How To Teach Domain Ontology-based Knowledge Graph Construction? An Irish Experiment

Subhashis Das*, Mayukh Bagchi** and Pamela Hussey***

*ADAPT, CeIC, Dublin City University, Dublin 9, Ireland <Subhashis.das@dcu.ie>
**DISI, University of Trento, Via Sommarive, 9 I-38123 Povo (TN), Trento, Italy <Mayukh.bagchi@unitn.it>
***ADAPT, CeIC, Dublin City University, Dublin 9, Ireland <Pamela.hussey@dcu.ie>

Dr. Subhashis Das is an ELITE-S Marie-Curie postdoctoral fellow at ADAPT Centre @DCU and Centre for e-Integrated Care (CeIC). Dr. Das is a member of NSAI-HISC. Subhashis was a data scientist in the Know-Dive Group at the Department of Information Engineering and Computer Science (DISI), Università degli Studi di Trento, Italy working on the InteropEHRate, an EU-Horizon 2020 project. He obtained his Ph.D. from the ICT-Doctoral School, University of Trento, Trento, Italy. Subhashis' research interests focus on ontology, information science, and health informatics. He has working experience in the healthcare sector having worked at NHS-Scotland, School of Informatics, University of Edinburgh, UK.

Mayukh Bagchi is a PhD student under the supervision of Prof. Fausto Giunchiglia at the KnowDive Group within the Department of Information Engineering and Computer Science (DISI), University of Trento, Italy. He is also the Teaching Assistant for the Knowledge and Data Integration (KDI) postgraduate course. He has a Bachelor of Science (BSc) degree with Honours in Mathematics [from the University of Calcutta, India] and a Master of Science degree in Library and Information Science (MSLIS) [from the Indian Statistical Institute (ISI), Bangalore, India]. His research interests include Knowledge Organization, Data and Knowledge Management, Knowledge Representation and Reasoning (KRR), and Semantic Web Technologies. His other research interests include Cognitive Sciences, Computational Intelligence, Human-Computer Interaction (HCI) and mathematical approaches to intelligent information management.

Dr. Pamela Hussey is PhD in Computer Science and has also completed an MEd in Education and Training Management and an MSc in Health Informatics. From 2018-2022 she was chair of the National Standards Authority of Ireland Health Informatics Standards Consortium (NSAI HISC) and Director of the Center for eIntegrated Care (CeIC) in the Faculty of Science and Health. She is also a funded investigator with Adapt Research Center with SFI 2025. Hussey's current research activity includes acting as an advisor to different agencies on topics relating to interoperability. From 2016-2022 she was the Director of the DCU SNHS ICNP User Group, an accredited research center for the International Classification for Nursing Practice in Ireland. Funded research in the past includes a HSE public services contract for concepts standards and information modelling for the national health and social care data dictionary and an Enterprise Ireland Disruptive Technology Fund with the Insight Research Center on uptake and use of digital IoT and self management support care delivery. Through Adapt SFI she is supervisor to an Elite S research fellow through the EU Marie Curie programme. Her teaching interests include: health informatics, self care management, interoperability and use of technology enhanced learning resources in both undergraduate and postgraduate education. She is currently supervising PhD students, and was lead editor and author on the 5th edition of An Introduction to Nursing Informatics published January 2021 through the Springer Health Informatics Series.

Das, Subhashis, Mayukh Bagchi and Pamela Hussey. 2023. "How To Teach Domain Ontology-based Knowledge Graph Construction? An Irish Experiment." *Knowledge Organization* 50(3): 182-201. 50 references. DOI:10.5771/0943-7444-2023-3-182.

Abstract: Domains represent concepts which belong to specific parts of the world. The particularized meaning of words linguistically encoding such domain concepts are provided by domain specific resources. The explicit meaning of such words are increasingly captured computationally using domain-specific ontologies, which, even for the same reference domain, are most often than not semantically incompatible. As information systems that rely on domain ontologies expand, there is a growing need to not only design domain ontologies and domain ontology-grounded Knowledge Graphs (KGs) but also to align them to general standards and conventions for interoperability. This often presents an insurmountable challenge to domain experts who have to additionally learn the construction of domain ontologies and KGs. Until now, several research methodologies have been proposed by different research groups using different technical approaches and based on scenarios of different domains of application. However, no methodology has been proposed which not only facilitates designing conceptually well-founded ontologies, but is also, equally, grounded in the general pedagogical principles of knowledge organization and, thereby, flexible enough to teach, and reproduce *vis-à-vis* domain experts. The purpose of this paper is to provide such a general, pedagogically flexible semantic knowledge modelling methodology. We exemplify the methodology by examples and illustrations from a professional-level digital healthcare course, and conclude with an evaluation grounded in technological parameters as well as user experience design principles.

Received: 10/06/2023 Accepted 10/07/2023

Keywords: Knowledge Organization Education, Healthcare Informatics Education, Domain Ontology, Digital Transformation, Domain Knowledge, Knowledge Graph.

1.0 Introduction

A domain is encoded as a composition of concepts representing real world referents which belong to a part of the world (Hjørland 2017). The particular meaning of terms which stand for concepts applied to that domain are provided by domain-specific resources (e.g., dictionaries, lexical-semantic resources, etc.). For example, the word *dressing* can have different meanings. A resource about the domain of medicine (aka healthcare) would encode the meaning of *dressing* as "*a piece of material used to cover and protect a wound*", a resource about the domain of gastronomy would encode the meaning of *dressing* as "*a sauce for salads, typically one consisting of oil and vinegar with herbs or other flavourings*", whereas, in the nursing domain, *dressing* can mean either *wound dressing* or "*the activity of getting dressed*" (Ådland and Lykke 2018). The explicit meaning of words, such as the one exemplified above, needs to be *conceptually disentangled* (Bagchi and Das 2022; Bagchi and Das 2023) which, in mainstream knowledge organization and representation, is often codified using a domain-specific ontology (Guarino, Oberle and Staab 2009).

Different ontologies in the same domain (e.g., healthcare) arise due to differences in several dimensions crucial for semantic knowledge modelling, for instance, differences in domain-languages (Ghosh, S. et al., 2020; Giunchiglia and Bagchi 2022b; Staines, Hussey and Das 2022), intended usage of domain (Bagchi 2021a; Bagchi 2021b), and domain perception (Giunchiglia, F. et al., 2014; Giunchiglia and Bagchi 2022a; Giunchiglia, Bagchi and Diao 2023). Due to the preceding differences, domain ontologies represent concepts, often, in very specific and eclectic ways leading to their mutual incompatibility. As knowledge-based information systems that rely on domain ontologies expand, there is an increasing need to merge domain ontologies into, often, more general ontological representations. The combination of the above dimensions result in increasing difficulties for domain experts (e.g., different levels of healthcare professionals) towards understanding and design facilitation of conceptually well-founded ontologies (Guizzardi 2007; Biagetti 2021). More precisely, the challenge is to develop a general knowledge modelling methodology which not only facilitate designing conceptually well-founded ontologies, but is also, equally, *grounded in the general pedagogical principles of knowledge organization* (Pattuelli 2010; Hudon 2014; Szostak 2014) and, thereby, flexible enough to *teach, appropriately tweak* and *reproduce vis-à-vis* domain experts (such as healthcare professionals) in collaboration with knowledge managers (Bagchi 2022).

The current paper is positioned exactly in the preceding challenging context and discusses the highlights and experiences of teaching a pedagogically flexible knowledge modelling methodology (Hussey et al., 2021; Das and Hussey 2022) and a digital healthcare course (Hussey and Das 2021) designed for healthcare professionals focused on understanding and developing semantic models such as ontology-driven Knowledge Graphs (KGs) in collaboration with knowledge engineers. The methodology and the course was incubated at the *Centre for eIntegrated Care* (CeIC), a research centre at Dublin City University, Ireland. To that end, the paper also elucidates the step-by-step development of the proposed pedagogically flexible general methodology and the selection of academic resources and references to support its instruction and implementation *vis-à-vis* healthcare professionals. It also provides appropriate illustrations, wherever required, to demonstrate how the methodology can be taught in courses similar to the one above, both for classroom as well as online teaching.

The rest of the paper is structured as follows: Section 2.0 provides a brief literature review on existing ontology development methodologies and discusses the lack of pedagogical foundation, assessment and validation on such methodologies. Section 3.0 explains the proposed pedagogically flexible knowledge modeling approach exemplifying its use in teaching the construction of ontology-based Healthcare Knowledge Graphs (HKGs). It emphasizes the implementation of HKGs with specific reference to its teaching and education dimensions. The evaluation of the proposed methodology, and therefore, of the course has been described in Section 4.0 and a concluding discussion is provided in Section 5.0.

2.0 Literature Review

The scientific literature on ontology and KG development methodologies is overloaded, with several researchers proposing their own theories and theory-backed approaches and processes. We briefly discuss some of them below.

IDEF5 Capture Method: IDEF5 project, one of the first applied ontology modelling projects, provided a set of guidelines to knowledge engineers, analysts, and researchers to develop and manage ontologies efficiently (Peraketh et al. 1994).

Uschold And King's Methodology: This methodology (Uschold and King 1995) has been proposed based on the experience of developing the Enterprise Ontology. This is the first technique of its kind to provide guidelines for developing ontologies, which are: 1) Identify the purpose and intended uses for the proposed ontology. 2) Build the ontology itself, which is again divided into three sub steps, which are: i) identification of the key concepts and relationship in the given domain and identification of their terminology; ii) the coding phase deals with the explicit representation of the knowledge acquired in the previous steps, and iii) the integration process analyzes the question of how to reuse existing ontologies. 3) The overall third phase is evaluation, i.e., to make a technical judgement of the ontologies. It can be checked against requirements specification, competency questions, or the real world (Gómez-Pérez et al. 1995), and, 4) Finally, the documentation for the whole ontology process, with, possibly, a write-up of all guidelines followed and justification in case of differences.

Gruninger And Fox's Methodology: Gruninger and Fox (1994) proposed a methodology using first order logic which is inspired from the development of knowledge-based systems. This design has been suggested as TOVE (TOronto Virtual Enterprise) project ontology within the domain of business processes and activities modelling. The steps in the methodology are: 1) Describing the motivating scenarios. 2) Formulation of informal competency questions, to set the scope of the planned ontology. 3) Formulation of formal competency questions, which specify the terminology with definition and constraints. 4) Specification of axioms and definition within a formal language, and 5) Finally, specifying the conditions under which the solutions to the questions are complete. In this procedure, the ontology can be created by using questions and answers for predefined motivating scenarios, which represents main concepts, properties, relations and axioms on the ontology. The methodology is very well structured and can be extended beyond its original scope.

METHONTOLOGY Methodology: This methodology proposed a structured method to build ontologies from scratch (Fernández-López et al., 1997). Seven steps are followed in the technique, which are: 1) specification phase, 2) knowledge acquisition, 3) conceptualization, 4) integration, 5) implementation, 6) evaluation and 7) lastly, documentation.

SENSUS Methodology: The SENSUS-based methodology was proposed while developing the SENSUS ontology at the ISI (Information Science Institute) Natural Language Research Group to provide a broad conceptual structure for developing automated machine translators (Knight *et al*., 1995 The five steps proposed by SENSUS to build a particular domain ontology are: 1) First, to collect a series of terms which are taken as seed. 2) These seed terms are then linked manually to SENSUS, 3) All the collected concepts in the path from the seed to the root of SENSUS should be included, and, 4) Finally, to check those nodes which have a large number of paths and add the entire subtree if needed. This is a manual process, since it seems to require good understanding of the domain to make the right decision (Swartout et al., 1996).

WordNet Methodology: WordNet is a lexical database for the English language. It groups English words into sets of synonyms called synsets, provides a short description of all terms, and captures the various semantic relations between these synonym sets (Miller G. A. et al. 1990). The purpose of WordNet is twofold: to produce a combination of dictionary and thesaurus that is more usable, and to support automatic word sense disambiguation and natural language processing tasks. The hypernym/hyponym relationships among the noun synsets can be interpreted as specialization relations between conceptual categories. In other words, WordNet can be interpreted and used as a lexical ontology in computer science applications. However, such ontology should be corrected before being used since it contains hundreds of basic semantic inconsistencies such as (i) the existence of common specializations for exclusive categories and (ii) redundancies in the specialization hierarchy. Furthermore, transforming WordNet into a lexical ontology usable for knowledge representation should normally also involve (i) distinguishing the specialization relations into *subtypeOf* and *instanceOf* relations, and (ii) associating intuitive unique identifiers to each category.

Ontology Development 101: This methodology was proposed by the Stanford Center for Biomedical Informatics Research (BMIR) to develop ontologies using the Protégé tool (Noy *et al*. 2001). The wine and food examples were used in the methodology guide, which is loosely based on an example knowledge base presented in the paper on a description-logics approach by (**Brachman** *et al*. 1991). However, the technique also clarified that ontology development is different from designing classes and relations in object-oriented programming. This guide tried to provide developers with an initial roadmap that would help a new ontology designer to develop ontologies. Finally, the authors conclude that there is no single correct ontology-design plan, and this methodology should be used as a reference for domain ontology development.

Integrated Methodology: Integrated ontology development methodology (Chaware and Rao 2010) is quite similar with the Gruninger and Fox approach (Gruninger, 1995), and, is divided into four steps: 1) Motivating user scenarios or keyword. 2) Formulation of informal/formal questions and answer modules, 3) extracting of terms and constraints module, and, 4) Finally, build ontology based on top-down approach. This methodology is validated only in a shopping mall scenario.

YAMO: YAMO is methodology for large-scale ontology development. The methodology is motivated by facet analysis (Ranganathan 1967) and an analytico-synthetic classification approach. The approach ensures the quality of the system precisely because of its flexibility, hospitable, extensible, dense, and complete. YAMO consists of two-way approaches: top-down and bottom-up. The YAMO paper took food as an example domain and used it to illustrate the methodology. A user interview was conducted with a group of people to gather a practical overview, which provided more insight into the theoretical understanding of the domain (Dutta et al., 2015).

The above knowledge modelling methodologies, while serving their own specialized purpose of formulating different genres of well-founded knowledge artifacts (e.g., formal classifications, ontologies, KGs, etc.), do have crucial *limitations* if they are analyzed from the perspective of the general pedagogical norms, principles and stress of knowledge organization (see, amongst others, Hudon 2011; Joudrey and McGinnis 2014; Salaba 2020) and, more general, of library and information science education (see, for instance, Satija, Tiwari and Bagchi 2022). Therefore, some of the general limitations (individually and cumulatively) which we have qualitatively analyzed are as follows.

- Some of the above methodologies are very formal and fit for very specialized, small-scale applications or contexts, thereby, out-of-bounds of the requirements of a pedagogically flexible knowledge modelling methodology which can be taught to domain experts.
- Some methodologies, like Methontology, while being better structured, lack the design principles which a pedagogically flexible knowledge modelling methodology should embody, this, being crucial to either adopt or adapt it depending on context (e.g., nursing within healthcare, hu-

Methodology	Common Dimension	Distinct Dimension	Pedagogical Dimension
Our methodology	middle-out approach, description logics	UX Evaluation + Top-level alignment + Committed to OntoClean	Online Teaching Material + Hybrid Classroom Experimentation + Quadruple Helix framework
IDEF5 Capture	top-down approach	No	No (As Per Literature)
Uschold And King's	top-down approach	No	No (As Per Literature)
Gruninger And Fox's	top-down approach, first-order logic	No	No (As Per Literature)
METHONTOLOGY	life cycle	No	No (As Per Literature)
SENSUS		No	No (As Per Literature)
WordNet	top-down approach, lexical relationship	No	No (As Per Literature)
OD 101	top-down approach,	No	No (As Per Literature)
Integrated	top-down approach,	No	No (As Per Literature)
YAMO	middle-out approach, description logics	No	No (As Per Literature)

Table 1: Comparison of Ontology Development Methodologies

man resource management within healthcare, etc.) within which domain experts are working in.

- None of the above methodologies provide clear guidelines as to how existing standards and conventions can be reused and integrated into an easy-to-teach knowledge modelling methodology (whether technological specifications or domain specific conventions).
- For almost all of the former methodologies, there is a complete lack of clarification on how to use the methodology to teach students (e.g., domain experts like healthcare professionals) the identification and extraction of concepts required for a knowledge model like an ontology and their linguistic labels. Further, no details exist on how to teach the clustering of synonymous words and how to handle knowledge organization codification systems used in specific domain applications (e.g., in healthcare).
- Existing methodologies are never grounded in the user experience (UX) point of view (Hartson and Pyla 2012). This means that they are never analyzed from the user perspective i.e., whether a methodology is pedagogically easy to understand, teach or tweak, etc. Further, a UX analysis of any knowledge modelling methodology can also reveal the extent to which its phases require interdisciplinary pedagogical interdependence.
- As per our knowledge, existing methodologies don't factor in the pedagogically crucial *Quadruple Helix Model* (see Note 2) as proposed by Carayannis and Campbell (2009).

Additionally, in the practice homepage (see https://www.ceic.ie/sandpit-practice), we include examples of how healthcare professionals can define key requirements for a Use Case. In the example, we provide a completed Use Case Template and Scenario Template based on the Antilope Report revised in 2015 by Vincent van Pelt and Michiel Sprenger from the Nictiz Organisation in the Netherlands. It details the process steps to develop a Use Case Report. The cited Antilope report provides a number of "tools" that can be used in solving interoperability challenges. Notice that this dimension, while being pedagogically flexible, has not been considered by any of the aforementioned methodologies.

3.0 The Pedagogically Flexible KG Construction Methodology

We now describe, step-by-step, our proposed pedagogically flexible methodology for teaching how to organize domain knowledge into concrete knowledge organization structures such as an ontology-driven knowledge graph (KG). To that end, we illustrate the steps of the methodology with a running example of designing an ontology-driven Healthcare Knowledge Graph (HKG) by a study group of healthcare professionals (anonymized here for privacy) who were registered with our digital healthcare course (Hussey and Das 2021). Notice that, besides elucidating and illustrating the detailed steps of the methodology, we also provide highlights, wherever relevant, of why they are relevant from the perspective of the principles of knowledge organization education.

3.1 Reference Scenarios

The first step of the proposed methodology is to teach the bootstrapping of a possible reference scenario which can be modelled for a chosen domain. For example, in the healthcare domain, a possible reference modelling scenario for the study group was:

Any entity working with healthcare related data, such as a public healthcare data controller (e.g., NSS (NHS National Services Scotland), Health Service Executive (HSC)), needs to optimize their data preparation pipelines in terms of time, effort, and competences invested into preparing data for research experiments. To that end, a key issue to address is data heterogeneity and the tedium associated with it in the daily work of data analysts (e.g., repeated solving of similar heterogeneity issues in order to maintain an acceptable quality of service to clients) who, often, have to crucially collaborate with diverse genres of domain experts (e.g., healthcare professionals specializing in different aspects of healthcare). Besides, some other related issues are interoperability with data from external sources and interfacing with clients not familiar with local conventions and practices.

3.2 Generalized Queries

The second step of the methodology is to teach the mechanism by which domain experts can generate well-formed competence queries (Grüninger and Fox 1995) for a chosen reference scenario in a chosen domain. This is crucial to pedagogically facilitate two key aspects. Firstly, by learning to define a set of general competence queries which different personas can have in a reference domain scenario, the domain expert has a head-start in conceptually determining the concepts and their interconnections which will ultimately compose the ontology-driven knowledge graph model. Secondly, the domain expert also has an excellent chance to learn, hands-on, the technique to instantiate such generalized competence queries into specific queries. We provide an illustration below which the study group, by learning and practicing our methodology, determined the concepts which would be key to design their HKG, starting with some generalized competence queries below.

- Q1 all patients (X) who are diagnosed with condition (D) from country (Z)

- Q2 all hospitals (X) and healthcare (Y) clinics which have (Z) clinical specialty.
- Q3 all available drug products (X) for a disease (Y) with their daily dose (Z).
- Q4 all patients (X) who visited hospital (Y) from time t1 to time t2.
- Q5 shows the geospatial distribution of patients (X) with their prescribed quantity (Z).

3.2.1 Identify the Domain Concepts

In the above context, some of the concepts which can be crucial for the healthcare professional towards learning to design the HKG were:

Electronic Health Record (EHR): EHR refers to an individual person's medical record in digital format. It may be made up of electronic medical records from many locations and/or sources. The EHR is a longitudinal electronic record of personal health information generated by one or more encounters in any care delivery setting. Included in this information are a person's demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data and radiology reports (Reich et al. 2017).

Electronic Medical Record (EMR): An electronic medical record is a computerized medical record created in an organization that delivers care, such as a hospital or outpatient setting. Electronic medical records tend to be a part of a local stand-alone health information system that allows storage, retrieval and manipulation of records. This document will reference EHR moving forward even if specific data sources might internally use EMR definition (Reich et al. 2017).

Person: The Person class contains records that uniquely identify each patient in the source data who is time-at-risk to have clinical observations recorded within the source systems (Reich et al., 2017).

Prescription: written instructions from a physician or dentist to a druggist concerning the form and dosage of a drug to be issued to a given patient (Miller, G. A. et al. 1990).

Visit: The visit class contains the spans of time a Person continuously receives medical services from one or more providers at a care site in a given setting within the healthcare system. Visits can be classified into four categories: outpatient care, inpatient confinement, emergency room, and long-term care. Persons may transition between these settings over the course of an episode of care (for example, treatment of a disease onset) (Reich et al. 2017).

Death: The death class contains the clinical event for how and when a Person dies (Reich et al. 2017).

Specimen: The specimen class contains the records identifying biological samples from a person.

3.3.2 Query Collection

In principle, there are various ways to teach the collection of competence queries. Domain experts collect such queries alone OR with the collaboration of a data/knowledge scientist. For example, query collection can be done by analyzing the millions of user queries stored in the query logs of existing search portals or from interviewing people. As intended clients, in our example, can be from national health services, such as, NHS (National Health Service) Scotland, HSE (Health Service Executive) Ireland and/or from pharmaceutical companies, a discussion of the study group with them was extremely helpful to list relevant queries. We illustrate some examples below, inspired by from the same example, as to how reference persona scenarios with their medical conditions can be described. See Figure 1 and the dedicated link https://www.ceic.ie/sandpit-citizen.

3.3 Informal Modelling Phase

This step of our proposed methodology concentrates on teaching domain experts the techniques behind the design of a conceptual model underlying the final output of the ontology-driven knowledge graph model. Again, we exemplify the step by providing indicative examples from the study group who registered with our digital healthcare course.

3.3.1 Pilot Reference Datasets

The first sub-step in the informal modelling phase is to choose well-formatted pilot reference datasets. For example, the group in our course analyzed four datasets as part of their preliminary study to understand how healthcare information is modelled in those datasets. The datasets also provided information regarding material disadvantage in the population in terms of lack of car ownership. They often contained sensitive personal information of a patient such as ethnic group, marital status related to a patient and prescriber (all of which were taught, as required by the standard conventions of the healthcare domain, to be suitably anonymized). Demographic information available in the dataset pertained to gender, address, date of birth and occupation of patients. The methodology also taught the detailed understanding of coding systems used for codifying disease names, such as, ICD-10, ICD-9, ICD-O-2 (Hong and Zeng 2022), as it natively teaches the reuse and adoption of existing technological standards and domain-specific conventions.

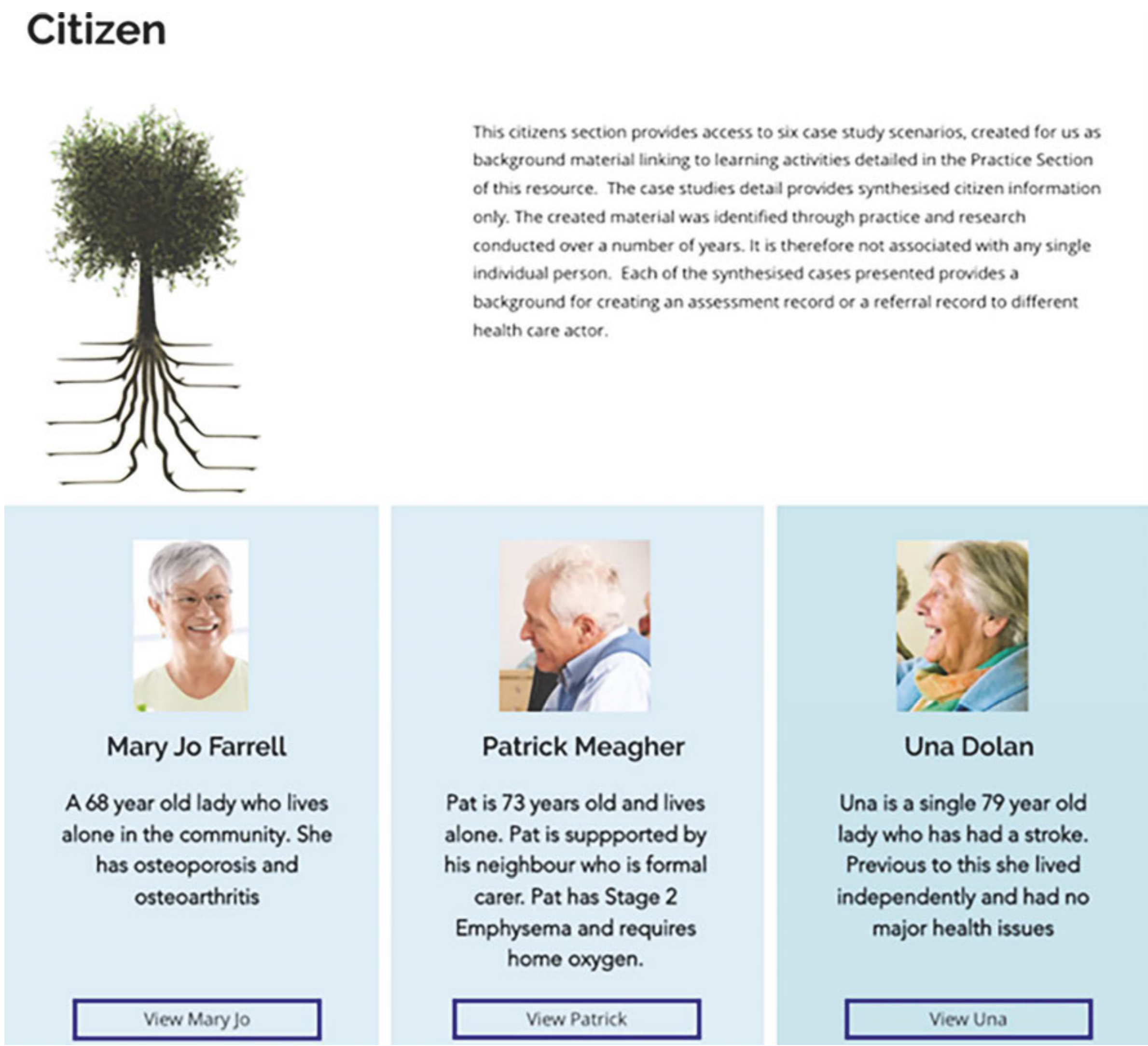

Figure 1: Some Indicative Reference Persona Scenarios

3.3.2 ExER Model

The next step of the methodology is to teach the domain experts to build conceptual models such as Entity Relationship (ER) models (Chen 1976) by mediating between the identified concepts and the analysed pilot reference datasets. The Extended Entity Relationship (ExER) model for healthcare developed by our example study group is shown in Figure 2. The *patient entity* has a community health index (CHI) number as an identifying attribute and age (at time of visit). For the privacy reason, they were taught to separate all personal information of a patient from a person and connected with the *role of* relation. This enables the user to access personal information of a patient if they have the right permission. The *person entity* has all demographic information, such as date of birth, ethnic group, and postal address. With relation to *country of birth,* they were taught to collocate all persons who were born in the same country. The *country entity* has the attributes such as ISO code, population, currency name etc. The *visit entity* stores all information about the patient visit to a health encounter site. The *hospital entity* has an attribute address (a structure attribute) which stores information about house number, street name, street name, postcode, country, and city.

Notice that, from a pedagogical perspective, the above step is very interactive and should be performed by domain experts such as healthcare professionals in collaboration with data and knowledge scientists.

3.4 Formal Modelling Phase

The last step of the methodology teaches the nuances of translating the informal conceptual model (ExER model) developed in the previous step into a completely formal ontological model (Guarino, Oberle and Staab 2009). Figure 3 shows the class hierarchy for the healthcare domain devel-

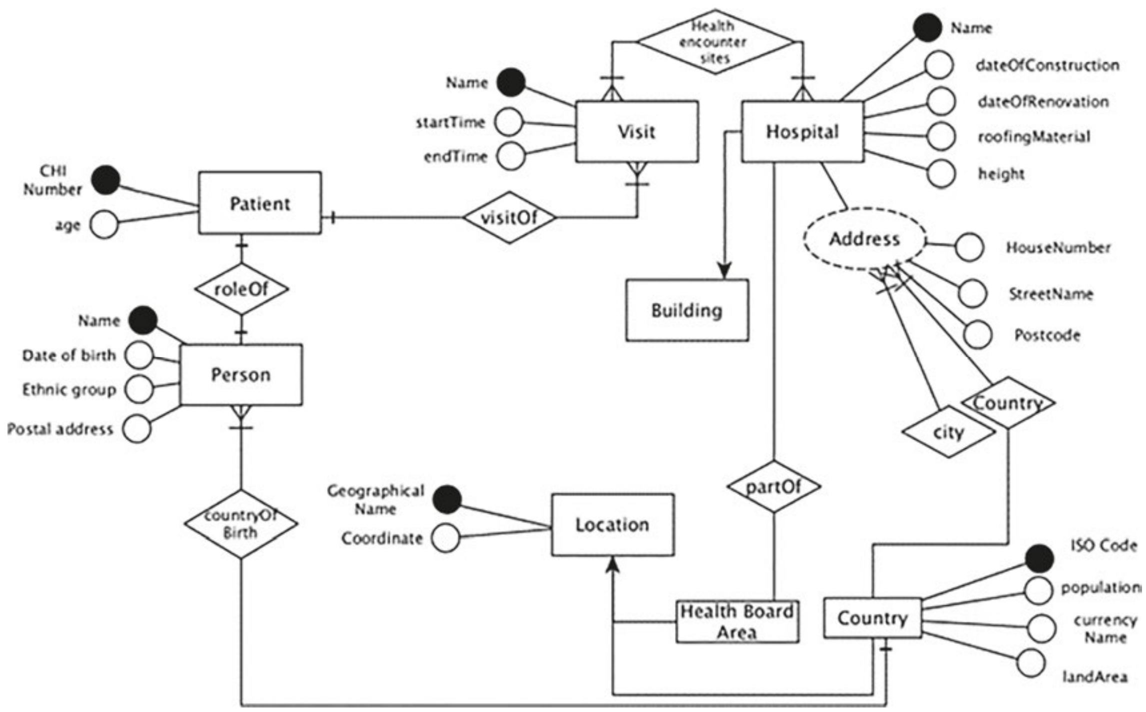

Figure 2: ExER Model for Healthcare

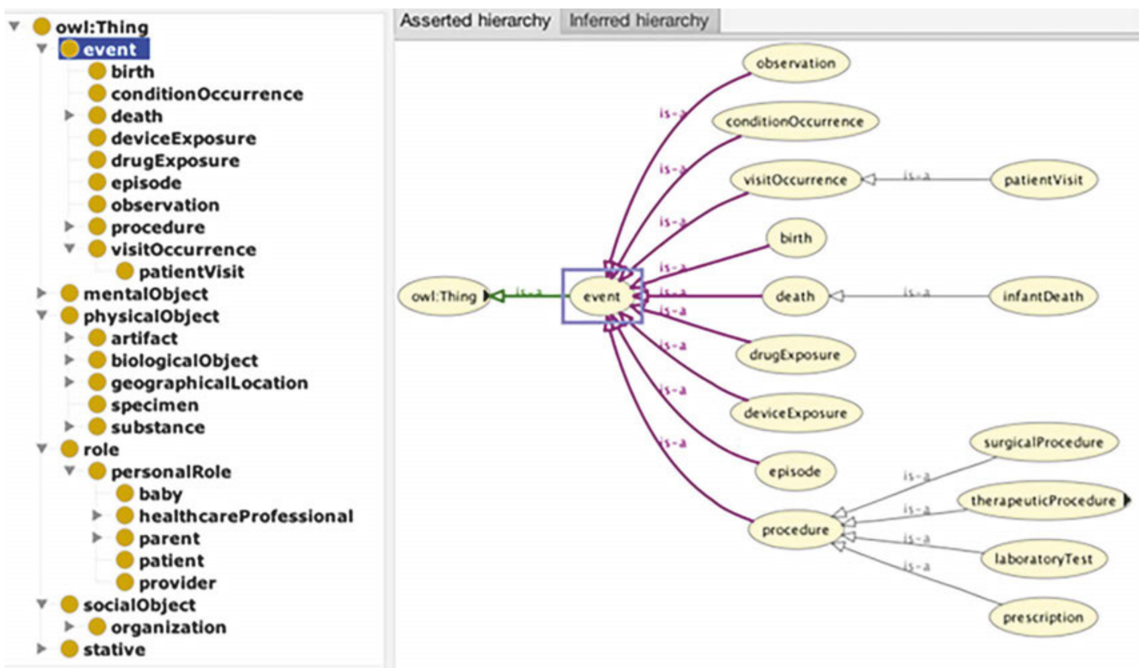

Figure 3: Class Hierarchy of Healthcare

oped by the study group. On the right-hand side of the figure, it displays all subclasses of the event class. It includes medical procedures which record all information related to medical treatment or laboratory tests of a patient. *patientVisit* stores all visiting information of a patient and *episode* is used for collocating all independent but interconnected events related to a patient visit who changes different wards of the same hospital without discharge from the hospital. The intermediate class hierarchy is modelled by the domain concepts. All domain concepts are generated to capture the healthcare domain. Some are reused from OMOP standard (Reich et al., 2017) and some other classes are created to accommodate a domain concept from SNOMED-CT and International Classification of Nursing Practice (ICNP).

We provide, as follows, a brief description of all domain concepts reused by the study group from the OMOP data model (Reich et al., 2017). This is *key* to understanding the pedagogical effectiveness of our proposed step-by-step guiding methodology.

Death: The death class contains the clinical event for how and when a person dies. A person can have up to one record if the source system contains evidence about the death, such as: 1) condition code in the header or detailed information of claims 2) status of enrollment into a health plan 3) explicit record in EHR data.

Visit occurrence: The visit occurrence class contains the duration of time a Person continuously receives medical services from one or more providers at a Care Site (i.e., Hospital) in a given setting within the health care system (e.g., NHS system). Visits are classified into four categories: outpatient care, inpatient, emergency, and long-term care. Persons may transition between these categories over the course of an episode of care (for example, treatment of a disease onset).

Device exposure: The device exposure class captures information about a person's exposure to a foreign physical object (or device) or instrument that is used for diagnostic or therapeutic purposes through a mechanism beyond chemical action. Devices include implantable objects (e.g. pacemakers, artificial joints), medical equipment and supplies (e.g. bandages, crutches, syringes), other instruments used in medical procedures (e.g. sutures, defibrillators) and material used in clinical care (e.g. adhesives, body material, dental material, surgical material).

Drug exposure: The drug exposure class captures records about the utilization of a Drug when ingested or otherwise introduced into the human body. A Drug is a biochemical substance formulated in such a way that when administered to a Person it will exert a certain physiological effect. Drugs include prescription and over-the-counter medicines, vaccines, and large-molecule biologic therapies. Radiological devices ingested or applied locally do not count as Drugs. Drug Exposure is inferred from clinical events associated with orders, prescriptions written, pharmacy dispensing, procedural administrations, and other patient-reported information, for example: in the prescription section of an EHR, the Medication section of an EHR or Drugs administered as part of a Procedure (or therapy), such as chemotherapy or vaccines.

Procedure occurrence: The procedure occurrence class contains records of activities or processes ordered by, or carried out by, a health-care provider on the patient to have a diagnostic or therapeutic purpose. Procedures are present in various data sources in different forms with varying levels of standardization. For example: 1) Medical Claims include procedure codes that are submitted as part of a claim for health services rendered, including procedures performed. 2) EHR that captures procedures as orders. The class relationship visualization developed by the study group has been shown in Figure 4. It is showing the dependencies between 9 classes. In particular, the prescription has 1545 relations in total which include 545 incoming relations and 1000 outgoing relations. This, additionally, testifies the pedagogical validity and learning flexibility provided by the steps of our proposed methodology.

We also briefly describe some highlights of the computer science and knowledge engineering foundations and practices which are *demonstratively taught,* as technical accompaniments to the proposed general methodology, to healthcare professionals as part of our digital healthcare course. We explain, simultaneously to the theoretical underpinnings of our methodology, how to create, using ICT tools and standards, an interconnected knowledge graph model to visualize healthcare data from initial development of data to connecting data with the open-source community. The entire process includes seven main *action steps*, six of which we outline in this guidance document. The activity between step 4 and step 5 cannot be effectively demonstrated, and therefore is not discussed in detail as part of the process in this guidance document. The remaining six steps are briefly outlined with associated diagrams and signposted. Notice that these *action steps* are completely aligned and harmonized with the proposed methodological phases that we teach.

Step 1: The first step in the process relates to teach, in collaboration with the data and knowledge scientists, the creation a demonstrator for **Data Collection**, the focus being to design electronic forms, which are able to capture healthcare data in a structured way (see Figure 5), thereby, moving from a *Model of Use* (a paper based form) to a *Model of Meaning* (a demonstrator cloud-based system which offers a framework for machine to machine communication). For instance, agreed information is broken into single data fields rather than clustered into a text box. Attributes presented to depict an address are on different rows. For example, the address format on Amazon is presented as *Street Name, Apartment Name, etc.* all of which are entered in different fields or rows. While designing the form for data collection, it is also important to give preference to capturing what terms are core, what terms are common, and finally consider what local terms must be used within the organization (Giunchiglia *et al*. 2022). These are entitled context specific terms. Once data has been agreed and structured for data collection in the demonstrator. The

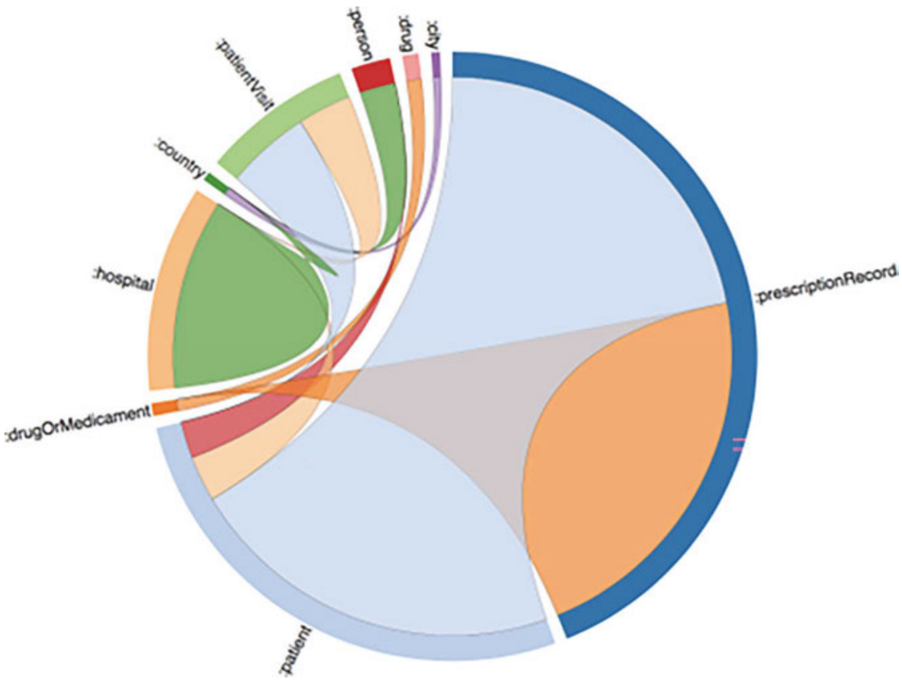

Figure 4: Class Relationship-Healthcare

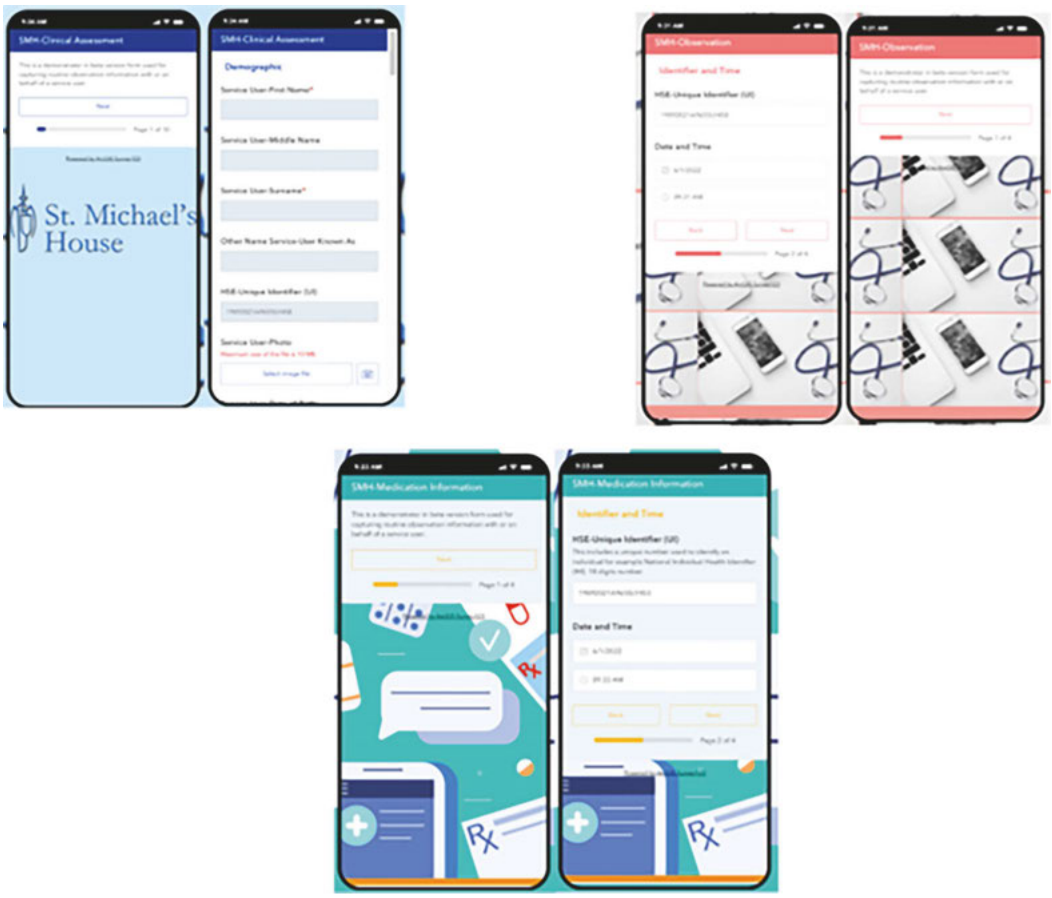

Figure 5: Prototype Mobile App for Data Entry and Collection

system is tested by entering synthesized data collected into the system through an online form presented on a mobile device. Once this process is completed, the demonstrating of the teaching can then proceed to check the quality of the data that was entered onto the form in **Step 2**.

Step 2: The second step teaches the technique of evaluating the quality of the data collected. For example, the data that is stored in your demonstrator database involves reviewing how your data is stored. While the data and knowledge scientists' team have applied a number of conformance rules initially, there is a need to revise and check any rules deployed are working according to the demonstrator system requirements and domain expertise. This may involve an iterative set of processes that you may wish to review and revise at the point of data entry. The subsequent step involves signing off on the processes, which underpin the data you are collecting in your demonstrator database. Specific activities that one needs to check include reviewing summary reports to ascertain if there are any missing values or anomalies within data, which one has agreed to collect. Checking summarized data, which is stored for any errors, is important and this can be done using administrative password and login software provided through the administrator authentication and permissions system. Data at this time is viewed through the summarized table view. Remember this view is restricted to only domain experts who are monitoring the data collated. Step 2 also provides an opportunity to see metadata (Satija, Bagchi and Martínez-Ávila 2020) that is the data (i.e. Submitted by and Submitted time), which describes the data collected in the demonstrator database. In this context, data storage is taught as the use of recording media to retain data using computers or other devices. The most prevalent forms of data storage taught are file storage, block storage, and object storage, with each being ideal for different purposes.

Step 3: The third step teaches the methodological translation of the required semantic information from the original Model of Use (data collected in everyday practice in paper format) to the agreed data for the Model of Meaning (data to be collected and transferred from machine to machine). Therefore, the pedagogical instructions of this step involves teaching the know-how of performing semantic mapping across different tools which requires translation and alignment of the agreed information now presented as formalized terms into a readable quantified language represented as named entity types and properties. This facilitates the computer-to-computer interactions which support semantic meaning and requires defined ontological schema. This translation process is an important step in methodological development but also one which is complex and specialized. The theoretical and scientific elaboration of the processes taught in this step is provided in Section 3.3 and Section 3.4 above.

In Figure 6, we present how we teach the data mapping to an ontological schema to generate the final ontology-based knowledge graph model. The upper part of the figure which is represented in the black box illustrates the classes or entity types associated with the ontological schema, whereas, in the bottom part of the figure, one can see all terms in a table format which we taught to gather during the processes associated with Steps 1 and Step 2. The arrow illustrated in Figure represents the direct mapping between data and concepts. It is a semi-automated process using the Karma tool (Knoblock et al. 2012) just to say that this mapping needs to be provided in supervision of a human expert.

Step 4: The fourth step connected with the previous step teaches the techniques of data integration. Data integration is the combination of technical and business processes used to combine data from disparate sources into meaningful and valuable information. It includes looking at data more carefully and finding out if there is any anomaly or whether all values collected are correct e.g., is there any null value or empty column? What about the date format? Are they all the same? Is ICD-10 code properly stored or entered? etc. This process is called data integration.

Step 5: The fifth step is all about teaching Data Visualization, more precisely, data visualization through a Knowl-

Figure 6: Semantic Data Mapping Example

edge Graph (more about different data visualization techniques and knowledge graphs can be found in the FAQ section of https://sites.google.com/dcu.ie/csdm/faq). Here, we teach how one can see how one name entity is related with another name entity using a directed arrow. Node represents class or type (i.e. Person, Hospital, Location etc.) of a name entity. Different colours are associated with the different classes of a name entity. This kind of graph visualization shows how different elements of a healthcare system are interconnected. It is very intuitive in nature to find out new information otherwise hidden in silos of different hospital systems or databases. Data Visualization is part of many business-intelligence tools and key to advanced analytics. It helps people make sense of all the information, or data, generated today. An illustration of data visualization is shown in Figure 7.

4.0 Evaluation of the HKG Model

4.1 Query Evaluation

As this step is very technical to teach, we don't always expect domain experts (e.g., healthcare professionals) to conduct it with their own effort. Rather, the focus of the pedagogical instruction at this phase is that the domain experts supervise and validate the query results performed by the data scientists (e.g., within a healthcare organization). For example, the main responsibilities of the healthcare professionals are to check and confirm what was their expectation from the given scenario, what they had in their mind when they defined persona (described in Section 3.1). To exemplify, for a healthcare domain project by the study group, we illustrate some query evaluations. To that end, we mention some of the SPARQL queries (Pérez et al. 2009) to show that the knowledge graph model was able to answer the necessary questions which might be useful for further research and analysis. This, additionally, provides a flavour of how our proposed methodology pedagogically interoperates between theory, practice and interdisciplinarity.

Q1. Select all female patient who visited hospital and prescribe quantity.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-
   schema#>
PREFIX shib: <http://www.semanticweb.org/shib>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-
   syntax-ns#>
SELECT DISTINCT  ?subtanceName ?Age
   ?prescribedQuantity
 WHERE {  ?patientVisit shib:ageInYears ?Age.
          ?patientVisit shib:visitOfPatient ?patient.
          ?patient shib:roleOf ?person.
          ?person shib:forename ?PersonForename.
        ?person shib:surname ?PersonSurname.
  OPTIONAL {?person shib:findingRelatedToBiological
   Sex ?Sex.}
      ?person shib:countryOfBirth ?Country.
      ?Country rdfs:label ?CountryOfBirth.
      ?patient rdfs:label ?PatientUPI.
```

Figure 7: Data Visualization (Fragment)

```
        ?patient shib:hasPrescription ?Prescription.
        ?Prescription rdfs:label ?PrescriptionID.
        ?Prescription shib:prescribedQuantity
    ?prescribedQuantity.
        ?Prescription shib:prescriptionDrug
    ?drugSubstance.
        ?drugSubtance rdf:type shib:drugOrMedicament.
        ?drugSubtance rdfs:label ?subtanceName.
        ?drugSubstance shib:legalDrugProduct ?drug
    Product.
        ?drugProduct rdfs:label ?ProductName.
FILTER regex(?Sex, "Female")
}
```

Q2. Select all male patient who visited hospital and prescribe quantity.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-
    schema#>
PREFIX shib: <http://www.semanticweb.org/shib>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-
    syntax-ns#>
SELECT DISTINCT   ?subtanceName ?Age
    ?prescribedQuantity
 WHERE {    ?patientVisit shib:ageInYears ?Age.
            ?patientVisit shib:visitOfPatient ?patient.
                  ?patient shib:roleOf ?person.
            ?person shib:forename ?PersonForename.
            ?person shib:surname ?PersonSurname.
   OPTIONAL {?person
    shib:findingRelatedToBiologicalSex ?Sex.}
        ?person shib:countryOfBirth ?Country.
        ?Country rdfs:label ?CountryOfBirth.
        ?patient rdfs:label ?PatientUPI.
        ?patient shib:hasPrescription ?Prescription.
        ?Prescription rdfs:label ?PrescriptionID.
        ?Prescription shib:prescribedQuantity
    ?prescribedQuantity.
        ?Prescription shib:prescriptionDrug
    ?drugSubstance.
        ?drugSubtance rdf:type shib:drugOrMedicament.
        ?drugSubtance rdfs:label ?subtanceName.
        ?drugSubstance shib:legalDrugProduct
    ?drugProduct.
        ?drugProduct rdfs:label ?ProductName.
              FILTER regex(?Sex, "Male")
 }
```

Q3. Retrieve the all prescribed drugs information of a patient along with the name of drugs manufacturer and patient personal information.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-
schema#>
PREFIX shib: <http://www.semanticweb.org/shib>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-
syntax-ns#>
SELECT ?PatientUPI ?PersonForename ?PersonSurname
?CountryOfBirth
?PrescriptionID ?subtanceName ?ProductName
   WHERE {  ?patient shib:roleOf ?person.
            ?person shib:forename ?PersonForename.
                  ?person shib:surname ?PersonSurname.
            ?person shib:countryOfBirth ?Country.
            ?Country rdfs:label ?CountryOfBirth.
            ?patient rdfs:label ?PatientUPI.
            ?patient shib:hasPrescription ?Prescription.
            ?Prescription rdfs:label ?PrescriptionID.
            ?Prescription shib:prescriptionDrug
?drugSubstance.
            ?drugSubtance rdf:type
shib:drugOrMedicament.
```

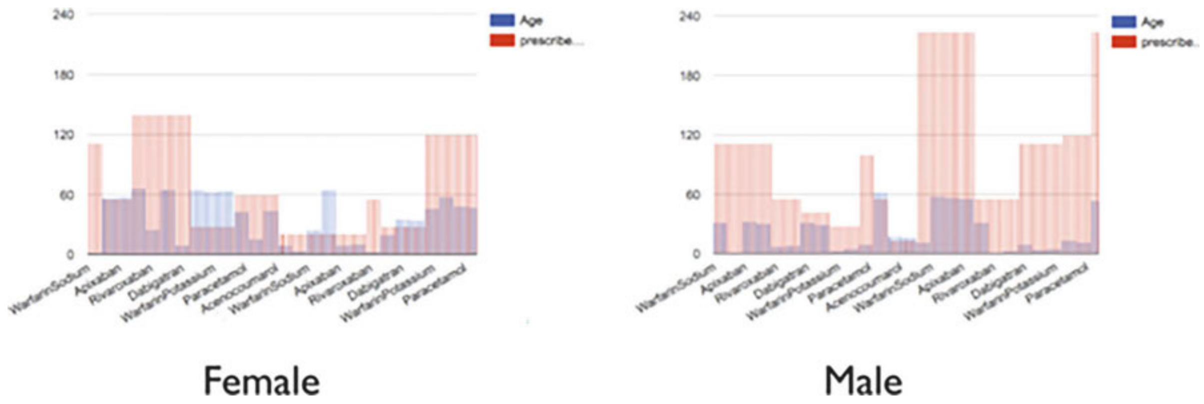

Figure 8: Comparison chart showing prescribed quantity of Drug X in respect with their age (Result of Q1 Q2)

```
        ?drugSubtance rdfs:label ?subtanceName.
        ?drugSubstance shib:legalDrugProduct
?drugProduct.
        ?drugProduct rdfs:label ?ProductName. }
```

Q4 Show all connection of a patient in a knowledge graph using SPARQL construct.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-
schema#>
PREFIX shib: <http://www.semanticweb.org/shib>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-
syntax-ns#>
CONSTRUCT { ?patientVisit shib:visitOfPatient
?patient.
        ?patient shib:roleOf ?person.
        ?person shib:countryOfBirth ?Country.
        ?patient shib:hasPrescription ?Prescription.
        ?Prescription shib:prescriptionDrug
?drugSubstance.
        ?drugSubstance shib:legalDrugProduct
?drugProduct.

    }
```

The knowledge graph (see Figure 9) shows how the patient 1234567928 (PatientId) is related with the person 38 (personID) with role of relation. It also visualizes the seven prescriptions which have been prescribed to the patient. All prescriptions prescribed the same drug, which is *warfarin sodium*. And for that generic drug, there are four brand products available in the market. As a validation of our methodology, teaching to showcase such information using knowledge graph (KG) visualization is immensely useful for healthcare professionals as they can easily understand and grasp patients' necessary information needed for their clinical assessment. The methodology not only helps healthcare professionals to interact and read the KG but also empowers them in theory and practice of knowledge organization and modelling.

4.2 User Experience (UX)

We conducted a user study to assess the pedagogical usability (Bagchi 2022b) of methodology for domain knowledge organization and modelling. During an entire period of a semester, students were asked to perform a series of steps described before to design a domain ontology based KG as an assignment for Knowledge and Data Integration Course (KDI) at the University of Trento, Italy (http://dit.unitn.it/~ldkr/ldkr2016/index.html), and CS6422 - eHealth System Standards & Digital Service in Dublin City University (DCU), Ireland (https://www101.dcu.ie/registry/module_contents_no_mod.php?function=2&subcode=NS5058). At the end of the semester, we asked all the participants to fill an online questionnaire. An open-ended group discussion was also conducted at the end of the session. Data from the discussion session were further analyzed and mapped with the online questionnaire. The questionnaire was designed to understand different UX dimensions (Laugwitz et al. 2008) along with the specific traits of the methodology. (Laugwitz et al. 2008) came up with more generic UX dimensions. They are:

1. Attractiveness relates to the overall pedagogical impression of the methodology.

Figure 9: Partial view of Healthcare Knowledge Graph (Q4)

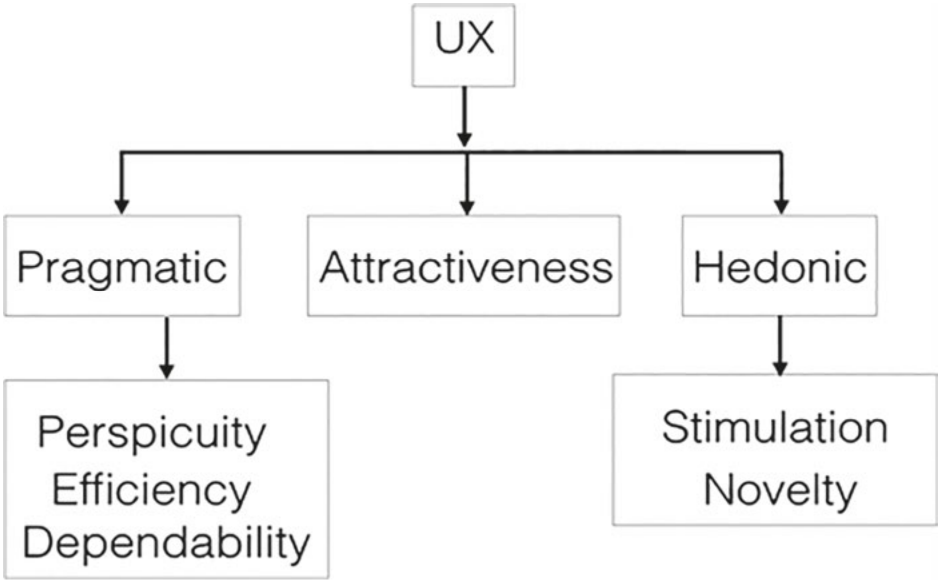

Figure 10: User Experience dimensions. Adapted from (Laugwitz et al., 2008)

2. Usefulness of the methodological ecosystem across three dimensions: i) Efficiency, which deals with how easily the methodology can be learned and used; ii) Perspicuity, which deals with the familiarity of the methodology, and iii) Dependability, which deals with the user's feeling of the control of the methodology.
3. Users stimulation composed of stimulation (deals with feeling of excitement while learning and using the methodology) and novelty (the pedagogical innovation of the methodology.).

These UX dimensions helped us perform a thorough assessment of the methodology using five scales with 20 terms. These scales were: Perspicuity, Efficiency, Dependability, Stimulation, and Novelty. These scales were further categorized as Pragmatic (the learnability and usability of the methodology and basically consists of Efficiency, Perspicuity, and Dependability) and Hedonic (learners' stimulation while seeing the methodology quality). It consists of 2 scales: Stimulation and Novelty. After that, to validate the pedagogical usefulness of the methodology, we asked the participants to point out the advantages and disadvantages of the methodology. A total number of 38 participants (13 Male, 25 Female; 17 within an age range of 22-25 and 21 within an age range of 26-49) took part in the user study. Participants' highest degree of education ranged from postgraduate (9) to professional (29) degree. 17 of them were studying digital health whereas one was from the linguistics department. Some of the participants were fully aware of the semantic technologies and tools whereas others had no prior knowledge regarding semantic knowledge modelling technologies. All of the participants possessed good knowledge of English and volunteered for the study. The consent form was signed prior to the beginning of the study. Throughout the semester, the students were asked to model an ontology based KG following the methodology.

The analysis of the result of user evaluation shows (see Figure 11 for a glimpse across the abstract UX dimensions) that the age, qualification and location had no impact on learning and using the methodology. The main motivation of the learners was to get good credits on the course and, principally, to learn data integration tools and techniques which they could exploit in the future for their own reference scenarios. The results obtained from the participants were mostly positive with a mean value of (0.86) for perspicuity, (1.181) for efficiency, (0.722) dependability, (1.185) stimulation and (0.056) for novelty. The overall Pragmatic quality was 0.92 whereas the Hedonic quality was 0.63. This shows that the methodology was user friendly and easy to learn. However, some of the participants didn't find the technologies involved stimulating (while being relatively negative, this is expected as there is a limited number of state-of-the-art semantic technologies to choose from). The elaborated examples and one to one teacher - student interaction made the methodology easy to learn and reproduce. The practical benefits were also noticed by the participants. Some participants, though, were unsure of the benefit ontology driven KG would bring to their precise working context. This, in fact, provides us food for thought to make the methodology even more learner-friendly.

The formalization process of the ontology driven KG was also considered difficult by some student professionals. Similarly, some participants felt that not all the generalized

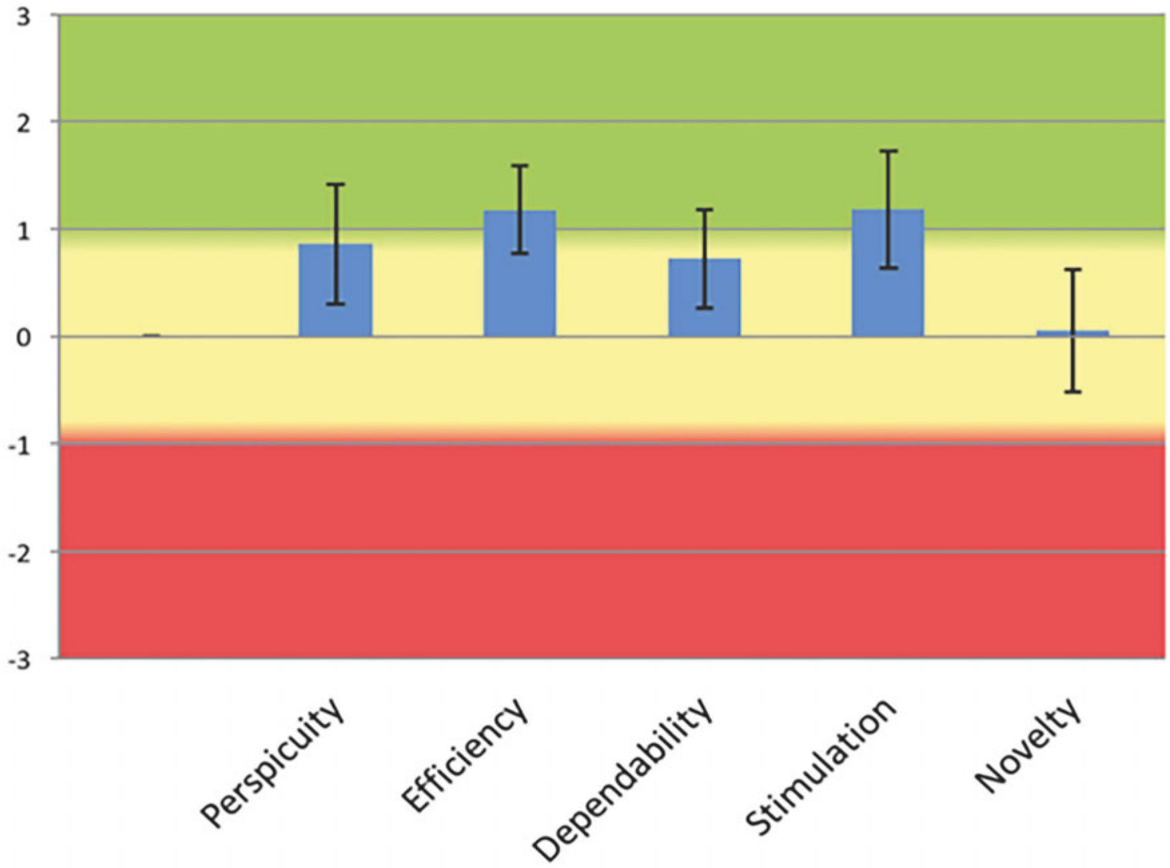

Figure 11: Result of Methodology Evaluation

queries were answered by the model. Some participants felt more emphasis had to be made on the linguistic level rather than on generalized queries. Some felt the development of the life cycle could be more streamlined by using the results. Considering the suggestions from the participants, we will plan to perform a new comparative enrichment and evaluation of the methodology from a pedagogical as well as scientific perspective. Similar questionnaires will be asked based on five UX dimensions to a new set of students taking the course. After that, we will perform the comparative analysis over the data and test the pedagogical efficiency of the methodology.

5.0 Conclusion

Before the inception of any pedagogical development using ICT, we must first determine the required components through an interactive design process. In this paper, we have not committed to an explicit implementation of the various "design thinking, "human centered design, or "integrative thinking processes", which, otherwise, are excellent models to follow, Instead, we aligned our methodology and course to the crucial recommendations by the Health Informatics Course run by MITx [https://sana.mit.edu/node/206], which we need to consider when integrating an informatics solution in a domain specific environment (e.g., a global health environment). In addition to the combined scientific-pedagogical elucidation of the methodology, we also provided exemplifications (see Note: (1)) to illustrate that the proposed pedagogically flexible knowledge modelling methodology works, both in theory and *practice*, and, it can be instantiated to develop knowledge organization models such as ontologies and KGs in any reference domain (healthcare, in our case). In order to accomplish these challenging yet conflicting goals, we recommend incorporating the following mechanisms into the knowledge modelling design process:

Stakeholder Analysis

Understanding the key players and users involved with an intervention is central to deploying a usable and effective solution. This goes beyond just end users to include anyone who is involved in the system's delivery process for your intervention.

Workflow Integration

Any technical innovation needs to be thoughtfully integrated within the workflow. Rather than designing around a novel technology, it is often constructive to begin with the clinical intervention and analyze where technology can play a role and improve processes. At the outset, it is helpful to investigate any significant barriers to implementation, such as pre-existing systems, user hesitance, immutable processes, or insufficient human resources.

An aspect of workflow integration is a strong emphasis on workflow efficiency. Clinicians are notoriously overworked, with limited time and significant amounts spent on documentation already. As much as possible, any innovation needs to avoid adding work or the solution won't be adopted by the clinician user base, nearly regardless of the potential benefit. Therefore, it is imperative to design efficient clinical interfaces that minimize the amount of additional work, with improving efficiency an ideal objective.

An important caveat comes from including the previous recommendation of end-user analysis, where the clinician users will likely have different biases toward what interface improves efficiency from the implementing engineers, so it is prudent to include clinicians in the design process. Most importantly, for any mHealth solution to yield a lasting impact, it must be directly tied to quality improvement. This requires closing the loop with a clinical intervention, so it is strongly recommended that a holistically designed intervention connect the informatics innovation with the clinical process.

This is a challenging proposition, and it requires clinicians to collaborate closely with technologists to push the envelope of combining what is technically feasible with what can make a clinical impact. In addition, these factors must be understood and debated when teaching the proposed knowledge modelling methodology.

Notes

1. The detailed course material is available in the following link: https://www.ceic.ie/ and the instruction to access the course is available in the following link: https://www.dcu.ie/ceic/interoperability-lab.
2. Quadruple Helix innovation models – involves institutional bodies, research sphere, business sector, and citizens in the process. This new generation of open innovation leads to stronger economic impact and better user experience in Europe. https://ec.europa.eu/digital-single-market/en/open-innovation-20.

Acknowledgements

The authors would like to thank the KnowDive Research Group, University of Trento, Italy for providing the academic and research environment within which this idea could flourish. This research has received funding from MF No: 222879, the European Union's Horizon 2020 research and in- novation programme under the ELITE-S Marie Skłodowska-Curie grant agreement No. 801522, by Science Foundation of Ireland and co-funded by the European Regional Development Fund through the ADAPT Centre for Digital Content Technology grant number 13/RC/2106 and DAVRA Networks.

References

Ådland, M. K., and Marianne Lykke. 2018. "Tags on Healthcare Information Websites". In *Challenges and Opportunities for Knowledge Organization in the Digital Age: Proceedings of the Fifteenth International ISKO Conference,* edited by Fernanda Ribeiro and Maria Elisa Cerveira, 684-692. Baden-Baden: Ergon-Verlag.

Bagchi, Mayukh. 2021a. "A Large Scale, Knowledge Intensive Domain Development Methodology." *Knowledge Organization* 48(1): 8-23. https://doi.org/10.5771/0943-7444-2021-1

Bagchi, Mayukh. 2021b. "Towards knowledge organization ecosystem (KOE)." *Cataloging & Classification Quarterly* 59(8):740-756. https://doi.org/10.1080/01639374.2021.1998282

Bagchi, Mayukh. 2022a. "Smart Cities, Smart Libraries and Smart Knowledge Managers: Ushering in the Neo-Knowledge Society." *ArXiv* abs/2202.08118.https://doi.org/10.48550/arXiv.2202.08118

Bagchi, Mayukh. 2022b. "A Diversity-Aware Domain Development Methodology". In *Proceedings of the 41st International Conference on Conceptual Modelling (ER Forum and Symposium, 2022)*, virtual. https://ceur-ws.org/Vol-3211/CR_096.pdf

Bagchi, M. and Subhashis Das. 2022. "Conceptually Disentangled Classificatory Ontologies". In *Proceedings of the 15th Seminar on Ontology Research in Brazil (ONTOBRAS)*, edited by Oliveira et al. 85-93, virtual. https://ceur-ws.org/Vol-3346/Paper7.pdf

Bagchi, M. and Subhashis Das. 2023. "Disentangling Domain Ontologies". In *Proceedings of the 19th Conference on Information and Research science Connecting to Digital and Library science* (*IRCDL)*, edited by Alessia et al., Bari, Italy. https://ceur-ws.org/Vol-3365/short1.pdf

Biagetti, Maria Teresa. 2021. "Ontologies as Knowledge Organization Systems." *Knowledge Organization* 48 (2): 152-176. https://doi.org/10.5771/0943-7444-2021-2-152.

Brachman, R. J., McGuinness, D. L., Patel-Schneider, P. F., Resnick, L. A., & Borgida, A. (1991). Living with CLASSIC: When and how to use a KL-ONE-like language. In Principles of semantic networks (pp. 401-456). Morgan Kaufmann.

Carayannis, E. G., and David F. Campbell, D. 2009. "Mode 3' and 'Quadruple Helix': toward a 21st century fractal innovation ecosystem". *International Journal of Technology Management*, 46(3-4), 201-234. https://10.1504/IJTM.2009.023374

Chaware, S., & Rao, S. (2010). Integrated approach to ontology development methodology with case study. International Journal of Database Management Systems, 2(3), 13-19.

Chen, Peter Pin-Shan. 1976. "The Entity-relationship Model: Toward a Unified View of Data." *ACM Transactions on Database Systems (TODS)* 1(1): 9-36. https://doi.org/10.1145/320434.320440

Das, S., and Pamela Hussey. 2022. "Development of an Interoperable-Integrated Care Service Architecture for Intellectual Disability Services: An Irish Case Study." In *Frontiers of Data and Knowledge Management for Convergence of ICT, Healthcare, and Telecommunication Services*, edited by Suman Paul, Sara Paiva and Bo Fu. Springer, Cham, 1-24.

Dutta, B., Chatterjee, U., & Madalli, D. P. (2015). YAMO: yet another methodology for large-scale faceted ontology construction. *Journal of Knowledge Management*, 19(1), 6-24.

Fernández-López, M., Gómez-Pérez, A., and Natalia Juristo. 1997. "Methontology: from Ontological Art Towards Ontological Engineering". In *Proceedings of the Association for the Advancement of Artificial Intelligence Spring Symposium*, 33-40.

Ghosh, S. S., S. Das and S. K. Chatterjee. (2020). "Human-centric Faceted Approach to Ontology Construction." *Knowledge Organization,* 47: 31-44.

Giunchiglia, F., B. Dutta and V. Maltese. (2014). "From Knowledge Organization to Knowledge Representation." *Knowledge Organization,* 41: 44-56.

Giunchiglia, Fausto and Mayukh Bagchi. 2022a. "Object Recognition as Classification Via Visual Properties." In *Knowledge Organization across Disciplines, Domains, Services and Technologies: Proceedings of the Seventeenth International ISKO Conference*, edited by Marianne Lykke et al., 87-100. Baden-Baden: Ergon-Verlag,

Giunchiglia, Fausto and Mayukh Bagchi. 2022b. "Representation heterogeneity". In: *1st International Workshop on Formal Models of Knowledge Diversity (FMKD) and the 8th Joint Ontology WOrkshops (JOWO)*, Jönköping University, Jönköping, Sweden, https://ceur-ws.org/Vol-3249/paper2-FMKD.pdf

Giunchiglia, Fausto, Simone Bocca, Mattia Fumagalli, Mayukh Bagchi, and Alessio Zamboni. 2022. "Popularity Driven Data Integration." In *Knowledge Graphs and Semantic Web*: *Proceedings of the 4th Iberoamerican Conference and third Indo-American Conference*, edited by Boris Villazón-Terrazas et al. 277-284. Springer International Publishing.

Giunchiglia, Fausto, Mayukh Bagchi, and Xiaolei Diao. 2023. "Aligning Visual and Lexical Semantics." In *Information for a Better World: Normality, Virtuality, Physicality, Inclusivity*: *Proceedings of the18th International Conference, iConference* 2023, Virtual Event, 294-302. Cham: Springer Nature Switzerland https://doi.org/10.48550/arXiv.2212.06629

Gómez-Pérez, A., Natalia Juristo Juzgado and Juan Pazos Sierra. 1995. "Evaluation and Assessment of Knowledge Sharing Technology." In *Towards Very Large Knowledge Bases*, edited by N. J. Mars. Washington, DC: IOS Press, 289-96.

Grüninger, Michael and Mark S. Fox. 1994. "The Design And Evaluation of Ontologies for Enterprise Engineering." In *Workshop on Implemented Ontologies, European Conference on Artificial Intelligence (ECAI)*. Amsterdam, NL. http://www.ie.utoronto.ca/EIL/public/onto_ecai94.ps

Grüninger, Michael, and Mark S. Fox. 1995. "The role of competency questions in enterprise engineering." *Benchmarking—Theory and practice IFIP Advances in Information and Communication Technology*, p. 22-31, edited by Rolstadås, Springer, Boston, MA. https://doi.org/10.1007/978-0-387-34847-6_3A.

Guarino, Nicola, Daniel Oberle and Steffen Staab. 2009. "What is an Ontology?." In *Handbook on ontologies*, edited by S. Saab and R. Studer, 1-17, Springer-Verlag: Berlin Heidelberg. https:// DOI 10.1007/978-3-540-9267 3-c2009

Guizzardi, Giancarlo. 2007. "Conceptualizations, Modeling Languages, and (Meta) Models." In *Databases and Information Systems IV: Selected Papers from the Seventh International Baltic Conference*, DB&IS'2006, vol. 155, 18. IOS Press.

Hartson, Rex, and Pardha S. Pyla. 2012. *The UX Book: Process and Guidelines for Ensuring a Quality User Experience*. Elsevier.

Hjørland, B. 2017. "Domain Analysis".*Knowledge Organization*, 44(6): 436-464. https://doi.org/10.5771/0943-7444-2017-6-436

Hong, Yi, and Marcia Lei Zeng. 2022. "International classification of diseases (ICD)." *Knowledge Organization* 49(7): 496-528. https://doi.org/10.5771/0943-7444-2022-7-496

Hudon, Michèle. 2011. "Teaching Classification in the 21st Century." *Knowledge Organization,* 38(4):342-351. https://doi.org/10.5771/0943-7444-2011-4-342

Hudon, Michèle. 2014. "KO and Classification Education In The Light of Benjamin Bloom's Taxonomy of Learning Objectives." In *Knowledge Organization in the 21st Century: Between Historical Patterns and Future Pro-*

spects. Proceeedins of the Thirteenth International Isko Conference. Krakow, Poland, 524-531. Baden-Baden:Ergon-Verlag.

Hussey, Pamela, Das, Sharon Farrell, Lorraine Ledger, and Anne Spencer. 2021. "A Knowledge Graph to Understand Nursing Big Data: Case Example for Guidance." *Journal of Nursing Scholarship* 53(3): 323-332. https://doi.org/10.1111/jnu.12650

Hussey, Pamela and Subhashis Das. 2021. A Micro Credential for Interoperability. *Open Research Europe*, 1(109), 109.

Joudrey, Daniel N., and Ryan McGinnis. 2014. "Graduate Education for Information Organization, Cataloging, and Metadata." *Cataloging & Classification Quarterly* 52(5): 506-550. https://doi.org/10.1080/01639374.2014.911236

Knight, Kevin, I. Chander, M. Haines, V. Hatzivassiloglou, E. H. Hovy, M. Iida, S.K. Luk, R.A. Whitney, and K. Yamada. 1995. Filling Knowledge Gaps in a Broad-Coverage MT System. *Proceedings of the 14th IJCAI Conference*. Montreal, Quebec. Morgan Kaufmann Publishers Inc, Francisco, CA, United States.

Knoblock, Craig A., Pedro Szekely, José Luis Ambite, Aman Goel, Shubham Gupta, Kristina Lerman, Maria Muslea, Mohsen Taheriyan, and Parag Mallick. 2012. "Semi-Automatically Mapping Structured Sources Into The Semantic Web." In *The Semantic Web: Research and Applications: Proceedings of the 9th Extended Semantic Web Conference, ESWC 2012*, edited by Elena Simperl Heraklion, Crete, 375-90. Springer: Berlin Heidelberg.

Laugwitz, B., Theo Held, and Martin Schrepp. 2008. "Construction and Evaluation of a User Experience Questionnaire". In *HCI and Usability for Education and Work. USAB 2008. Lecture Notes in Computer Science,* edited by A. Holzinger, vol. 5298. Springer: Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-89350-9_6

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. International journal of lexicography, 3(4), 235-244.

Noy, N. F., & McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology.

Pattuelli, M. Cristina. 2010. "Knowledge Organization Landscape: A Content Analysis of Introductory Courses." *Journal of Information Science* 36 (6) 812-822. http:// 10.1177/0165551510388118

Peraketh, Benjamin, Christopher P. Menzel, Richard J. Mayer, Florence Fillion, and Michael T. Futrell. 1994. "Ontology Capture Method (IDEF5)". *Air Force Materiel Command*, Wright-Patterson Air Force Base, Ohio. https://apps.dtic.mil/sti/tr/pdf/ADA288442.pdf

Pérez, J., Marcelo Arenas, and Claudio Gutierrez. 2009. "Semantics and Complexity of SPARQL". *ACM Transactions on Database Systems* (TODS) 34(3): 1-45. http://doi.acm.org/10.1145/1567274.1567278.

Ranganathan, S. R. 1967. *Prolegomena to library classification*. Asia Publishing House: New York.

Reich, C., Rimma Belenkaya, Karthik Natarajan, and Clair Blacketer. 2017. *OMOP Common Data Model v5. 2 Specifications*. https://www.ohdsi.org/web/wiki/doku.php?id=documentation:cdm:single-page

Salaba, Athena. 2020. "Knowledge Organization Requirements in LIS Graduate Programs." In *Knowledge Organization at the Interface*, edited by Marianne Lykke et al., 384-393. Baden-Baden: Ergon Verlag.

Satija, M. P., Mayukh Bagchi, and Daniel Martínez-Ávila. 2020. "Metadata Management and Application." *Library Herald* 58(4):84-107. https:// 10.5958/0976-2469.2020.00030.2

Satija, M. P., Amit Tiwari, and Mayukh Bagchi. 2022. "Documentation Research and Training Center (DRTC): A Historical Perspective." *Library Progress International* 42(1):220-30.

Staines, Anthony, Pamela Hussey, and Subhashis Das. 2022. "Terminologies Matter- The Case of ICNP and SNOMED-CT." *European Journal of Public Health* 2(3): ckac129-364.https:// 10.1093/eurpub/ckac129.364

Szostak, Rick. 2014. "The importance of knowledge organization." *Bulletin of the Association for Information Science and Technology* 40(4): 37-42. https://doi.org/10.1002/bult.2014.1720400414

Uschold, Michael and Martin King. 1995. *Towards a Methodology for Building Ontologies*. Edinburgh: Artificial Intelligence Applications Institute, University of Edinburgh.

Appendix

Q1 Describe patient's demographic information.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX shib: <http://www.semanticweb.org/shib>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT ?PatientUPI ?PersonForename ?PersonSurname
?Sex ?EthnicGroup ?CountryOfBirth ?DateOfBirth
        WHERE { ?patient shib:roleOf ?person.
              ?person shib:forename ?PersonForename.
              ?person shib:surname ?PersonSurname.
              ?person
shib:findingRelatedToBiologicalSex ?Sex.
              ?person shib:dateOfBirth ?DateOfBirth.
              ?person shib:countryOfBirth ?Country.
              ?Country rdfs:label ?CountryOfBirth.
              ?patient rdfs:label ?PatientUPI.
```

```
            ?patient shib:ethnicGroup ?EthnicGroup

  }
```

Q2 Describe patient's demographic information along with prescribed drugs.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX shib: <http://www.semanticweb.org/shib>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT ?PatientUPI ?PersonForename ?PersonSurname
?CountryOfBirth ?PrescriptionID ?subtanceName
?ProductName
  WHERE { ?patient shib:roleOf ?person.
      ?person shib:forename ?PersonForename.
      ?person shib:surname ?PersonSurname.
      ?person shib:countryOfBirth ?Country.
      ?Country rdfs:label ?CountryOfBirth.
      ?patient rdfs:label ?PatientUPI.
      ?patient shib:hasPrescription ?Prescription.
      ?Prescription rdfs:label ?PrescriptionID.
      ?Prescription shib:prescriptionDrug
?drugSubstance.
      ?drugSubtance rdf:type shib:drugOrMedicament.
      ?drugSubtance rdfs:label ?subtanceName.
      ?drugSubstance shib:legalDrugProduct
?drugProduct.
      ?drugProduct rdfs:label ?ProductName.

  }
```

Q3 Describe spatial distribution of prescribed quantity.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX shib: <http://www.semanticweb.org/shib>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT DISTINCT  ?CountryOfBirth
?prescribedQuantity
 WHERE {  ?patientVisit shib:ageInYears ?Age.
      ?patientVisit shib:visitOfPatient ?patient.
      ?patient shib:roleOf ?person.
      ?person shib:forename ?PersonForename.
      ?person shib:surname ?PersonSurname.
   OPTIONAL {?person
shib:findingRelatedToBiologicalSex ?Sex.}
      ?person shib:countryOfBirth ?Country.
      ?Country rdfs:label ?CountryOfBirth.
      ?patient rdfs:label ?PatientUPI.
      ?patient shib:hasPrescription ?Prescription.
      ?Prescription rdfs:label ?PrescriptionID.
      ?Prescription shib:prescribedQuantity
?prescribedQuantity.
      ?Prescription shib:prescriptionDrug
?drugSubstance.
      ?drugSubtance rdf:type shib:drugOrMedicament.
      ?drugSubtance rdfs:label ?subtanceName.
      ?drugSubstance shib:legalDrugProduct
?drugProduct.
      ?drugProduct rdfs:label ?ProductName.

}
```

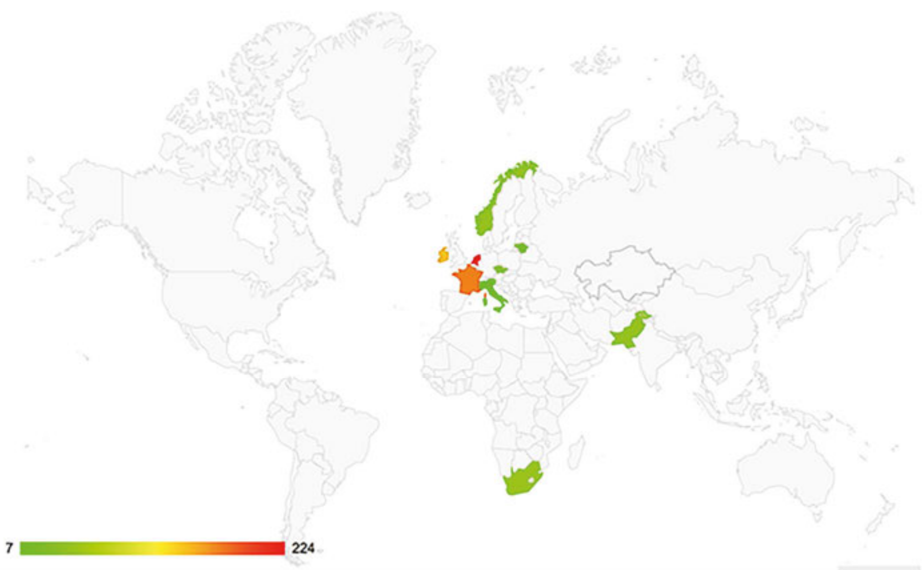

Figure A1: Spatial Distribution of patient with their Prescribe Quantity (Result of Q3)