology should be based on (Fox M.S., 2016) and implements the design from the DCO philosophy of data collection. This would be the ontology that one developed

given the DCO and the requirements of a Survey Ontology.

3.1 Evaluation Method

Traditional methods of evaluation are important for comparison to existing ontology designs and identify key components in the ontology community about what makes a reasonable design. There are several different methods for evaluation (Bandeira et al., 2016) with somewhat competing views. Therefore we are choosing an aggregate method, the FOCA method (Bandeira et al., 2016) which combines ideas of several existing methods to derive evaluative measures for ontology design. The FOCA method introduces several measures as well as a framework for evaluating the measures making it possible for an ontology developer of any experience level to evaluate their design. Secondly, it allows ontologies to be compared regardless of their level or purpose making it suitable to evaluate the DCO against even domain level ontologies as well as DCO derived ontologies with domain level ontologies.

3.2 The FOCA Methodology

FOCA has several parts to the methodology that includes determining the type of ontology, a questionnaire to evaluate components, a framework to follow based on ontology type, and finally a statistical model that calculates the quality of ontologies. Most of these components are fairly common to the Applied Ontology community but previous methods lack the questionnaire and framework which is why focus will be placed on these components (Bandeira et al., 2016). The FOCA method breaks down its questionnaire into relevant roles that separate measures based on Ontological Commitments, Intelligent Reasoning, Efficient Computation, Human Expression, and Substitution (Bandeira et al., 2016). Each of these goals evaluates a particular part of the ontology and has questions associated with them which will be defined. Additionally, it groups questions based on the following metrics: Completeness, Adaptability, Conciseness, Consistency, Computational Efficiency, and Clarity (Bandeira et al., 2016). These goals and metrics are found through many Applied Ontology communities (Bandeira et al., 2016) which is why we feel this method is a strong indicator of an ontologies quality.

To ensure readers who are unfamiliar with the method have an understanding of what it evaluates we will examine each question and how it will be evaluated. Additionally, there are a few cases where we have altered the evaluation which will be explained on a question by question basis. FOCA divides questions into major goals which is what we will use to divide the questions into groups.

Table 1: FOCA Goal 1.

ID	Question	Description/Evaluation Criteria
Q1	Were the competency questions defined?	If the ontology does not have competency questions defined, assign 0. If they do exist there are three subquestions: Does the document define the objective of the ontology? Second: Does the documentation define stakeholders? Third: Does the document define use cases? Each sub question receives a grade of one of: 0, 25, 50, 75, 100. The overall grade is the mean of the 3 sub questions.
Q2	Were the competency questions answered?	This grade is 0 if competency questions were not defined. Otherwise determine if the ontology has satisfied the competencies. Grades: 0, 25, 50, 75, 100.
Q3	Did the ontology reuse other ontologies?	If the ontology reuses other ontology(s) assign 100, 0 otherwise.

3.2.1 FOCA Goals and Questions Defined

Goal 1 centres around the ontology design ensuring that one has competency questions defined and answered, and that the ontology has some form of reuse. See Table 1 for full descriptions.

Goal 2 centres around the ontologies design and structure, specifically ensuring that a goal is met based on its high level goal (domain or upper level ontology) and whether it is met with its class definitions. For example, it ensures that if an ontology claims to be upper level that it includes an inheritance structure before it defines domain particular terms. Goal 2 also checks the domain of the ontology, ensuring that it does not define terms outside of the domain it claims to represent. See Table 2 for full descriptions.

Goal 3 centres around looking through the ontology and determining if it has contradictions or invalid reuse of terms. Contradictions are issues where the properties on relations (functional, transitive, reflexive etc.) are not applicable to the term in the ontol-

Table 2: FOCA Goal 2.

ID	Question	Description/Evaluation Criteria
Q4	Did the ontology impose a minimal ontological commitment?	Answer this if it is type 2 (domain ontology). Ensure that the ontology does not define high level abstractions and content that is not specific to a domain. i.e. a Facebook ontology does not need to define a computer system. Grades are 0, 25, 50, 75, 100.
Q5	Did the ontology impose a maximal ontological commitment?	Answer only if the ontology is type 1 (high level ontology). Ensure the ontology defines high level abstractions such that domain level elements have more general parents. Grades are 0, 25, 50, 75, 100
Q6	Are the ontology properties coherent with the domain?	Determine if the ontology contains elements that are not coherent with te domain. For example, a car ontology should not contain lion. Grades 0, 25, 50, 75, 100.

ogy. Redundant axioms are cases where reuse should not occur as a term with the same name has two different meanings. See Table 3 for full definitions.

Goal 4 centres around reasoning and reasoner performance and this is where our first modification is made. Question 10 (Q10) (Bandeira et al., 2016) is based on the speed of ontological reasoning where it uses the verification of stopping being a grade of 0, any delay being a grade of 25, 50, or 75, and running quickly being 100, see Table 4. This does not allow for easy comparison as *quick* is relative. Furthermore, it does not consider the number of ontology components (classes, relations, properties etc) or the expected number of individuals which is important where equivalence relations are used. For speed our grade will be boolean, 0 for reasoner failure, or 100 for successful reasoning.

Goal 5 (see Table 5) centres around documentation that is internal to the ontology as well as ensuring the modelled ontology is the same as what is described in the design documentation. The evaluation is straightforward but we make modification for clarification purposes. Question 12 (Q12) (Bandeira et al., 2016) is based on definitions and descriptions in the ontology and scores based on language used with deductions made for using a language other than English, we are going to change this justification so that as long as all terms are defined in English full score is assigned and we will not deduct points for ontologies that include definitions in other languages even if they do not cover all terms. The explanation in (Bandeira et al., 2016) gives the impression that one should deduct for using other languages, this will not be case in our evaluation.

Table 3: FOCA Goal 3.

ID	Question	Description/Evaluation Criteria
Q7	Are there contradictory axioms?	Check if the classes and relations contradict the domain for example if the hasSocialSecurityNumber is not functional this would be a problem because a person can only have one. Based on the number of contradictions give grades between 0, 25, 50, 75 or 100 if there are none.
Q8	Are there redundant axioms?	Determine there are classes or relations that model the same thing with the same meaning (ie.e using mouse for computer hardware and the animal). If there are many redundancies grade 0, if there are some assign one of 25, 50, 75 and 100 if there are no redundancies.

Table 4: FOCA Goal 4 (Bandeira et al., 2016).

ID	Question	Description/Evaluation Criteria
Q9	Does the reasoner bring modelling errors?	Check if the reasoner returns errors. If there are many errors or the reasoner stops assign 0, if there are some errors assign 25, 50, 75 and 100 if there are no errors.
Q10	Does the reasoner perform quickly?	Determine if the reasoner runs quickly. If the reasoner stops assign 0, if there is any delay assign one of 25, 50, 75 or 100 if it runs quickly.

4 RESULTS

Our comparison for the Survey Ontology variants will be from the structural perspective since any docu-

Table 5: FOCA Goal 5.

ID	Question	Description/Evaluation Criteria
Q11	Is the documentation consistent with the modelling?	Determine if there are definitions in the ontology. If there are none assign 0. Check that each class and relation has a definition and it is to the same detail as the document. Secondly, determine if the documentation explains each term and justifies it. For each sub question assign 25, 50, 75, or 100. Calculate the mean of the two sub questions.
Q12	Were the concepts well written	Determine if the classes or relations are written in an understandable and correct form (according to English or another language). If the ontology is difficult to understand or full of poorly written terms assign 0. If there are a mix of languages, assign one of 25, 50, 75. If the ontology is well written and one language as used assign 100.
Q13	Are there annotations in the ontology bringing the concepts definitions?	In this question check existing annotations being the definitions of the modelled concepts. If there are no annotations assign 0. If there are some annotations assign 25, 50, or 75. If all classes have annotations have annotations, assign 100.

mentation evaluated in the FOCA methodology is not something that DCO will provide benefit to. Additionally, this is something that can easily be changed by the Survey Ontology developers through releasing an additional document with competency questions and updating the ontology file. Furthermore, it is the structural differences and reuse differences that form the advantage of DCO. These differences are reflected in the scoring with improvement seen from the Survey Ontology to the Integrated Survey Ontology and further improvement seen from the Integrated Survey Ontology to the DCO Survey Ontology. These results can be seen in Tables 7, 8, 9, 10, and 11 at the end of this paper.

The major advantages are seen in the reuse question (Q3) and again in Q5 where the DCO's reuse avoids redefining common terms and relations while providing a structure for domain specific terms to have parents. Furthermore, there are no terms in the

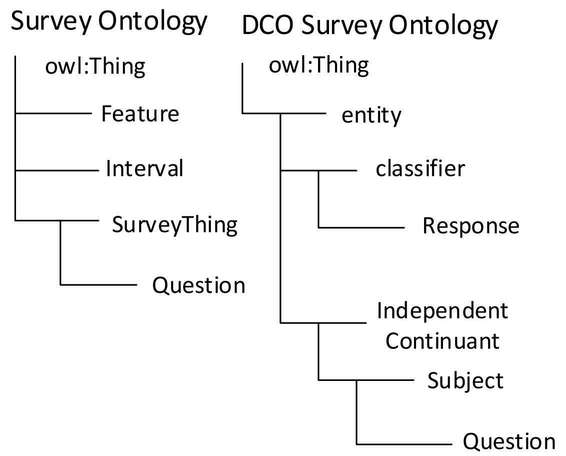

Figure 2: An example of the Structural Differences between the Survey Ontology and the DCO Survey Ontology.

DCO based survey ontologies that have owl:Thing as a parent (see Figure 2). Classes always have some parent either through the BFO or through the DCO's own terms. This creates familiar hierarchies allowing developers less familiar with the domain to understand where terms exist in the larger world view.

The FOCA scores for documentation or competency questions are not an area where the DCO provides benefit but it did avoid issues regarding duplicate object and data properties since existing terms were reused instead of being defined helping to avoid potential error. Results for Q1 and Q2 (Table 7) presented no major differences due to the fact that all documentation was based on the Survey Ontology so any issues regarding competency questions were carried through. The major difference is seen with Q3 where the DCO based variants see improvement in score due to their reuse of a higher level ontology, the DCO which in turn reuses the BFO. Goal 2 (Table 8) presents some notable changes beginning with Q5 where improvements are seen due to the fact that DCO includes hierarchies for terms which FOCA states is important for non-application ontologies or Type 1 ontologies (Bandeira et al., 2016). Since the goal of each of these ontologies is to establish a high level domain that is not application specific they should define terms in a hierarchy to place them within the larger world. This is where major advantages start to be seen with DCO since it allows people new to the survey domain to place particular terms and those familiar with BFO will have an even easier time with the basic structure. There are no differences for axioms or redundancies therefore Goal 3 does not present anything significant between ontology implementations (Table 9). Goal 4 (Table 10) presented only one improvement over the Survey Ontology since it defines a relation that is used as both an object relation and a data relation. However due to equivalences in relation purpose these relations were not migrated to the DCO variants which provided benefit in scoring as the error was no longer present. Goal 5 (Table 11) is based on components where again we see little difference with the only being no domain specific content in the DCO Survey Ontology.

With success demonstrated using the traditional FOCA method we consider the structural numbers of each ontology and contrast them with our criteria for choosing an ontology which will be outlined below. This is the criteria that was used when choosing a foundational ontology to base the DCO on.

The first criteria we want to define is based on the number of terms and relations in the ontology, where we prefer to have fewer of each for two main reasons. Firstly, upper level ontologies are meant to be derived into a domain level ontology and thus will have more terms and relations added over time and large ontologies introduce performance penalties potentially resulting in an ontology that is intractable for a reasoner (Horrocks, 2005). Secondly, in terms of understandability, the fewer terms a person must know to use an ontology the easier it is to get started. Additionally, it will reduce reliance on documentation and expert knowledge making it easier to design and organize derived ontologies. Furthermore large ontologies may deter usage of the ontology altogether.

The second criteria we care about is usage and popularity. Popularity of an upper level ontology is important when considering its purpose for unifying ontologies (Herre, 2010). We want to look at what people are using to see what is working and how many domains are being captured by the upper level ontology. If only one domain is using a particular ontology it is possible that it has not met the needs of others. Additionally greater popularity increases the likelihood that ontology developers will have experience with the ontology.

Finally, we move on to a more formal definition for upper level ontologies which is used for the purpose of ensuring that the base is kept generic, again to satisfy our definition. Thus we say an upper level ontology must be free from any domain specific terms or relations. We are not interested in ontologies that take the role of defining thousands of terms to satisfy a large number of domains since it is unlikely such an ontology could satisfy each domain realistically.

Starting with the first criteria we compared ontology sizes. The sizes are summarized in Table 6. As one would expect the class count is larger in both versions of the DCO based Survey Ontology variants, however, what is noted is that in both derived forms the number of relations were dramatically reduced. Therefore when looking at both classes and relations the size difference is insignificant. This means any implementation is not too large by our standards and this is because when the world of ontologies is considered as a whole, ontologies with thousands of classes are not uncommon making a difference in size of 70 relatively small and unlikely to deter ontology developers.

When looking at relations one can see the dramatic effect reuse makes. For object relations the Derived Survey Ontology added only 25 of the 41 relations defined in the Survey Ontology. The other relations had equivalents that already existed in the DCO. Similarly, the Derived Survey Ontology adds only 21 of the 26 data properties again using equivalents declared in the DCO. Therefore in terms of growth the DCO maintains a small number of relations through reuse of major data collection requirements such as time, units, and data structures.

Our second criteria involves looking at domain specific content. We note that the Survey Ontology includes domain specific content as of this writing but we acknowledge that it is not tied into the structure of the ontology and could be removed without major refactoring. Therefore, we will not dive deeply into this criteria as it is not something the DCO will affect.

Lastly, is usage and popularity. Both ontologies are in their infancy, however, the DCO's base provides familiarity since any OBO user or developer is already familiar with our design to an extent. An OBO user would already know of the basic design, classes, and how the ontology is structured since it based on the BFO. To use the DCO they would only need to learn about Subjects, Classifiers, and Datums. We argue this provides significant benefit in terms of usage and popularity since the core of the ontology is well established. The Survey Ontology does not have this advantage since it uses no base ontology and defines terms exclusively in its own hierarchy. This would mean developers would need to completely learn the structure before use.

4.1 FOCA Evaluation Table Notation Defined

Here are some brief definitions for table components so one can more easily navigate the results. Major differences between versions are **bolded** to show changes across the variants. All Justifications are presented in the same order as the score columns meaning the first justification for a question refers to the Survey Ontology while the last justification refers to the DCO Survey Ontology. Any dashes present means refer to the row above for explanation. See below for more terminology.

SO refers to the Survey Ontology's score, **ISO** refers to the Integrated Survey Ontology's score, **DSO** refers to the DCO Survey Ontology's score. For Question ids when a dash proceeds the number it refers to the subsection question in FOCA for example, Q1-2 refers the second subquestion of question 1. N/A manes the question was not scored.

4.2 Evaluating Ontology Hypotheses

We also define several hypotheses about DCO variants that are presented below. These are properties an ontology developed using the DCO will present with.

The first hypothesis states that we expect overlap in terms meaning that the DCO should contain terms and relations that can be used by the Survey Ontology variants. Overlap is found through integration of the Survey as a Process which uses the control flow for repeating, and branching directly. Similarly, Questions, Answers, and Person are considered Subjects as they are what is studied in Surveys. Relations for time were used directly in place of those defined in the Survey Ontology itself demonstrating direct overlap. Therefore we can say that our hypothesis is true that there is domain overlap with the DCO components.

The second hypothesis states that we expect terms to be at a lower level. We note that none of the terms placed into the DCO derived Survey Ontology were placed at a level in the hierarchy that was above any existing DCO term meaning there were no terms that were of greater generality than those defined in the DCO. Therefore our hypothesis is true.

Lastly, we are concerned with coverage meaning we do not want terms defined outside of the DCO's hierarchy and in the construction of the Survey Ontology derivatives it was the case that all terms fit within the DCO hierarchy meaning nothing was subclassed as owl:Thing. Therefore this hypothesis was true.

4.3 Survey Ontology Evaluation Conclusions

The Survey Ontology implementations meet our hypotheses. Therefore, with the greater scores of the FOCA evaluation and no major detriment to our criteria we can determine that DCO provided some notable improvements with reuse especially though relations which we were able to reduce as well as through the hierarchies and organizational benefits to the classes of the Survey Ontology. Within DCO we used Subjects, and Processes to describe Survey Ontology terms so they could be understood and motivated at a higher level and as a result reused definitions and relations imposed on those classes to reduce the amount of restrictions that needed to be created.

Table 6: Survey Ontology Sizes Compared.

Measure	SO	ISO	DSO	DCO Base
Class Count	37	102	107	73
Object Relation Count	41	50	34	25
Data Relation Count	26	33	16	12

Table 7: Goal 1 Questions and Justifications for Survey Ontologies.

Q	SO	ISO	DSO	Justification
Q1	0	0	0	No competency questions are defined in the Survey Ontology.
Q1 Subquestions				
Q1-1	100	100	100	The document does contain the ontology objective which is to "represent the logic of the Survey, including contingent questions and repeated sections" (Bandeira et al., 2016) .
Q1-2	100	100	100	Yes, the scenarios are cases when you would normally use an XML or other Survey representation.
Q1-3	100	100	100	Yes, the stakeholders are defined throughout the document making it obvious who would be interested in such an ontology.
Q2	0	0	0	There were no competency questions in the ontology or in the document.
Q3	0	**100**	**100**	No the ontology does not reuse other ontologies. The ontology does however, reuse the schema:Person class.
				Yes, the DCO and the BFO are reused. -

Table 8: Goal 2 Questions and Justifications for Survey Ontologies.

Q	SO	ISO	DSO	Justification
Q4	N/A	N/A	N/A	Not Applicable, all ontologies are Type 1
Q5	25	**75**	**75**	The Survey Ontology provides very little in the way of abstraction by going from owl:Thing to SurveyThing. It does not define the time or space a survey is represented in. Additionally, the ontology incorporates terms such as schema:Person as owl:Thing as well as Population and Interval. The Integrated Survey Ontology does define hierarchies for each of its objects with Surveys represented as Processes and SurveyParts being Process Parts and all objects organized according to their place in time. No objects are derived directly from owl:Thing. Similar to the Integrated Survey Ontology, the DCO Survey Ontology also defines a hierarchy for survey terms.
Q6	0	0	**100**	The ontology does include elements that are not coherent with the domain such as Disorder and Medication which are not elements of all surveys. - The DCO Survey Ontology only includes relevant terms for the Survey domain.

5 CONCLUSIONS

We have presented two ontologies derived from the Survey Ontology each with benefits over the other iteration. Based on the design of using an upper level the FOCA evaluation has shown benefit in reuse and in defining more general hierarchies to help ontology developers place domain specific terms within the hierarchy and providing a place for non-survey terms within the ontology.

Lastly, we presented a working example of the DCO in the case of integrating an existing ontology showing that its possible to take an existing ontology and map it into the DCO if one is part way through development and still see benefit and secondly that the DCO is capable of integrating a separate philosophy into its design and that it is of a high enough level to not encumber such an integration.

Table 9: Goal 3 Questions and Justifications for Survey Ontologies.

Q	SO	ISO	DSO	Justification
Q7	100	100	100	There are no contradictory axioms.
Q8	100	100	100	There are no redundant terms in the ontology.

Table 10: Goal 4 Questions and Justifications for Survey Ontologies.

Q	SO	ISO	DSO	Justification
Q9	75	**100**	**100**	There is one error with a redeclaration of *targetProperty* as both an object property and a data property. There are no errors in the ISO. -
Q10	100	100	100	Yes the reasoner completed and did so at an average of just under 500ms based on our testing.

6 FUTURE WORK

A useful addition to any Survey Ontology would be an Analysis Ontology that focuses on tools that are used to analyze the results of captured survey data. This might be integrated into the DCO or the DCO Survey Ontology itself whichever way the developer sees fit. The ontology should focus on utilizing and mapping data between survey instances so that reuse can be conducted using similar surveys from different studies to pool data instead of continually recapturing the same or similar data.

A major goal of this ontology would be to create some form of mapping of compatible answer formats. As an example using our questions in Figure 1 it would be possible to map the age in years or months to the ranges such that 6 months would map to 0-1

Table 11: Goal 5 Questions and Justifications for Survey Ontologies.

Q	SO	ISO	DSO	Justification
Q11	50	50	50	The documentation explains each term very well along with examples how it will work which results in a Score of 100. However, the definitions in the document are not the same as those in the ontology so we award a Score of 50.
Q11 Subquestions				
Q11-01	0	0	0	No, the definitions in the ontology are not the same as some detail is missing and not all have annotations in the modelling.
Q11-02	100	100	100	Yes, the terms and the design of the ontology are well documented and explained.
Q12	100	100	100	The annotations are well written and easy to understand with no errors.
Q13	50	50	**100**	Not all elements were annotated. Many of the object and data relations were not annotated. Classes that were not annotated include: Feature, Interval, SurveyThing, Mental Disorder, Physical Disorder, Medication, Confidence QT, ExcuseQT, FrequencyQT, satisfactionQT, YearQT, YesNoQT, TextQEQ, Experience Statement, Intrinsic Statement, and Non Repeated Survey Part.
				All Elements were transferred over along with annotations, therefore the missing annotations were included.
				All elements were annotated with descriptions and purposes.

years and 5 years would map to 2-10 years etc. Allowing this type of mapping to be inferred by a reasoner would allow for users to compare two similar surveys and determine if and for which responses can the data can be pooled and reused for a further study.

REFERENCES

Bandeira, J., Bittencourt, I. I., Espinheira, P., and Isotani, S. (2016). Foca: A methodology for ontology evaluation. *arXiv preprint arXiv:1612.03353*.

Cummings, J. and Stacey, D. (2017). The mid level data collection ontology(dco). *Proceedings of KEOD 2017*.

Fox, M. (2016). Enterprise integration laboratory eil. http://www.eil.utoronto.ca/. (Accessed on 04/23/2017).

Fox M.S., K. M. (2016). An ontology for surveys. *Proceedings of the Association for Survey Computing*, 7.

Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *Int. J. Hum.-Comput. Stud.*, 43(5-6):907–928.

Herre, H. (2010). *General Formal Ontology (GFO): A Foundational Ontology for Conceptual Modelling*, pages 297–345. Springer Netherlands, Dordrecht.

Horrocks, I. (2005). Description logics in ontology applications. In *International Conference on Automated Reasoning with Analytic Tableaux and Related Methods*, pages 2–13. Springer.