A Demonstration of 4D Digital Avatar Infrastructure for Access of Complete Patient Information

Project acronym: MyHealthAvatar

Deliverable No. 4.1
Requirements Analysis for Semantic Core Ontology

Grant agreement no: 600929

Dissemination Level		
PU	Public	**X**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

COVER AND CONTROL PAGE OF DOCUMENT	
Project Acronym:	MyHealthAvatar
Project Full Name:	A Demonstration of 4D Digital Avatar Infrastructure for Access of Complete Patient Information
Deliverable No.:	D4.1
Document name:	Requirements analysis for semantic core ontology
Nature (R, P, D, O)*1*	R
Dissemination Level (PU, PP, RE, CO)*2*	PU
Version:	0.3
Actual Submission Date:	31/07/2013
Editor: Institution: E-Mail:	Haridimos Kondylakis FORTH kondylak@ics.forth.gr

ABSTRACT:
This deliverable focuses on the analysis of the requirements for the semantic core ontology that will be developed for the MyHealthAvatar project. The analysis, initially reviews related approaches from currently running research project. Then it specifies the methodology that will be followed and presents the first two phases: i.e. the purpose and scope specification and the knowledge acquisition. In the first phase the description of the work and the use-cases are analyzed to identify the domain of interest whereas in the second phase relevant domains are analyzed and evaluated. In the following months the outcomes of this deliverable will be used to specify the semantic core ontology.

KEYWORD LIST:
Requirement Analysis, Semantic Core Ontology

1 **R**=Report, **P**=Prototype, **D**=Demonstrator, **O**=Other

2 **PU**=Public, **PP**=Restricted to other programme participants (including the Commission Services), **RE**=Restricted to a group specified by the consortium (including the Commission Services), **CO**=Confidential, only for members of the consortium (including the Commission Services)

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 600929.

The author is solely responsible for its content, it does not represent the opinion of the European Community and the Community is not responsible for any use that might be made of data appearing therein.

MODIFICATION CONTROL			
Version	**Date**	**Status**	**Author**
1.0	01/05/2013	TOC	Haridimos Kondylakis
2.0	01/06/2013	First Draft	Haridimos Kondylakis
3.0	31/07/2013	Final	Haridimos Kondylakis, Manolis Spanakis, Manolis Tsiknakis Kostas Marias

List of contributors

- Haridimos Kondylakis, FORTH-ICS
- Manolis Spanakis, FORTH-ICS
- Manolis Tsiknakis, FORTH-ICS
- Kostas Marias, FORTH-ICS

Contents

1 Executive Summary

This deliverable focuses on the analysis of the requirements for the semantic core ontology that will be developed for the MyHealthAvatar project. Initially, similar approaches from research projects are explored to identify guidelines and possible reusable technologies, modules and methodologies. Then the analysis specifies the methodology that will be followed and presents the first two phases, i.e. the purpose and scope specification and the knowledge acquisition. In the first phase the description of the work and the use-cases are analyzed to identify the domain of interest whereas in the second phase relevant domains are analyzed and evaluated. In the following months the outcomes of this deliverable will be used to specify the semantic core ontology.

2 Introduction

The main goal of Semantic Web is to add meaning to data, to allow users to find, share and combine information. It is a vision of information that can be readily interpreted by machines, so that the machines can perform more of the tedious work involved in finding, combining, and acting upon information on the web. Towards this vision, a sheer syntactic level is added to the data, and in addition a semantic value to them. This is done by utilizing semantic resources such as thesauri or ontologies.

Originally introduced by Aristotle, ontologies are formal models about how we perceive a domain of interest and provide a precise, logical account of the intended meaning of terms, data structures and other elements modeling the real world. As such, they are often viewed as the key means through which the Semantic Web vision (Berners-Lee et al. 2001) can be realized and have already found several applications in the area of Knowledge Representation (KR) and in the Semantic Web. Ontologies are so important in the Semantic Web because they provide a means to formally define the basic terms and relations that comprise the vocabulary of a certain domain of interest (Lambrix & Edberg 2003), enabling machines to process information provided by human agents. As a result, ontologies can help in the representation of the content of a web page in a formal manner, so as to be suitable for use by an automated computer agent, crawler, search engine or other web service.

However, if ontologies are developed to support scientific research, particular attention must be paid to both, the reasoning capabilities as well as the correct meaning of the representation. At the same time, the creation of new ontologies must be avoided if there are already ontological resources which can be re-used and integrated for different purposes. Furthermore, their development must be driven by clearly defined-methodology. They should be released to their users only after they have been subjected to a rigorous process of evaluation, in order to assure a high degree of accuracy.

2.1 Purpose of this document

According to the aforementioned claims, the purpose of this deliverable is to establish the methodology for the development of the semantic core ontology that will be used in the MyHealthAvatar project. The first step towards this direction is to identify the purpose and the scope of this ontology and then to reuse existing semantic resources that can model the needed information. The development of the semantic core ontology will be based on the following three principles:

- **Reuse**: Avoid “reinventing the wheel” and reuse already established high quality ontologies.
- **Granularity**: Usually in health-care domain, annotations or mappings cannot be extracted from a single ontological resource. So, multiple ontologies should be used.
- **Modularity**: Create a framework where different ontologies would be able to integrate many modules through mappings between ontologies.

Those principles are already extensively used in similar research projects.

2.2 Structure of this document

The structure of this document is the following: Sections 3 reviews similar approaches from research projects. Then Section 4 defines the methodology and the procedure specification that we will use to define the semantic core ontology. Section 5 describes the first step of this procedure which is the

purpose and scope specification and Section 6 describes and evaluates the relevant knowledge sources relevant to the domain of MyHealthAvatar. Finally, Section 7 concludes this deliverable.

3 Recent Approaches of related Research Projects

In this Chapter we select some currently running EU projects with similar goals in closely related domains and we try to benefit from their experiences. In the end of this Chapter we include one another one approach that seems interesting and relevant to our use-cases.

3.1 p-Medicine

The Health Data Ontology Trunk (HDOT) is the middle-layer ontology being developed by the Institute for Formal Ontology and Medical Information Science (IFOMIS) of the University of Saarland for the p-Medicine EU project[3]. HDOT is published in OWL-DL format online[4] and its goal is to constitute the semantic core of the p-medicine semantic layer. It achieves this goal by specifying how the meaning of the data relevant to the project can be managed and stored in computer format in order to have semantic standards for clinical needs.

HDOT is designed as a middle-layer ontology, with the aim not to create new classes, properties or terms, but rather improving and integrating already existing and well found ontologies in a single ontological resource. Their approach rests on two major principles:

- Several existing biomedical ontologies represent different portions of biomedical reality. Ontologists and scientists providing semantic resources should *avoid to "reinvent the wheel"* as far as possible.
- Clinicians and researchers in the health-care domain usually need to store and manage information, for which semantic content, annotations or mappings cannot be extracted from a single ontological resource. By providing an axiomatic framework for the integration of different ontologies, HDOT delivers a general semantic grid to represent complex and composite biomedical terms and corresponding mid-level classes together with relations defined on them.

[3] http://www.p-medicine.eu/

[4] http://code.google.com/p/hdot/

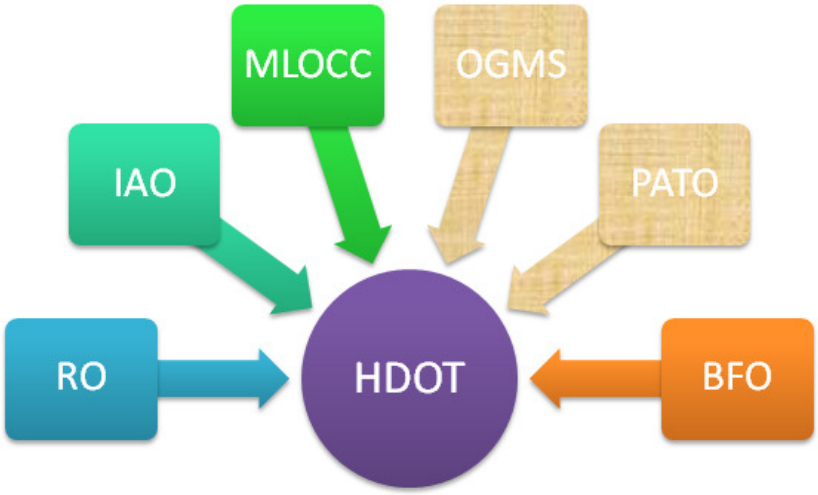


Figure 1: HDOT Structure

In order to provide p-medicine with a high degree of semantic accuracy, the development of HDOT has been driven by three core ideas:

1. reflect the granularity of the biomedical domain;
2. represent such differences through a modular approach, and
3. assure to p-medicine's data a high degree of semantic interoperability.

It integrates under the same "semantic umbrella" the Basic Formal Ontology (BFO), the Relational Ontology (RO), the Information Artifact Ontology (IAO), the Middle Layer Ontology for Clinical Care (MLOCC), part of the Phenotypic Quality Ontology (PATO) and part of the Ontology for General Medical Science (OGMS). All of these ontologies are pre-existing modules of the OBO Foundry and the general idea is shown in Figure 1.

Despite the nice work on ontology design and development the HDOT is still under development and has not been tested in real world scenarios. These two drawbacks make it not ideal choice for the MyHealthAvatar project. However, the principles adopted and the three core ideas give us the directions for our approach and could be re-used in our case as well.

3.2 INTEGRATE

The INTEGRATE project[5] aims to develop innovative infrastructures to enable data and knowledge sharing and to foster large-scale collaboration in biomedical research. Key step to this vision is semantic interoperability. To be able to reuse previous efforts in data sharing, modelling and knowledge generation, and to access relevant external sources of data and knowledge it is beneficial to adhere whenever possible to widely accepted standards and ontologies.

To provide homogeneous access to different data sources, the semantic interoperability layer should provide a Common Information Model (CIM) to represent the information. Thus, a common query endpoint can be provided to retrieve semantically uniform data. The CIM proposed for the

[5] http://www.fp7-integrate.eu/

INTEGRATE platform semantic layer comprises two components: (i) the core dataset and (ii) the Common Data Model (CDM). CDM refers to the schema of the data warehouse and it is based on HL7 RIM. The core dataset is the domain vocabulary of the INTEGRATE platform. Although different candidates were considered for core dataset such as SNOMED CT, LOINC and MedDRA, ultimately SNOMED-CT was selected.

Although SNOMED-CT consists of over 400.000 medical concepts and more than one million relationships cannot be used solely for the MyHealthAvatar. In short, it is too big to have efficient reasoning, there are no formal definitions and does not cover the different domains of MyHealthAvatar.

3.3 Researching Interoperability using Core Reference Datasets and Ontologies for the Virtual Physiological Human (RICORDO)

RICORDO[6] is a FP7 project focused on the study and design of a multi-scale ontological framework in support of the Virtual Physiological Human community to improve the interoperability amongst its data and modelling resources.

Such interoperability can be achieved by using both vocabularies and ontologies, where vocabularies are primarily devoted to support explanation of metadata for humans and ontologies are used to promote automated reasoning and machine processing. The main ontologies used in the RICORDO core are taken from the OBO foundry and they are PATO, OPB, FMA, GO, Cell-type and CHEBI.

However, the RICORDO community is inclined to think that the combination of ontologies is intractable and cannot be readily applied in software systems. So in RICORDO either the data is annotated with a pre-composed term from the RICORDO ontology or the term is post-composed using the RICORDO grammar which is being developed. RICORDO offers a number of potential advantages to clinical data management by performing and maintaining annotation of resources while respecting their integrity and confidentiality constraints, bridging clinical terminologies to ontology-based semantics, supporting semantic integration in the physiology and clinical domains and the semantic interoperability of their data and model resources.

RICORDO is still an on-going project and its results remain to be validated. However, the idea of using and combining different ontologies to achieve interoperability is useful for MyHealthAvatar.

3.4 Virtual Physiological Human Network of Excellence (VPH NoE)

The VPH NoE[7] is a FP7 project which aims to help, support and progress European research in biomedical modelling and simulation of the human body. The goal of is to achieve a more efficient and effective healthcare system and to create new economic opportunities for European healthcare industries.

[6] http://www.ricordo.eu/

[7] http://www.vph-noe.eu/

One of the main objectives is to allow experts to search, classify, integrate and share information about data and model resources based on the biomedical knowledge they represent. The main ontologies that are used are the following

Biological structure:

- The Foundational Model of Anatomy (FMA)
- The Edinburgh Mouse Atlas (EMAP) and the Mouse Anatomical (MA) Dictionary
- The Cell Type Ontology (CL)
- The Gene Ontology Cell Component (GO)
- The Protein Ontology (PRO)
- Chemical Entities of Biological Interest (ChEBI)

Biological processes:

- The GO Biological Process
- The Mammalian Pathology

Qualities:

- The GO Molecular Function
- The Phenotypic Qualities Ontology (PATO)
- The Ontology for Physics in Biology (OPB)

Classes of entities in Experiments, Modeling and Simulation:

- Units Ontology (UO)
- Ontology for Biomedical Investigation (OBI)

VPH NoE is developing a system of mappings between these existing ontologies in order to improve clinicians' ability to predict, diagnose and treat disease, in order to have a real efficient impact on the future of healthcare, the pharmaceutical and medical device industries.

Despite the fact that the target of this program is different than MyHealthAvatar, it shows that interoperability lies in the combination of several pre-existing ontologies and in establishing mappings between them.

3.5 eHealthMonitor

The eHealthMonitor[8] project provides a service-oriented platform used in the process of generating a Personal eHealth Knowledge Space (PeKS) as an aggregation of all knowledge sources relevant for the provision of individualized personal eHealth services. The platform supports end users in two hospital-based scenarios – covering dementia and cardio-vascular domain as well as one prevention-based scenario in the health insurance domain.

The project adopted and extended the Translational Medicine Ontology (TMO) and Translational Medicine Knowledge Base (TMKB) initially produced by the Translational Medicine task force of the World Wide Web Consortium's Health Care and Life Sciences Interest Group (Luciano et al., 2011).

[8] http://www.ehealthmonitor.eu/

TMO has been developed as a unifying ontology that integrates chemical, genomics and proteomic data with disease, treatment, biomedical processes and electronic health records. So TMO is a high level ontology trying to integrate the ontologies included in the TMKB via mappings.

The TMKB consists of the TMO, mappings from TMO to other terminologies and ontologies, and data in RDF format spanning discovery research and drug development, which are of therapeutic relevance to clinical research and clinical practice. The ontologies that are already included in the TMKB and the extensions of the eHealthMonitor project are presented in Figure 2.

The TMO provides a foundation for types declared in Linking Open Drug Data (LODD) and EHRs. It captures a core high-level terminology to bridge existing open domain ontologies and provides a framework to relate and integrate patient-centric data across the knowledge gap from bench to bedside. With the TMO and TMKB, patient and clinical research are bridged and valuable translational knowledge pertinent to clinical practice is developed.

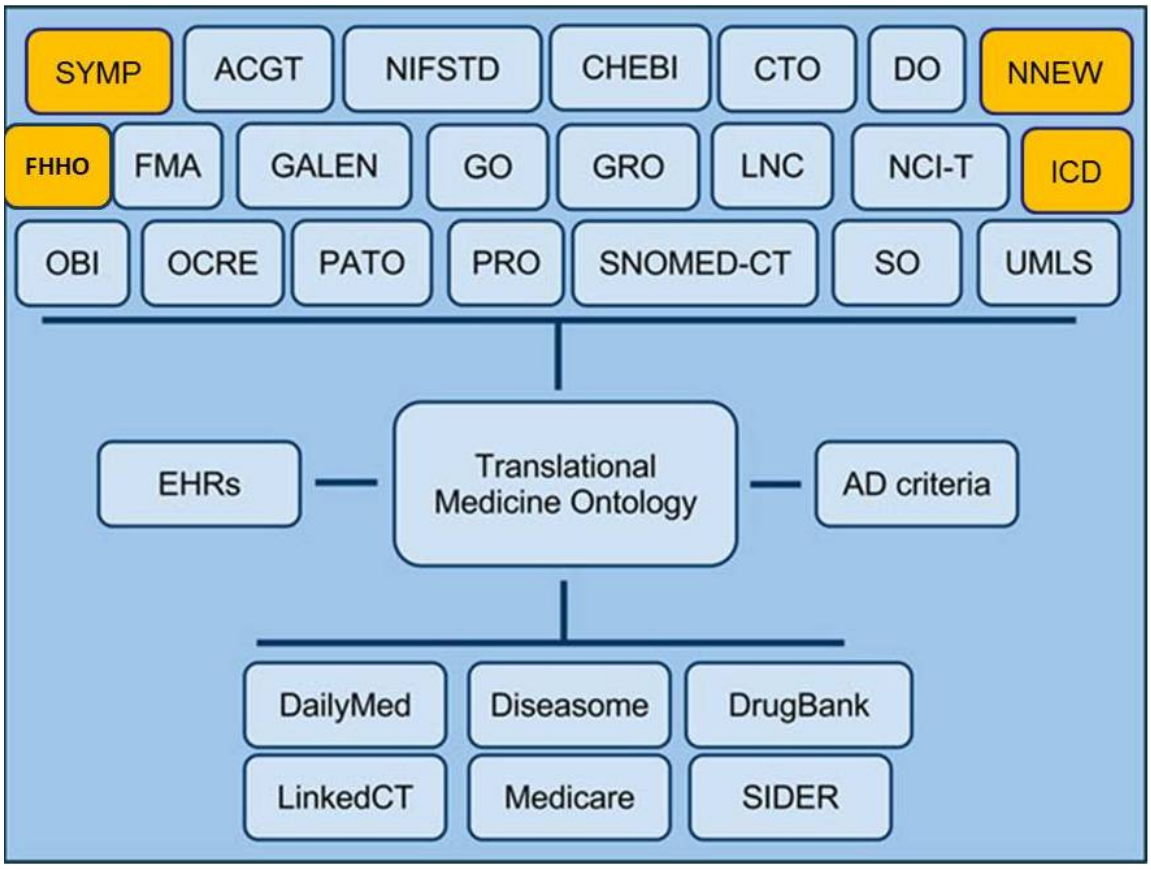


Figure 2. Overview of the Ontologies included in the eHealthMonitor ontology suite

Abbreviations: ***SYMP*** *– Symptom Onology,* ***ACGT****- ACGT Master Ontology,* ***NIFSTD*** *– Neuroscience Information Framework Standardized ontology,* ***CHEBI*** *– Chemical Entities of Biological Interest,* ***CTO*** *– Clinical Trial Ontology,* ***DOID*** *– Human Disease Ontology,* ***NNEW*** *– New Weather Ontology, W****FMA*** *– Foundation Model of Anatomy,* ***FHHO*** *– Family Health History Ontology,* ***Galen*** *– Galen Ontology,* ***GO*** *– Gene Ontology,* ***GRO*** *– Gene Regulation Ontology,* ***LNC*** *– Logical Observation Identifier Names and Codes,* ***MSH****- Medical Subject Headings,* ***NCIt*** *– NCI theraurus,* ***ICD*** *– International Classification of Diseases,* ***NDFRT*** *– National Drug File,* ***OBI*** *– Ontology for Biomedical Investigation,* ***OCRe*** *- Ontology for Clinical Research,* ***PATO*** *– Phenotypic Quality Ontology,* ***PRO***

– Protein Ontology, ***SNOMED CT*** *- SNOMED clinical terms,* ***SO*** *– Sequence Ontology,* ***UMLS*** *– Unified Modeling Language System.*

TMO/TMKB is mature enough with a large community behind it, and is being developed using the principles of 1) modularity, 2) reuse existing ontologies as much as possible, 3) map between existing ontologies.

3.6 Discussion

All approaches described above show that ontologies are increasingly being uses in research projects in the bio-informatics domain. However, ontology construction is deemed to be a labour-intensive and a time-consuming process. Moreover, the development of new ontologies does not necessarily tap the full potential of existing domain-relevant knowledge sources. Due to these problems the latest years the tendency is not to create new ontologies from scratch but to try to integrate high quality, domain-specific ontologies that have already prove their value. The lessons learnt from the above projects is that in MyHealthAvatar we should target for a modular high-level ontology being able to integrate through mappings a set of high quality already existing domain ontologies. The next Chapters describe the methodology for doing so and present an initial selection and evaluation of relevant domain ontologies.

4 Methodology and Procedure Specification

There are currently several methodologies for developing an ontology. Those methodologies give a set of guidelines about how to carry out the activities identified in the ontology development process and what kind of techniques are most appropriate in each activity.

There have been proposed several of them the latest years and the most well-known are the following:

- The state of the art ontology methodology presented in (Staab & Sruder, 2004), (Corcho et al. 2003) and (Fernandez-Lopez, 1999),
- The methodology by Uschold, Gruniger and King (Uschold & Gruninger, 1996), (Uschold, 1996) and (Uschold & King, 1995),
- The TOronto Virtual Enterprise (TOVE) methodology[9] (Gruninger & Fox, 1995)[10],
- The Sensus Methodology [10] (Hovy, 2005),
- The METHONTOLOGY methodology (Fernandez-Lopez et al., 1997),
- The Kactus methodology[11] (Schreiber et al., 1995), (Schreiber et al., 1995),
- The Dolce methodology[12] (Gomez-Perez et al. 1996)

All these approaches have many steps in common. The steps that are described in two of them, namely MENTHONTOLOGY and the one from Uschold, Gruniger and King are presented in Figure 3. The arrows between them show the equivalent phases between these two methodologies. We can see for example that the "Purpose and scope identifications" is the same with the "Build requirements specification document" etc. Independent of the specific methodology selected, the life cycle of an ontology development process is composed of the following iterative processes:

- **Purpose and Scope Specifications:** The goal of this phase is to determine what is expected from the ontology and to define its scope. This includes the set of terms, its distinct characteristics and its granularity. The intended users and what is their purpose has to be determined. This purpose can be identified by listing typical queries that the ontology has to answer or by describing usage scenarios. In our case we will use the initial use-case scenarios to identify the types of data that the platform needs to store and access.
- **Knowledge acquisition:** This phase begins by gathering all available knowledge resources describing the domain of the ontology. These resources can be:
    - **Other Ontologies**
    - **Terminologies:** A terminology is a collection of terms. It is a broad expression which may refer to any collection of terms. In that sense any semantic resource is, generally speaking a terminology.
    - **Controlled vocabularies:** It is a simple collection of terms without any other semantic information.

[9] http://www.eil.utoronto.ca/enterprise-modelling/
[10] http://www.isi.edu/~hovy/
[11] http://hcs.science.uva.nl/projects/NewKACTUS/
[12] http://www.loa.istc.cnr.it/DOLCE.html

- **Coding Systems:** Coding Systems are used when codes, usually code numbers, are applied. This is for example done when a diagnosis is referred to a diagnostic code. Coding systems have all advantages of a controlled vocabulary, which uses natural language terms but they further support interoperability since they do not depend on the use of natural language terms bur they use semantic free identifiers. In that way, they can be internationally used. However, for a coding system to work, documentation and coding keys must be available in different languages.
- **Taxonomies:** The main feature of any taxonomy is a hierarchical structure which is generated by the subsumption (or so called) is-a relation, i.e. the ordering of classes and subclasses. The original and until today most famous taxonomy is the classification of organisms in the Linnaean taxonomy. The term "taxonomy" in information technology may refer to any classification which has the typical structure of the Linnaean taxonomy. It is supposed that every class has only one superclass on the level directly above it. An ontology provides much more and richer relations and connections than a taxonomy.
- **Thesauri:** A thesaurus is a terminological tool that includes at least a controlled vocabulary. Additionally, thesauri provide definitions. They point out synonymy but also broader and narrower terms. Furthermore, it is possible to mark a related term which is no equivalent, sub- or super-term but otherwise semantically dependent from a given term. One famous example for a thesaurus in information technologies is the Art & Architecture Thesaurus.

  A Thesaurus is a much stronger terminological tool than a controlled vocabulary or a coding system. It is also more sophisticated than a taxonomy insofar as more than the subclass relation is representable. Erroneously, thesauri are sometimes called ontologies, but ontologies provide a more expressive syntactic and semantic description of terms than thesauri.
- **Messaging Standards:** The main goal of messaging standards is interoperability. The language and data types defined by such standards determine the way in which an information is transferred. Medical messaging standards are created e.g. by HL7. For example, HL7 v2.x provides six different message types with segments and fields which contain specific determined information: If a patient is admitted to a hospital one segment contains fields with data on the identity of the patient and another one the data containing the case etc. Like an ontology, a messaging standard advances interoperability but unlike an ontology it does not provide resources for automated reasoning.
- **Dataset repositories:** A dataset repository is a catalogue of datasets. In a dataset repository datasets can be identified by a code and named. In that way they are easily accessible. Dataset repositories help to organize data. Their aim is not to represent reality or to produce models like it is done in an ontology. Furthermore, they do not give any semantic explanation of terms.
- **Tools/Algorithms:** tools and algorithms used in the domain might be also a good source of information about the domain.
- **Technical documentations:** Usually the tools and the algorithms in the domain have a technical documentation presenting in details the aforementioned information. This information is usually unstructured text.

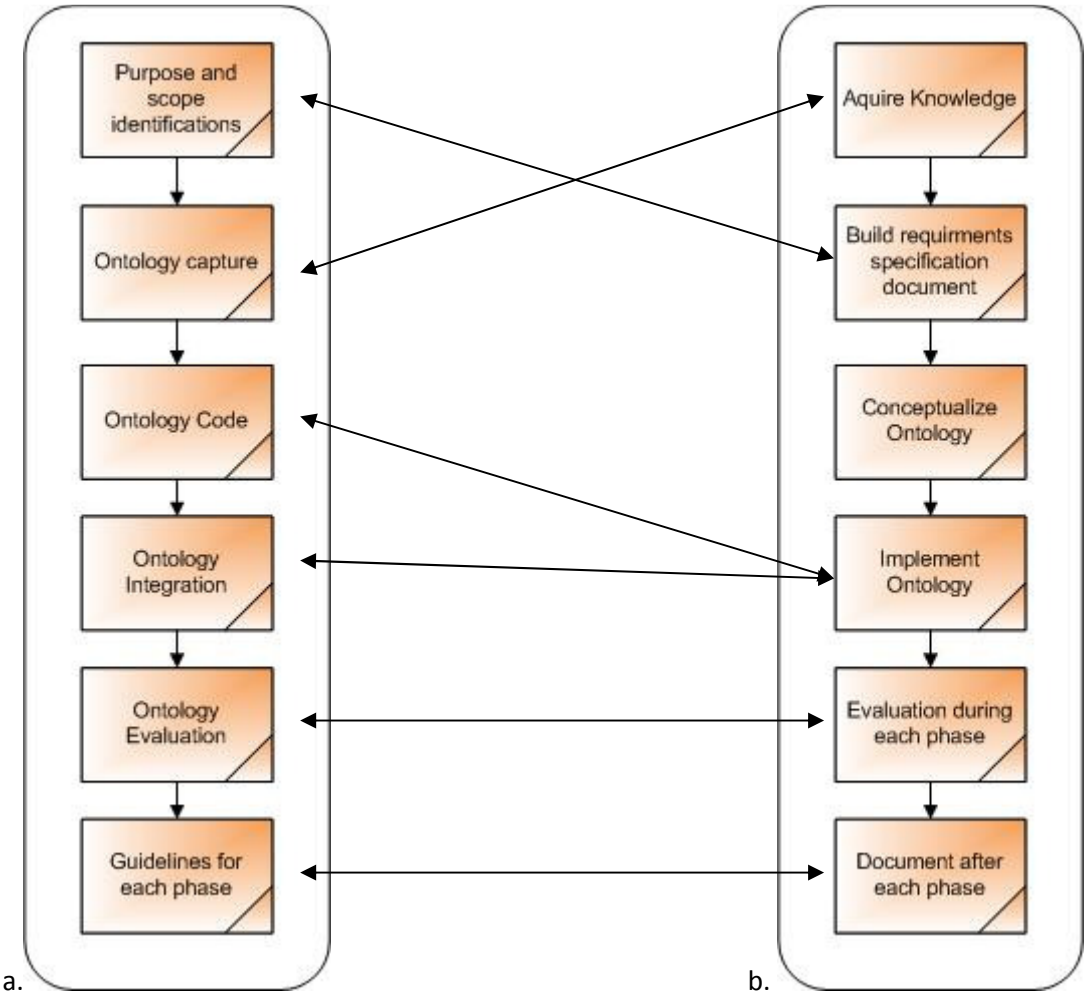


Figure 3. Comparison of two methodologies (a. the approach from Uschold, Gruniger and King and b. Menthontology)

The result of this phase is to identify the most important terms of the domain and define them according to the more consensual definitions

- **Conceptualization:** In this stage, concepts are detected, defined and organized. During this phase a concept is no more a term, but a concept is a definition. Metadata can be added to those concepts to characterize them. In some cases an initial structure
- **Implementation:** The goal of this phase is to build the formal representation of an ontology. Thus the ontology engineer has to choose a language to capture the content of the intermediary representation already built. The next stage is to populate the ontology to build the knowledge base.
- **Evaluation:** This phase evaluates the ontology build according to several metrics. In those metrics the satisfaction of users when testing the ontology might be included, the completeness of the domain representation, or the correctness of the knowledge base and its inference engine.

- **Documentation:** Each choice or problem occurred in the previous phases has to be documented and explained. All the definitions found has to be documented too in order to be precise the source documentation and the authors.

In this deliverable we analyse the first two phases until the conceptualization phase. In the forthcoming months the Semantic Core Ontology will be constructed and implemented. The evaluation and the extensions occurring from this evaluation will be reported in D4.2.

5 Phase 1: Purpose and Scope Specification

As already mentioned the goal of this chapter is to identify what kind of information is expected from the ontology to cover. As a source for this identification we used the description of work and b) a preliminary version of use-cases collected by all partners for WP2 that will be reported in D2.2. Although these use-cases are not yet finalized, and we expect more to be added, they provide a good starting point for collecting the different types of data that will be used within the MyHealthAvatar platform.

5.1 Description of Work

According to the DOW (MyHealthAvatar Consortium, 2013) the target of the project is to create an innovative representation of the health status of citizens for future healthcare. It is "*a unique interface that will allow data access, collection, sharing and analysis by utilizing modern ICT technology, overcoming the shortcomings of the existing resources in Europe, which is highly fragmented*". It has to store and retrieve information about long-term health status of the individual citizen, along with a timeline representing the citizen's life, starting from birth. So, the following types of data should be collected and accessed

- Personal Information
- Lifestyle
- Health Status & Clinical Information
- Medical Information
- Predictive Models
- Molecular Data

The system has to access internal data repositories to store individual data and models for the avatars and external sources such as EHRs and external data and model repositories.

5.2 Use-Cases

This Section provides examples of MyHealthAvatar use cases.

Main Menu specifications

Users should be able to import medical data and to be able to search for patients with similar health issues. Moreover, it specifies that the following data should be available

- Personal Data (gender, birthday etc.)
- Medical Data and Conditions (examinations, hospital admissions etc.)
- Laboratory Values
- Patient Diary, Treatments, Symptoms
- Social Information (followers, friends interests, profile activity)

Users should be able to exam their own data and to find patients with similar conditions, symptoms and treatments.

Enter, import, store and export personal medical data

Personal medical data have to be entered, imported stored and exported. Example such data are conditions, treatments and symptoms, whereas the following system should be accessed

- Personal Health Record Systems
- Hospital Information Systems
- Clinical Trial Management Systems

Moreover, this use case proposes to define and describe a minimum medical dataset compatible and/or similar to Continuity of Care Record (CCR) and the Continuity of Care Document (CCD) formats/standards.

Informed Consent and Privacy

Data about consent management should be stored, entered and accessed. Moreover, the following type of information should be included:

- Biographical information, e.g. photograph, biography, gender, age, location (city, state and country), general notes.
- Condition/disease information, e.g. diagnosis date, first symptom, family history.
- Treatment information, e.g. treatment start dates, stop dates, dosages, side effects, treatment evaluations.
- Symptom information, e.g. severity, duration.
- Primary and secondary outcome scores over time, e.g. ALSFRS-R, MSRS, PDRS, FVC, PFRS, Mood Map, Quality of Life, weight, InstantMe.
- Laboratory results, e.g. viral load, creatinine.
- Genetic information, e.g. information on individual genes and/or entire genetic scans;
- Individual and aggregated survey responses.
- Information shared via free text fields, e.g. the forum, treatment evaluations, surveys, annotations, journals, feeds, adverse event reports.
- Connections to other Avatars.

Interactive 3D Model of the Human Body (Patient Education & Serious Game)

According to this use-case, data about medical images (MRIs, radiographs etc.) should be stored in the platform. In this use-case data from EHR and PHR system should be accessed as well.

Collecting, saving and sharing data from third party social networks

This use focuses on data from third party social networks such as Facebook and Twitter. Data from social networks should be accessible, collected, saved and shared.

Life style monitoring

This focuses on wellness, fitness and prevention information about the most common chronic diseases. Lifestyle and medical data, daily activities and social media will be used in a dual mode allowing the users to insert information about themselves (like they do in common social media technologies) but also will be a mean of supporting personalized services to them from the system in the form of alerts and guidance

Remote patient monitoring (Diabetes, blood sugar level)

The MyHealthAvatar platform should support devices compatible to Continua Health Alliance. Moreover it should be able to track blood sugar levels, and glucose.

Tools for browsing & analysing medical images in avatar

Imaging data should be available in the MyHealthAvatar platform to be browsed and analysed. Data capturing the image analysis should be stored in the platform.

Multi-scale visualization of biomedical data

One of the key challenges for MyHealthAvatar is the interactive visualization of multi-scale biomedical data. The typical data will be a 3D+time dataset of which multiple instances at different scales will have to be displayed together. Information will be on very different spatial and temporal scales going from the molecule up to body level, in different forms (medical images, computer models, signals etc) and of heterogeneous dimensionality (2D, 3D, 3D+t). So the following types of data should be supported by MyHealthAvatar platform:

- Patient history
- Physical examinations
- Laboratory tests
- Computed tomography (CT) data
- Magnetic resonance imaging (MRI) data
- Electroencephalography (EEG) data
- Electrocardiogram (ECG or EKG) data

In addition the results of the following tests should be able to be stored in the MyHealthAvatar platform: Neuropsychological testing, Positron emission tomography (PET) scan, Single photon emission computed tomography (SPECT) scan and Magnetic resonance spectroscopy imaging (MRSI).

Clinical application: 3D virtual lung

This uses a smart-phone application that will interface a breathing classification component, a lung capacity estimation component and a 3D visualization component. All the components will be integrated in order to produce appropriate output including the.

Clinical application: Simulations using a biological models and clinical data

This focuses on biological simulation models and in sets of clinical data available in Clinical Repositories. So information describing models should be stored and the clinical data of the patients should be available as well.

Clinical application: Utilization of personal genomic information for the individualization of MHA platform

Since this use-case has to do with the interpretation and integration of personal genomic information into health medical history record, personal genome (or exome) data should be able to

be stored in the platform. Moreover, access to disease and pharmacogenomic profiling databases should be available to compare the identified genome variations. Moreover, data describing in silico models and the availability of personal medical data are critical dependencies in this scenario.

Clinical application: Anti-platelet therapy in pre-operating period (The example of decision making tool regarding emergency situations in clinical practice)

This case refers in the administration of drugs in emerging situations in clinical level such as pre-operative period and for patients in intensive care units. It represents a typical example of how data can be created through in silico clinical trials approaches especially in clinical cases where clinical trials cannot be performed. It also tries to represent how personalized information regarding drugs, diseases and health status information can be introduced and exploited through MHA in order to create decision making tools and approaches. So the following types of data should be stored:

- Drug data
  - Pharmacokinetic properties
  - Pharmacodynamic properties
- Population data
  - Demographic
  - Genetic
  - Physiology
  - Pathology
- Clinical trials protocols and parameters (as they described in regulatory organizations FDA and EMA)

5.3 Conclusions

The Description of Work and the aforementioned scenarios focus on the following type of information that the MyHealthAvatar platform should allow integrating.

- Personal Information, Lifestyle
- Health Status & Clinical Information
- Medical Information
- Predictive Models
- Molecular Data
- Social Data

In the next Section for each one of these categories we will identify existing ontologies for these types of data and check their applicability in our platform.

6 Phase 2: Knowledge acquisition

The result of this phase is to identify the most important concepts of a selected domain. To do that we collect the most well-known relevant ontologies and terminologies and knowledge sources describing the domain identified in the previous section. Those knowledge resources will be analysed and evaluated.

6.1 Evaluation Methodology

Despite the fact that there are already several ontologies created and maintained by different groups there is no standardized method for performing ontology evaluation. This is for the following reasons

- There is not a single correct way to model a domain
- The knowledge about a single domain changes over time
- Ontologies usually are initially developed outside any real scientific debate

According to Gomez-Perez (Gomez-Perez, 2004) , an ontology evaluation can be thought as a technical judgement of the content of the ontology with respect to a frame of reference during and between each phases of their lifecycle. So, an evaluation is a consideration about the whole content of the ontology's terms, definitions, taxonomy and axioms, in such a way to avoid the spread of mistakes. It should be carried out throughout the entire lifetime of the ontology development process and continuous assessments should occur also after its release.

On the other hand, Hartmann et al. (Hartmann et al., 2005) names three possible stages for the evaluation:

- Evaluating an ontology in its pre-modeling stage: evaluation of the material that the ontology developer has at disposal for building the ontology;
- Evaluating an ontology in its modeling stage: evaluation of the quality of the ontology and use of equivalent ontologies as a reference point;
- Evaluating an ontology after its release: evaluation of the quality of the ontology in specific works and compare it with others different but equivalent modules.

In our case we will evaluate the ontologies after their release. So we are going to look the ontologies from outside. More specifically we are going to examine the following for each ontology:

- Modeling Language: Whether the conceptual model is written in a standard language
- Standardization Body: Whether the conceptual model has been approved by a standardization community
- Open Source: Whether the conceptual model and its documentation is offered free for usage.
- Available Mappings: Relations with other sources
- Human Readable Text: Whether the conceptual model provides human readable text together with computer readable text
- Inter-linguistic operability: Whether the conceptual model covers multi-linguistic frame

- Coherent/Consistent Semantics: “No” when no semantics exists in the model, “Low” when only a few formal definitions/relations are provided, “Medium” for many formal definitions provided, “High” when complete ontologies.
- Interoperability level: “No” when interoperability cannot be reached in the model, “Syntactic” when only syntactic interoperability is provided, “Semantic” when syntactic and semantic interoperability is provided.
- Usage: Whether the conceptual model is being used in other projects.
- Covered MyHealthAvatar domains: It lists the domains that are covered by each semantic resource.

A comparison table will be provided at the end of this Section summarizing the results of our evaluation.

6.2 Domain Ontologies

6.2.1 Symptom Ontology (SO)

The Symptom Ontology (SO) was designed around the guiding concept of a symptom being: "A perceived change in function, sensation or appearance reported by a patient indicative of a disease". The Symptom Ontology[13] captures and documents the semantics of two sets of terms, the term “Sign” and the term “Symptom”. The ontology is open source.

It was developed for part of Gemina project[14] starting in 2005 at the Institute Genome Sciences (IGS) at the University of Maryland. Work ended on the project on 2009. In July 2008 the Symptoms Ontology was submitted for inclusion and review to the OBO Foundry and today the standardization body for the Symptom Ontology is the OBO Foundry. The ontology also provides human readable text together with computer readable format. The available computer readable formats of the ontology are in OBO and OWL. The semantics of the Ontology are coherent, consistent and there is a rigid domain specification. Last but not least the symptom ontology reaches high level of interoperability.

This ontology could be used to cover non-measurable vital parameters of patients for the MyHealthAvatar use cases. Moreover, symptoms and signs can be captured.

6.2.2 Human Disease Ontology

The Disease Ontology[15] (DO) was initially developed as part of the NUgene project[16] starting in 2003 at Northwestern. It is an open source ontology that is designed to link disparate datasets through disease concepts. The aim is to facilitate the connection of genetic data, clinical data, and symptoms through the lens of human disease.

[13] http://symptomontologywiki.igs.umaryland.edu/wiki/index.php/Main_Page
[14] http://gemina.igs.umaryland.edu
[15] http://www.obofoundry.org/cgi-bin/detail.cgi?id=disease_ontology
[16] http://www.nugene.org/

The DO enables the cross-walk between disease concepts, genes contributing to disease, and the 'cloud' of associated symptoms, findings and signs. As such, it can be exploited for modelling multi-scale data in the MyHealthAvatar environment. The understanding of disease and the association of disease with phenotype, environment, and genetics is dynamic and a reflection of current knowledge.

DO is a formally valid, it encapsulates a comprehensive theory of disease, and has a general domain, the health domain. The standardization body of the DO is the OBO Foundry. Terms in DO use standard references such as SNOMED, ICD-10, MeSH, and UMLS. The ontology also provides human readable text together with computer readable format and thus shows syntactic and semantic interoperability. The available computer readable formats of the ontology are in OBO and OWL. The semantics of the Ontology are coherent, consistent and there is a rigid domain specification.

As mentioned before this ontology has a broad domain, the health domain, so it can be used to model general information to model a disease.

6.2.3 The Foundational Model of Anatomy (FMA)

The Foundational Model of Anatomy[17] is a computer-based, open source ontology available for general use. It is created for biomedical informatics and it has to do with the representation of classes, types and relationships necessary for the symbolic representation of the phenotypic structure of the human body in a form that is understandable to humans and is also navigable, parse-able and interpretable to machine-based systems. It is a domain ontology that represents a coherent body of explicit declarative knowledge about human anatomy. It is composed of the following components:

1. **Anatomy taxonomy (At)**: classifies anatomical entities according to the characteristics they share (genus) and by which they can be distinguished from one another (differentia)

2. **Anatomical Structural Abstraction (ASA)**: specifies the part-whole and spatial relationships that exist between the entities represented in **At**;

3. **Anatomical Transformation Abstraction (ATA)**: specifies the morphological transformation of the entities represented in At during prenatal development and the postnatal life cycle;

4. **Metaknowledge (Mk)**: specifies the principles, rules and definitions according to which classes and relationships in the other three components of FMA are represented.

It contains approximately 75,000 classes, over 120,000 terms, over 2.1 million relationship instances and over 168 relationship types. As the FMA should serve as an ontology, its classes are defined in structural terms and grouped into classes on the basis of the structural properties that they share. In such a way, it is possible to aggregate such data in a taxonomy format.

The FMA is the best candidate for serving as a foundation and reference for the correlation of other ontologies in biomedical informatics. Clearly, the FMA is not an application ontology as it is not

[17] http://sig.biostr.washington.edu/projects/fm/AboutFM.html

intended as an end-user application and does not target the needs of particular user groups. Due to the diverse and implied meanings associated with the term "ontology", the FMA is described (Rosse & Mejino 2003) as a symbolic model rather than an ontology.

6.2.4 Ontology of Adverse Events (AEO)

The Adverse Event Ontology (AEO) (He et al., 2011) is a realism-based biomedical ontology for adverse events. Currently AEO has 484 representational units annotated by means of terms including 369 AEO-specific terms and 115 terms from existing feeder-ontologies. In AEO, the term "adverse event" is used exclusively to denote pathological bodily processes that are induced by a medical intervention.

The development of AEO follows the OBO Foundry principles such as openness, collaboration, and use of a common shared syntax. AEO is thus aligned with the Basic Formal Ontology (BFO) and the Relation Ontology (RO). The AEO is up-to-date since the last version release was version 1.1.64 in July 2012. The available computer readable format of the ontology is OWL and it is being used in 2 projects: the OntoCAT[18] project and the OntoMaton[19] project.

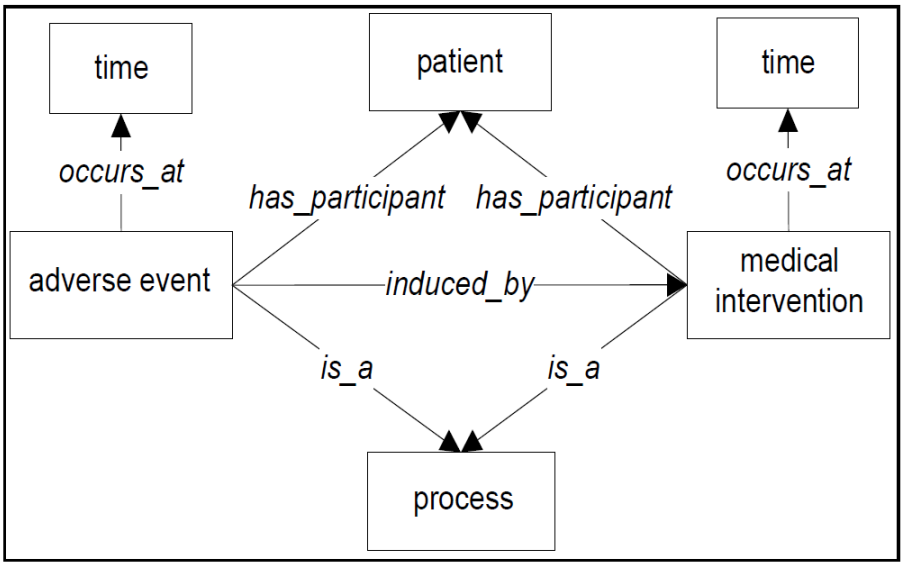


Figure 4: Basic AEO adverse event design pattern.

This ontology can be used to model the adverse events for the medications stored within MyHealthAvatar platform.

6.2.5 Experimental Factor Ontology

The Experimental Factor Ontology (EFO) is an application focused ontology modeling the experimental variables in the Gene Expression Atlas[20]. The ontology describes cross-product classes

[18] http://www.ontocat.org/
[19] http://isa-tools.org/
[20] http://www.ebi.ac.uk/gxa/

from reference ontologies in area such as disease, cell line, cell type and anatomy. EFO combines (Malone et al., 2010) parts of several biological ontologies, such as anatomy, disease and chemical compounds. The scope of EFO is to support the annotation, analysis and visualization of data handled by the EBI Functional Genomics Team. The EFO is application ontology – an ontology engineered for domain specific use or application focus and whose scope is specified through testable use cases and which maps to reference or canonical ontologies. The ontology is kept up-to-date since the last updated version of the EFO ontology is version 2.30 on November 2012.

The EFO methodology reuses reference ontologies (full list available at http://www.ebi.ac.uk/efo/metadata), where they exist, and where they describe classes that are in scope for EFO. To promote interoperability with the OBO Foundry ontologies, EFO is using the BFO as an upper ontology. Furthermore, the EFO Ontology is used in the following projects:

- ISA software suite[21]
- NCBO Annotator[22]
- NCBO Resource Index[23]
- OntoCAT[24]
- MeRy-B[25]
- Neural ElectroMagnetic Ontologies[26]
- OntoMaton [27]

6.2.6 Clinical Care Classification System (CCC)

The Clinical Care Classification System[28],[29] is a standardized framework and a coding structure for assessing, documenting, and classifying patient care by nurses and other clinical professionals in any health care setting. The CCC system consists of two interrelated terminologies, the CCC of Nursing Diagnoses and the CCC of Nursing Interventions/Actions. The two terminologies are both classified by 21 Care Components that represent the Functional, Health Behavioural, Physiological, and Psychological Patterns of Patient Care (Table 1: CCC Care Components).

[21] http://isa-tools.org/
[22] http://www.bioontology.org/annotator-service
[23] http://www.bioontology.org/resources-index
[24] http://www.ontocat.org/
[25] http://services.cbib.u-bordeaux2.fr/MERYB/
[26] http://nemo.nic.uoregon.edu/wiki/NEMO
[27] http://isa-tools.org/
[28] http://en.wikipedia.org/wiki/Clinical_Care_Classification_System
[29] http://www.sabacare.com/

Table 1: CCC Care Components[30]

Care Components		
Activity	Medication	Self-Care
Bowel/Gastric	Metabolic	Self-Concept
Cardiac	Nutritional	Sensory
Cognitive/Neuro	Physical Regulation	Skin Integrity
Coping	Respiratory	Tissue Perfusion
Fluid Volume	Role Relationship	Urinary Elimination
Health Behavior	Safety	Life Cycle

The CCC System is being used to document nursing care in the electronic health record (EHR) computer-based patient record (CPR) and Personal Health Record (PHR) Systems. It serves as a language for nursing and other health care providers such as physical, occupational, and speech therapists, medical social workers, etc. The CCC System is used to:

- Document integrated patient care processes
- Classify and track clinical care
- Develop evidence-based practice models
- Analyze patient profiles and populations
- Predict care needs, resources, and costs

In 2007, the CCC was accepted by the US Department of Health and Human Services[31] as the first national nursing terminology. The coding structures for the terminologies are based on the ICD-10 consisting of five alphanumeric characters for information exchange among health care terminologies promoting interoperability. They are used to track and measure patient/client care holistically over time, across settings, population groups, and geographic locations. The CCC has open architecture and is specially designed for computer-based systems – EHR, CIS and PHR. Furthermore, CCC was tested as an international nursing standard based on the An Integrated Reference Terminology Model for Nursing, approved by the International Organization for Standardization (ISO/TC 215: Health Informatics) in October 2003. The computable structure of the CCC is protected under copyright permission.

6.2.7 American Medical Association's Current Procedural Terminology Codes (AMA CPT)

The Current Procedural Terminology[32] (CPT®) is a medical nomenclature used to report medical procedures and services under public and private health insurance programs. It describes medical,

[30] http://www.sabacare.com/About

[31] http://www.hhs.gov/

[32] http://www.ama-assn.org/ama/pub/physician-resources/solutions-managing-your-practice/coding-billing-insurance/cpt.page

surgical, and diagnostic services and is designed to communicate uniform information about medical services and procedures among physicians, coders, patients, accreditation organizations, and payers for administrative, financial, and analytical purposes.

There are three types of CPT codes: Category I, Category II, and Category III. There are six main sections for Category I:

- Codes for Evaluation and Management
- Codes for Anaesthesia
- Codes for Surgery
- Codes for Radiology
- Codes for Pathology & Laboratory
- Codes for Medicine

Category II and III contain optional Codes for Performance Measurement and Emerging Technology respectively. Last, but not least it is necessary for users of the CPT code to pay license fees for access to the code.

6.2.8 Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT)

The Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) is a clinical terminology, which has been promoted as a reference terminology for electronic health record (EHR) systems. Its purpose is to serve as a standardized terminology in healthcare software applications, as it enables clinicians, researchers and patients to share comparable data. SNOMED CT is owned, maintained and distributed by the International Health Terminology Standard Development Organization [33]. It is open source and its current version was released in July 2011. SNOMED CT is used by the College of American Pathologists[34], the UMLS Metathesaurus[35], the European project epSOS[36] and the European project SemanticHealthNet[37]. SNOMED CT is designed to support translation. This multi-lingual resource is used in more than 50 countries. Available mapping to SNOMED CT exist with ICD-9-CM, ICD-03 and ICD-10.

SNOMED CT is the result of the combination of SNOMED Reference Terminology (SNOMED RT), developed by the College of American Pathologist, with the Clinical Terms Version 3 (CTV3), developed by the National Health Service of the United Kingdom. It consists of concepts, descriptions and relationships between concepts:

Concepts

- SNOMED CT *concepts* represent clinical ideas, ranging from abscess to zygote.
- Every concept has a unique numeric code known as the "concept identifier".

[33] http://www.ihtsdo.org/about-ihtsdo/
[34] http://www.cap.org/apps/cap.portal
[35] http://www.nlm.nih.gov/pubs/factsheets/umlsmeta.html
[36] http://www.epsos.eu/
[37] http://www.semantichealthnet.eu/

- Concepts are organized in hierarchies, from the general to the specific. This allows detailed clinical data to be recorded and later accessed or aggregated at a more general level.

Descriptions

- SNOMED CT *descriptions* link appropriate human-readable terms to *concepts*. A concept can have several associated *descriptions*, each representing a synonym that describes the same clinical idea.
- Each translation of SNOMED CT includes an additional set of *descriptions*, which link terms in another language to the same SNOMED CT *concepts*.

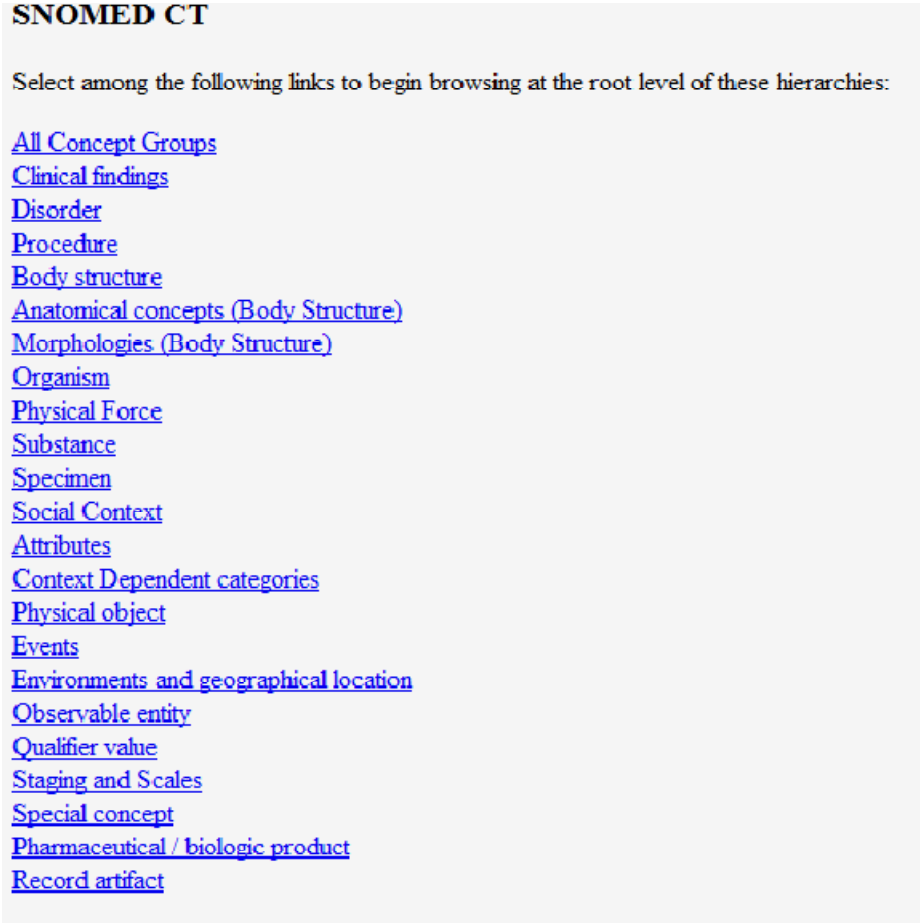


Figure 5: SNOMED CT hierarchies

Relationships

- SNOMED CT *relationships* link each *concept* to other *concepts* that have a related meaning. These *relationships* provide formal definitions and other characteristics of the *concept*.
- One type of link is the "is a" *relationship* which relates a *concept* to its more general *concepts*. For example, the *concept* "viral pneumonia" has an "is a" *relationship* to

the more general *concept* "pneumonia". These "is a" *relationships* define the hierarchy of SNOMED CT *concepts*.

- Other types of *relationship* represent other aspects of the definition of a *concept*. For example, the *concept* "viral pneumonia" has a "causative agent" *relationship* to the *concept* "virus" and a "finding site" *relationship* to the *concept* "lung".
- There are well over a million *relationships* in SNOMED CT.

The amount of data stored in SNOMED CT is organized into top-level hierarchies (Figure 5) and as they are descendent, the concepts within them become increasingly specific.

Although SNOMED is widely used it is also criticized as having a vague domain (SemanticHEALTH Report, 2009). However, it is widely used and this is the reason that will be integrated also in the core ontology of the MyHealthAvatar platform.

6.2.9 Gene Ontology

The Gene Ontology (GO) project is a collaborative effort to develop and use ontologies to support biologically meaningful annotation of genes and their products in different databases. At the beginning, the project began as collaboration between three model organism databases, the FlyBase (Drosophila), the Saccharomyces Genome Database (SGD) and the Mouse Genome Database (MGD). To date, the GO Consortium includes several databases[38].

The GO Consortium is formed by both domain experts and knowledge engineers, which aim to develop and maintain the ontologies and ontology tools to be used for collaboration between GO databases. It is worth noting that GO is neither a database of gene sequences nor a catalogue of gene products; rather it describes how gene products behave in a cellular context.

In its first appearance the GO was reported to have several problems (e.g. an erroneous treatment of formal relations between classes in the ontology; (Smith et al., 2004)). The ontology, then, has been subjected to a series of reforms designed according to the realistic approach of the OBO Foundry[39].

The GO covers three domains in order to represent gene product properties and contains 35038 terms (ontology version 1.2297, dated 03.10.2011):

1. Cellular component: the parts of a cell or its extracellular environment. Such parts can be anatomical structures (e.g. rough endoplasmic reticulum or nucleus) or a gene product group (e.g. ribosome, proteasome). The cellular component domain has 2900 terms;
2. Biological process: operations or sets of molecular events with a defined beginning and end, pertinent to the functioning of integrated living units: cells, tissue, organs and organisms. For example, cellular physiological process, signal transduction, pyramid metabolic process. The biological process domain has 21443 terms;
3. Molecular function: This describes activities that occur at the molecular level. Such functions correspond to activities performed by individual and complex gene products. For instance,

[38] http://www.geneontology.org/GO.consortiumlist.shtml

[39] http://www.obofoundry.org/

catalytic activity, transporter activity, adenylate cyclase activity. The molecular function domain has 9109 terms.

The GO does not cover:

- gene products;
- processes, functions and components related to mutants or diseases;
- attributes of sequence;
- protein domain or structural features;
- protein-protein interactions;
- environment, evolution and expression;
- anatomical or histological features above the level of cellular component, including cell types.

The GO is one of the most successful bioinformatics ontology projects; it has been embraced by the scientific community, and it is being used in a multitude of projects which makes it ideal to be used in MyHealthAvatar as well.

6.2.10 Anatomical Therapeutic Chemical Classification System (ATC/DDD)

The ATC/DDD[40] system is an instrument for presenting active ingredient utilization statistics with the aim of improving drug use. The system is suitable for international comparisons of active ingredient utilization, for the evaluation of long term trends in drug use, for assessing the impact of certain events on drug use and for providing denominator data in investigations of drug safety.

- ATC - Anatomical Therapeutic Chemical (ATC) classification system
  Within the ATC system active substances are divided into different groups according to the organ or system on which they act and their therapeutic, pharmacological and chemical properties. The drugs are classified in groups (five different levels).
- DDD - Defined Daily Dose
  'The DDD is the assumed average maintenance dose per day for a drug used for its main indication in adults.'

  The DDD will only be assigned for drugs that already have an ATC code.

Latest versions are obtainable at http://wido.de/amtl_atc-code.html in the file format of Excel (xls). Linkage to commercial drugs with brand or generic names can be realized using mapping tables.

The classification does not fulfil criteria for a semantically meaningful classification since the ATC classification is rather vague (Group V for example contains various entities i.e. allergens for hypersensitisation but also surgical dressings). Furthermore the DDD is actually given without consideration of personal background (age, gender, etc.). The purpose of this classifying is simply to determine a letter for a code. Hence, ATC with DDDs is best described as a coding system. So, it

[40] http://www.dimdi.de/static/de/klassi/atcddd/index.htm

cannot be taken into consideration as a resource of knowledge representation or a foundation for automated reasoning.

6.2.11 UMLS

UMLS[41], the Unified Medical Language System, is a unifying framework which integrates different terminologies which are relevant to medicine and biomedical information technologies. It consists mainly of three parts. The Metathesaurus and the Semantic Network are the most important ones. The third part, the SPECIALIST lexicon, is a source of lexical information and language processing programs.

The Metathesaurus is currently distributed in two versions, the Rich Release Format (RRF) is provided since 2004. The Original Release Format (ORF) is older. Since RRF is more accurate and precise than ORF it is the preferable option. The Metathesaurus is the core of UMLS. With over five million names for over one million concepts and about 12 million relations between these concepts it is a very broad scoped but also detailed resource for the domain of biomedicine. The purpose of the Metathesaurus is not to give a new terminology but to give an extensive lexicon of existing vocabularies and coding systems. According to a ranking of source vocabularies one of the different terms which belong to the same concept is designated as a preferred term. Whatever is contained in the Metathesaurus has a unique identifier. For example, concepts are attached to a CUI (Concept Unique Identifier), terms get a LUI (Lexical Unique Identifier) and relationships are named by a RUI (Relationship Unique Identifier).

The second part of UMLS is the Semantic Network. Its aim is “to provide a consistent categorization of all concepts represented in the UMLS”. The network is a system of abstract categories and provides the foundation for the categorization of the concepts in the Metathesaurus. Every concept in the Metathesaurus is associated to at least one of the categories, usually to the most specific available category. Currently, the Semantic Network has 133 broad categories and 54 relationships. The *is_a* relation, i.e. subsumption, is essential for the hierarchical structure. The Semantic Network does not aim to be a complete characterization of the world but it is rather limited to medical purposes. This becomes obvious with respect to granularity. Narrow classes are only provided for the domain of biomedicine. Further relation are "physically related to", "spatially related to", "temporally related to", "functionally related to", "conceptually related to" and relation which are subtypes of these five relations. Relations between entities are usually inherited to the terms which are subsumed.

The UMLS has been under development by the US National Library of Medicine (NLM) since the eighties. As an integrating framework its goal is to unite the knowledge expressed in currently over 100 source terminologies for diseases, procedures, supplies and diagnoses, including for example the ICD terminologies and SNOMED, and, thereby, to support interoperability. All parts of UMLS are machine readable. Using UMLS is free of charge but a license agreement is necessary. The UMLS is a global and comprehensive source for manifold medical terminologies and it is hardly possible to ignore it when working on interoperability.

[41] http://www.nlm.nih.gov/research/umls/

6.2.12 MeSH

The Medical Subject Headings (MeSH)[42] are a medical thesaurus published and annually updated by the US National Library of Medicine (NLM). It is used for cataloging of the library holdings and for indexing of the databases that are produced by the NLM (e.g. MEDLINE).

It consists of sets of terms naming descriptors in a hierarchical structure that permits searching at various levels of specificity. MeSH descriptors are arranged in both an alphabetic and a hierarchical structure. At the most general levels of the hierarchical structure are very broad headings such as "Anatomy" or "Mental Disorders." More specific headings are found at more narrow levels of the twelve-level hierarchy, such as "Ankle" and "Conduct Disorder." There are 26,853 descriptors in 2013 MeSH. There are also over 199,000 entry terms that assist in finding the most appropriate MeSH Heading, for example, "Vitamin C" is an entry term to "Ascorbic Acid." In addition to these headings, there are more than 205,000 headings called Supplementary Concept Records (formerly Supplementary Chemical Records) within a separate thesaurus.

The MeSH thesaurus is used by NLM for indexing articles from 5,400 of the world's leading biomedical journals for the MEDLINE®/PubMED® database. It is also used for the NLM-produced database that includes cataloging of books, documents, and audiovisuals acquired by the Library. Each bibliographic reference is associated with a set of MeSH terms that describe the content of the item. Similarly, search queries use MeSH vocabulary to find items on a desired topic.

6.2.13 International Classification of Functioning, Disability and Health (ICF)

The International Classification of Functioning, Disability and Health, known more commonly as ICF, is a classification of health and health-related domains. These domains are classified from body, individual and societal perspectives by means of two lists: a list of body functions and structure, and a list of domains of activity and participation. Since an individual's functioning and disability occurs in a context, the ICF also includes a list of environmental factors. ICF is a WHO framework to measure health and disability at both individual and population levels.

ICF puts the notions of ‘health’ and ‘disability’ in a common understanding in acknowledging that every human being may experience a decrement in health and thereby experience some degree of disability. By shifting the focus from cause to impact it places all health conditions on an equal footing allowing them to be compared using a common metric – the ruler of health and disability. Furthermore ICF takes into account the social aspects of disability and does not see disability only as a 'medical' or 'biological' dysfunction. By including Contextual Factors, in which environmental factors are listed ICF allows to records the impact of the environment on the person's functioning[43].

However, the whole model suffers from shortcomings (Kumar, 2010). The classification is not coherent, as the criteria are sometimes based on the anatomic structure which has a function and sometimes on the process which is supported by the function. Another critical point is the

[42] http://www.ncbi.nlm.nih.gov/mesh

[43] http://www.who.int/classifications/icf/en/

"overemphasis on subsumptions", i.e. the restriction to the *is-a* relation. Though the categories from ICF are useful, one should put more effort in the definition the relations which hold between them and add more ontological power and expressivity.

ICF can be used online[44] or file-based[45] .

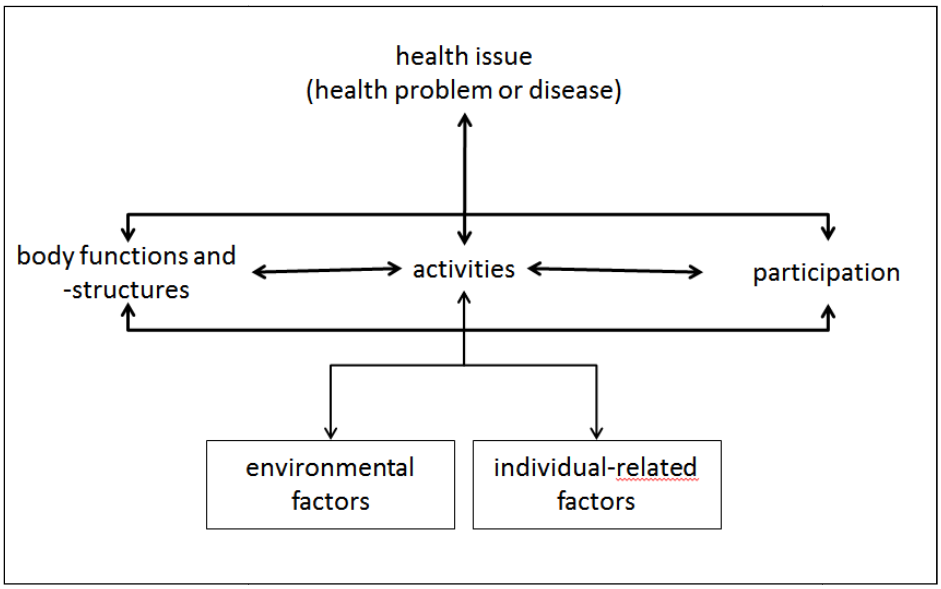


Figure 6: ICF as classification of components for health issues

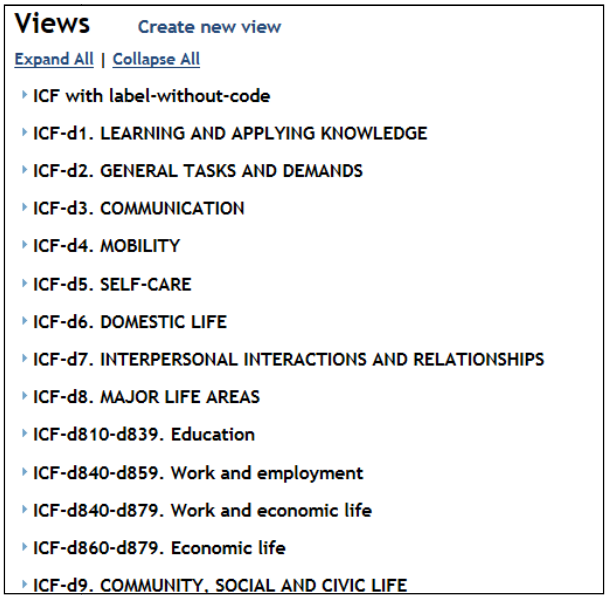


Figure 7: View of ICF components

[44] http://apps.who.int/classifications/icfbrowser/

[45] http://bioportal.bioontology.org/ontologies/40865

6.2.14 Ontology of medically related Social Entities

This ontology covers the domain of social entities that are related to health care and user scenarios in eHealthMonitor, such as demographic information (social entities for recording gender (but not sex) and marital status, for example) and the roles of various individuals and organizations (patient, caregiver, hospital, etc.) A subset of this ontology may be helpful to implement shared decision making between end-users and automated reasoning.

The last version of the Ontology was released in June 2012 and it consists of 119 classes and only 6 ObjectProperties. It is a rather small ontology with little semantic impact. It is available in OWL format[46].

6.2.15 Neuroscience Information Framework Standardized Ontology (NIFSTD)

The Neuroscience Information Framework Project (NIF)[47] according to (Imam, 2012) has been developing tools and strategies for creating resources that can be integrated across neuroscience domains. The end product is a semantic search engine and a knowledge discovery portal for describing neuroscience resources and provides access to multiple types of information organized by relevant categories. Through its resource catalog and data federation, NIF represents the source of neuroscience information available on the web.

The semantic framework through which these diverse resources are accessed is provided by the NIF Standardized Ontologies (NIFSTD) (Bug, 2008). NIFSTD represents a collection of terms and concepts from the domains of neuroscience.

The NIFSTD ontologies are built in a modular fashion, where each module covers a distinct, orthogonal domain of neuroscience (Bug, 2008). Modules covered in NIFSTD include anatomy, cell types, experimental techniques, nervous system function, small molecules, and so forth. The upper-level classes in NIFSTD modules are carefully normalized under the classes of Basic Formal Ontology (BFO)[48]. The Ontology is open source available online[49].

6.2.16 Biocaster Ontology (BCO)

The BioCaster Ontology (BCO) (Collier, 2010) aims to (a) describe the terms and relations necessary to detect and risk assess public health events in the grey literature at an early stage; (b) bridge the gap between the (multilingual) grey literature and existing standards in biomedicine; (c) to be open source and freely available for general usage.

In contrast to other ontologies that describe infectious diseases, the BCO focuses on the usage of terms and relations within informal unstructured reports which are often made at a pre-diagnostic stage of a disease outbreak by non-medically trained reporters. This is done to provide monitoring

[46] http://omrse.googlecode.com/svn/trunk/omrse/omrse.owl
[47] http://neuinfo.org
[48] http://www.ifomis.org/bfo/
[49] http://bioportal.bioontology.org/ontologies/40510

and early warning about public health hazards from online media reports. An example of its usage can be seen in the BioCaster Global Health Monitor[50].

The BCO is maintained by Dr. Nigel Collier's group at the National Institute of Informatics in Tokyo with the collaboration of partners in the international life science and computational linguistics communities. This ontology was developed to cover the need of a specific project, the Biocaster project[51] and thus has a very broad and vague domain. It contains models for the analysis of Internet news and research literature for public health workers, clinicians and researchers interested in communicable diseases. So it does not exactly conform to our needs in MyHealthAvatar uses cases.

6.2.17 Family Health History Ontology (FHHO)

The FHHO (Peace & Brennan, 2007) is representing the family health histories of persons related by biological and/or social family relationships (e.g. step, adoptive) who share genetic, behavioral, and/or environmental risk factors for disease. Projects that are linked with this Ontology, as well as other ontologies mapped to FHHO can be found online[52].

This ontology is very close related to the public health use case since the knowledge of Family health history is a way to prevent possible illnesses, and thus it is necessary to model this kind of data in the eHealthMonitor project.

6.2.18 Advancing Clinico-Genomic Trials Master Ontology (ACGT MO)

Advancing Clinico-Genomic Trials on Cancer (ACGT)[53] was a project financed by the European Union within the 6th Framework Program, which aimed at enabling the rapid sharing of data gained in both clinical trials and associated genomic studies. In order to meet such a goal, ACGT provided a grid-based infrastructure, designed to transmit the data between different groups of users in real time according to their needs with data integration being achieved by means of an ontology-based mediator. The system had been designed to enable the smooth and prompt transfer of laboratory findings to the clinical management and treatment of patients.

The ACGT consortium developed its own Master Ontology (MO) in order to address the goal of data integration for the domains of clinical studies, genomic research and clinical cancer management and care (Brochhausen et al., 2011). The MO has been grounded on the Basic Formal Ontology (BFO), which is the Open Biomedical Ontologies (OBO) Foundry's upper level ontology. BFO assured to MO's classes a high level of semantic specification.

The ACGT MO was the core of the ACGT Semantic Mediation Layer (ACGT-SM) which comprised a set of tools and resources working together to serve processes of Database Integration and Semantic Mediation. The ACGT-SM followed a Local-as-View Query Translation approach in order to cope with the problem of database integration. In such a way, the data is not actually integrated but it is made accessible to users via a virtual repository. This repository represents the integration of the

[50] http://born.nii.ac.jp

[51] http://ontolog.cim3.net/cgi-bin/wiki.pl?action=browse&id=BioCaster&revision=1

[52] http://bioportal.bioontology.org/ontologies/1126

[53] www.eu-acgt.org/

underlying databases and ACGT-MO acts as database schema, providing resources for formulation of possible queries.

The MO was constructed in modular fashion with Clinical Trial and Patient Management Ontology modules designed to be reused for different clinical domains. As such, the Patient Management modules could be used as relevant to the eHealthMonitor. However, although this ontology is already a well-established ontology it has not been widely used.

6.2.19 Glossary of Terms for Community Health Care and Services for Older Persons

The Ageing and Health Program of the WHO[54] Kobe Centre, in collaboration with the WHO Collaborating Centre for Population Ageing: Research, Education and Policy in Adelaide, Australia, initiated a project to develop an international glossary of terms applying to community health care and services for older persons through consultation with global experts, both via the Internet and in face-to-face meetings.

It aims to define and standardize the basic concepts and functions of community health care for older persons and organize them into a glossary[55], utilizing existing WHO definitions where appropriate, promote a common language for cross-program description and information dissemination.

The limitation of this glossary is that it just focuses on older persons, while the MyHealthAvatar project does not focuses on a specific age group. Furthermore and the semantic of this glossary are very poor.

6.2.20 The Weather Ontology- NNEW

The Nextgen Network Enabled Weather (NNEW)[56] ontology is a meta-model that represents areas of weather domain. The ontology can be used to describe and reason about entities within the domain. It was built upon a number of relevant ontologies and taxonomies shown in Figure 8. Most of these modules correspond to individual subdomains such as humidity or pressure. They are complete ontologies on their own and when combined make up the complete NNEW ontology. The package also includes units of measure ontology which is utilized by the other components. The merged ontology uses 514 unique terms (concepts and properties).

[54] http://www.who.int/

[55] http://whqlibdoc.who.int/wkc/2004/WHO_WKC_Tech.Ser._04.2.pdf

[56] https://wiki.ucar.edu/display/NNEWD/Data+Models+and+Formats

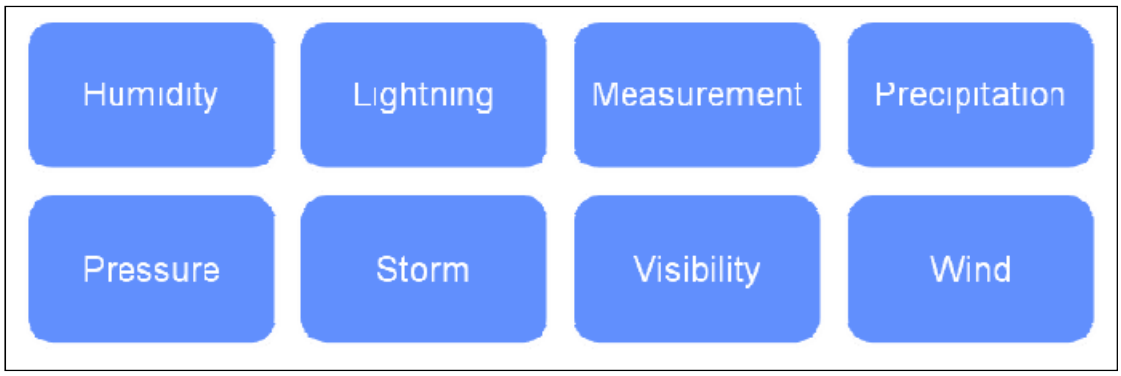


Figure 8: Modules of the NNEW ontology

Since we expect that environmental factors will be included in the MyHealthAvatar data this ontology could be used to model these data.

6.2.21 Systems Biology Ontology (SBO)

The Systems Biology Ontology[57] is a collaborative effort led my Biomodels.Net and it is a set of controlled, relational vocabularies of terms commonly uses in Systems Biology, and in particular in computational modelling. There are several orthogonal vocabularies in the ontology defining the following:

- Reaction participants roles (e.g. substrate)
- Quantitative parameters (e.g. pi)
- Classification of mathematical expressions describing the system (e.g. mass action rate law)
- Modelling framework used (e.g. logical framework)
- The nature of the entity (e.g. molecule)
- The type of the interaction (e.g. process)
- The different types of metadata present in a model

The ontology is defined in various formats such as OBO, OWL and XML. Moreover, to allow programmatic access to the resources Web Services have been implemented and libraries, documentation and samples as well.

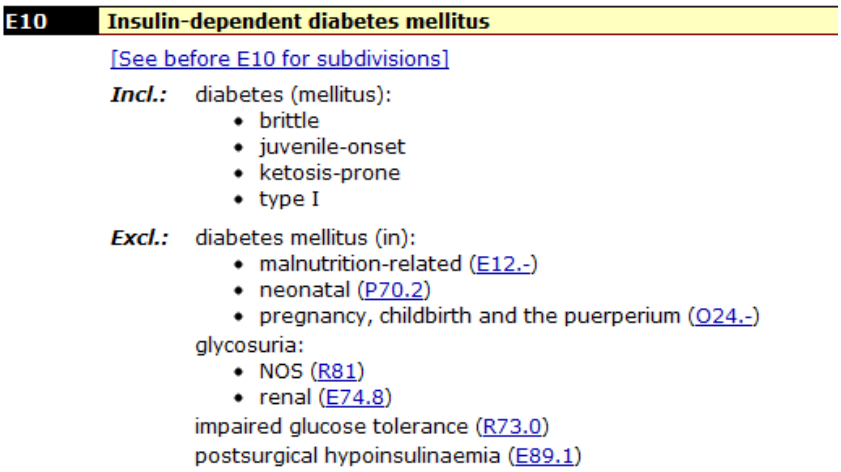
6.2.22 ICD-10

The International Classification of Diseases[58] is the world's standard tool to capture mortality and morbidity data. Generally speaking, the ICD contains information related to diagnoses, symptoms, abnormal laboratory findings, injuries and poisonings, external causes of morbidity and mortality, factors influencing health status from all the different branches of medicine: Oncology, Dentistry and Stomatology, Dermatology, Psychiatry, Neurology and so on.

[57] http://www.ebi.ac.uk/sbo/main/

[58] http://www.who.int/classifications/icd/en/

The basic ICD-10 is a single coded list of three-character categories (from A00 to Z99), each of which can be further divided into up to ten four-character subcategories (for example, A00.0, A02.2, B51.9 and so on). Thus, a disease is related to three main data: namely Chapter, Block and Title. An example concept is shown in Figure 3.



E10 Insulin-dependent diabetes mellitus

[See before E10 for subdivisions]

Incl.: diabetes (mellitus):
- brittle
- juvenile-onset
- ketosis-prone
- type I

Excl.: diabetes mellitus (in):
- malnutrition-related (E12.-)
- neonatal (P70.2)
- pregnancy, childbirth and the puerperium (O24.-)

glycosuria:
- NOS (R81)
- renal (E74.8)

impaired glucose tolerance (R73.0)

postsurgical hypoinsulinaemia (E89.1)

Figure 9. ICD-10 Concept E10

ICD-10 treats diseases as health problems which have been recorded, for example, on health records, or death certificates. However, there are several cases where the ontology is not coherent with wrong human labels.

6.2.23 Logical Observation Identifiers Names and Codes (LOINC)

LOINC[59] is a database and a universal standard for identifying medical laboratory and clinical observations. It was developed and maintained by the Regenstrief Institute. The LOINC vocabulary provides a set of universal names and ID codes for identifying laboratory and clinical test results in the context of existing HL7, ASTM E1238, and CEN TC251 observation report messages. The LOINC codes are mainly intended to identify test results and clinical observation. Other fields in the LOINC message can transmit, for example, the identity of the source laboratory or other special details about the sample. A formal, distinct and unique name (composed by six parts) is given to each LOINC component term.

The LOINC codes were released in April 1996 and, to date, thirteen revisions of LOINC, now including over 30,000 observation concepts, were released. LOINC contains fields for each of the six parts of the name, synonyms and comments for all observations in order to facilitate searches for individual laboratory test and clinical observation results. The database is dived into four categories: Lab, Clinical, Attachments, and Surveys. Such categories are not rigidly fixed and users can freely sort the database by whatever class is convenient in their application.

LOINC uses HL7 codes (see paragraph 6.2.5) for clinical documents aiming at avoiding the development of a new terminology. According to the LOINC Guide 2011, the component terms used in the creation of the names of document type codes will be mapped to either the UMLS Metathesaurus, or SNOMED CT as soon as possible.

[59] http://en.wikipedia.org/wiki/LOINC

6.2.24 Medical Directory for Regulatory Activities (MedDRA)

Medical Dictionary for Regulatory Activities[60] is a clinically validated international medical terminology for diagnoses, symptoms, surgeries and other medical procedures. It is used by regulatory authorities and the regulated biopharmaceutical industry during the regulatory process, from pre-marketing to post-marketing activities, and for data entry, retrieval, evaluation, and presentation. In addition, it is the adverse event classification dictionary endorsed by the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). It is translated into several languages and according to the European Union, MedDra is the mandatory tool for coding and transmitting information on product characteristics and side-effects.

So since terms of MedDra are mandatory in the regulatory process, it has an impotant impact. However, it is not a useful tool for automated reasoning since clarification is missing for the relations which hold between the links to upper and lower term. It is not a real hierarchy and it is not a simple vocabulary with some links and connections which carry no semantic content at all. In that way MedDRA can be a useful controlled vocabulary but doesn't provide semantic relations.

6.2.25 Thesaurus of the National Cancer Institute (NCIT)

The Thesaurus of the National Cancer Institute (NCI)[61] covers vocabulary for clinical care, translational and basic research and public information and administrative activities It can be browsed online and downloaded in OWL-DL or OBO format and currently contains over 34,000 concepts, structured into 20 taxonomic trees. The NCI Thesaurus provides concept history tables to record changes in the vocabulary over time as the science changes.

NCIT is published under an open content license. It covers a broad domain of entities which are related to cancer, e.g. in genetics, anatomy and medication. The vocabulary is related to some other terminologies. For example, the semantic type of the concepts from the UMLS Semantic Network is given. There exists also a NCI Metathesaurus[62] which integrates terms from over 70 terminologies.

Although it lacks many qualities of a good ontology design, i.e. objective, descriptive definitions and a high level of formal exactness it is easy to understand for human domain experts according to the OBO foundry it provides the most granular and consistent terminology available today.

6.3 Overall technical analysis

The following table summarizes the overall technical analysis for the conceptual models that were identified as relevant to the MyHealthAvatar project. We remind the different columns of the following table:

- Modelling Language: Whether the conceptual model is written in a standard language

[60] http://www.meddra.org/
[61] http://ncit.nci.nih.gov/
[62] http://ncim.nci.nih.gov/ncimbrowser/

- Standardization Body: Whether the conceptual model has been approved by a standardization community
- Open Source: Whether the conceptual model and its documentation is offered free for usage.
- Available Mappings: Relations with other sources
- Human Readable Text: Whether the conceptual model provides human readable text together with computer readable text
- Inter-linguistic operability: Whether the conceptual model covers multi-linguistic frame
- Coherent/Consistent Semantics: “No” when no semantics exists in the model, “Low” when only a few formal definitions/relations are provided, “Medium” for many formal definitions provided, “High” when complete ontologies.
- Interoperability level: “No” when interoperability cannot be reached in the model, “Syntactic” when only syntactic interoperability is provided, “Semantic” when syntactic and semantic interoperability is provided.
- Usage: Whether the conceptual model is being used in other projects.
- Covered MyHealthAvatar domains: It lists the domains that are covered by each semantic resource.

Conceptual Model	Modeling Language	Standardization Body	Open source	Available Mappings	Human / Computer Readable Text	Interlinguistic Operability	Coherent/ Consistent Semantics	Interoperability Level	Usage	Covered MyHealthAvatar Domains
BCO	Ontology (OWL)	-	No	No	Yes	Yes	High	Semantic	No	Health Status & Clinical Information, Medication Data
Glossary of Terms for Community Health Care & Sevices for Older Persons	-	WHO	Yes	No	No	No	Low	Syntactic	No	Personal Information & Lifestyle, Health Status & Clinical Information
NNEW	Ontology	-	Yes	No	Yes	No	High	Semantic	Yes	Personal Information & Lifestyle
FHHO	Ontology	-	Yes	Yes	Yes	No	High	Semantic	Yes	Personal Information & Lifestyle, Health Status & Clinical Information
ATC DDD	Classification in pdf document	-	Yes	No	Yes	Yes	No	Syntactic	Yes	Medication Data
UMLS	Terminology (RRF)	US National Library of Medicine	Yes	Yes	Yes	Yes	High	Semantic	Yes	Health Status & Clinical Information, Medication Data
MeSH	Terminology	US National Library of Medicine	Yes	No	Yes	Yes	Medium	Syntactic	Yes	Health Status & Clinical Information, Medication Data
ICF	Ontology (OWL)	WHO	Yes	Yes	Yes	Yes	Medium	Semantic	Yes	Personal Information & Lifestyle, Health Status & Clinical Information, Medication Data
Ontology of Medically Related Social Entities	Ontology (OWL)	-	Yes	No	Yes	No	Medium	Semantic	No	Personal Information & Lifestyle
NIFSTD	Ontology (OWL)	-	Yes	Yes	Yes	No	High	Semantic	Yes	Molecular Data

Conceptual Model	Modeling Language	Standardization Body	Open source	Available Mappings	Human / Computer Readable Text	Interlinguistic Operability	Coherent/ Consistent Semantics	Interoperability Level	Usage	Covered MyHealthAvatar Domains
SO	Ontology (OWL/OBO)	OBO Foundry	Yes	No	Yes	No	High	Semantic	Yes	Health Status & Clinical Information
DO	Ontology (OWL/OBO)	OBO Foundry	Yes	Yes	Yes	No	High	Semantic	Yes	Health Status & Clinical Information, Molecular Data
OAE	Ontology (OWL)	-	Yes	Yes	Yes	No	High	Semantic	Yes	Medical Information, Health Status & Clinical Information
EFO	Ontology (OWL)	-	Yes	Yes	Yes	No	High	Semantic	Yes	Health Status & Clinical Information, Molecular Data, Medication Data
CCC	Terminology (Not provided)	US Department of Health and Human Services ISO/TC 215:Health Informatics	No	No	Yes	No	No	Syntactic	Yes	Health Status & Clinical Information, Medication Data
AMA CPT	Terminology (Not provided)	American Medical Association	No	No	Yes	No	No	Syntactic	Yes	Health Status & Clinical Information, Medication Data
SNOMED CT	Terminology (EL++ formalism)	IHTSO	Yes	Yes	Yes	Yes	Medium	Semantic	Yes	Health Status & Clinical Information, Medication Data
ICD-10	Terminology (Not provided)	WHO	Yes	Yes	Yes	Yes	Medium	Semantic	Yes	Health Status & Clinical Information, Medication Data
GO	Ontology (OWL/OBO)	OBO Foundry	Yes	Yes	Yes	No	High	Semantic	Yes	Molecular Data
ACGT MO	Ontology (OWL/OBO)	OBO Foundry	Yes	No	Yes	No	High	Semantic	Yes	Health Status & Clinical Information, Molecular Data, Medication Data

Conceptual Model	Modeling Language	Standardization Body	Open source	Available Mappings	Human / Computer Readable Text	Interlinguistic Operability	Coherent/ Consistent Semantics	Interoperability Level	Usage	Covered MyHealthAvatar Domains
SBO	Ontology (OWL/OBO)	BioModels.net	Yes	No	Yes	No	Yes	Semantic	Yes	Molecular Data, Systems Biology Models
FMA	Ontology (OWL/OBO)	University of Washington	Yes	Yes	Yes	No	High	Semantic	Yes	Health Status & Clinical Information
LOINC	The Silver Book (IUPAC); the International Federation of Clinical Chemistry (IFCC)	International Standard	Yes	Yes	No	Yes	No	Syntactic	Yes	Health Status & Clinical Information
MeDRA	Textfiles in ASCII: *.asc	MSSO, ICH, IFPMA	Yes	Yes	Yes	Yes	No	Syntactic	Yes	Health Status & Clinical Information
NCI-T	Ontology (OWL)	NCI	Yes	Yes	Yes	No	Medium	Semantic	Yes	Health Status & Clinical Information

7 Conclusion

In the present deliverable we focused on the requirement analysis for the semantic core ontology that will be developed the following months for the MyHealthAvatar project. Initially we reviewed similar approaches from related research projects to identify guidelines. Then we established the methodology that we will follow on the development of the ontology. The basic principles of the overall ontology development methodology that we will follow are reuse, granularity and modularity.

In a next step we analysed the first two steps of the aforementioned methodology, i.e. the Purpose and Scope Specification and the knowledge acquisition. In the first step we used the description of the work and the preliminary versions of the use-cases to identify the domain of interest and then we listed the most common ontologies in the aforementioned domain. An evaluation about their applicability in the MyHealthAvatar project was performed and the results show that many of them can be used in the MyHealthAvatar Semantic Core Ontology.

The following months we shall define the ontology, which will be tested using the final use-cases. Then we expect that the ontology will require extensions which will be reported in D4.2 and finally it will be evaluated thoroughly in D4.3.

8 References

[1] Berners-Lee, T., Hendler, J. & Lassila, O. 2001. "The Semantic Web" Scientific American, 284 (5), pp. 34-43.

[2] Brochhausen M. et al.: The ACGT Master Ontology and Its Applications – Towards an Ontology-Driven Cancer Research and Management System. IN: Journal of Biomedical Informatics, 44(1), pp 8-25, 2011.

[3] Bug W. J., Ascoli G. A., Grethe J. S., Gupta A., Fennema-Notestine C., Laird A. R., Larson S. D., Rubin D., Shepherd G. M., Turner J. A., Martone M. E. (2008). The NIFSTD and BIRNLex vocabularies: building extensive ontologies for neuroscience. Neuroinformatics 6, 175–194. doi: 10.1007/s12021-008-9032-z.

[4] Collier, N., Matsuda Goodwin, R., McCrae, J., Doan, S., Kawazoe, A., Conway, M., Kawtrakul, A., Takeuchi, K. and Dien, D.: An ontology-driven system for detecting global health events, Proc. 23rd International Conference on Computational Linguistics (COLING), pp.215-222, 2010.

[5] Corcho, O., Fernandez-Lopez, M., Gomez-Perez, A.: Methodologies, tools and languages for building ontologies: Where is their meeting point? Data Knowl. Eng., 46, pp 41–64, July 2003.

[6] Fernández-López, M.: Overview of Methodologies for Building Ontologies. In: The IJCAI Workshop on Ontologies and Problem-Solving Methods, Stockholm, Sweden, 1999.

[7] Fernandez-Lopez M., Gomez-Perez A., Juristo N.: METHONTOLOGY: from Ontological Art towards Ontological Engineering, Proceedings of the AAAI97 Spring Symposium, Stanford, USA , pp. 33 - 40, 1997.

[8] Gómez-Pérez, A.: Ontology Evaluation. IN: Staab, S. / Studer, R. (eds.): *Handbook on Ontologies*, Springer-Verlag Berlin Heidelberg, 2004.

[9] Gomez-Perez, A., Fernandez, M. and De Vicente, A.J.: Towards a Method to Conceptualize Domain Ontologies. In: ECAI-96 Workshop on Ontological engineering, Budapest (1996).

[10] Grüninger M., Fox, M. S.: Methodology for the Design and Evaluation of Ontologies, Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI-95, Montreal, 1995.

[11] Hartmann, J. et al.: *D1.2.3 Methods for ontology evaluation*, available from http://knowledgeweb.semanticweb.org/semanticportal/deliverables/D1.2.3.pdf

[12] He, Y., Xiang, Z., Sarntivijai, S., Toldo, L., Ceusters W.: AEO: A Realism-Based Biomedical Ontology for the Representation of Adverse Events, International Conference on Biomedical Ontology • Buffalo, NY, USA Representing Adverse Events Workshop, July 26, 2011.

[13] Hovy, E.: Methodologies for the Reliable Construction of Ontological Knowledge, In Proceedings of ICCS, pp. 91-106,Springer 2005.

[14] Imam, F.T., Larson, S.D., Bandrowski, A., Grethe, J.S., Gupta, A., Martone, M.E.: Development and use of Ontologies Inside the Neuroscience Information Framework: A Practical Approach, Published online 2012 June 22. DOI: 10.3389/fgene.2012.00111

[15]Kumar, A., Smith, B.: The Ontology of Processes and Functions. A Study of the International Classification of Functioning, Disability and Health, http://ontology.buffalo.edu/medo/ICF.pdf

[16]Lambrix, P. & Edberg, A. 2003. "Evaluation of Ontology Merging Tools in Bioinformatics" Proceedings of the 8th Pacific Symposium on Biocomputing, pp. 589-600.

[17]Luciano, J.S., Andersson, B., Batchelor, C., Bodenreider, O., Clark, T., Denney, C., Domarew, C., Gambet, T., Harland, L., Jentzsch, A., Kashyap, V., Kos, P., Kozlovsky, J., Lebo, T., Marshall, M.S., McCusker, J.P., McGuiness, D.L., Ogbuji, C., Pichler, E., Powers, R.L., Prud'hommeaux, E., Samwald, M., Schriml, L., Tonellato, P.J., Whetzel, P.L., Zhao, J., Stephens, S., Dumontier, M.: The Translational Medicine Ontology and Knowledge Base: Driving personalized medicine by bridging the gap between bench and bedside. Journal of Biomedical Semantics, 2011.

[18]Malone, J., Holloway, E., Adamusiak, T., Kapushesky, M., Zheng, J., Kolesnikov, N., Zhukova, A., Brazma, A., Parkinson, H.: Modeling sample variables with an Experimental Factor Ontology, Bioinformatics, 26(8), pp. 1112–1118, 2010.

[19]MyHealthAvatar Consortium: A Demonstration of 4D Digital Avatar Infrastructure for Access of Complete Patient Information - Description of Work, 2012.

[20]Peace, J, Brennan, P.F.: Ontological representation of family and family history, at AMIA Annu Symp Proc. 2007.

[21]Rosse, C., Mejino Jr., J.L.V: *A reference ontology for biomedical informatics. The Foundational Model of Anatomy, J Biomed Inform. 2003 Dec;36(6):478-500.*

[22]SemanticHEALTH project: Semantic Interoperability for Better Health and Safer Healthcare, European Commission, Information Society and Media, available from: http://www.empirica.com/publikationen/documents/2009/semantic-health-report.pdf

[23] Smith, B., Kohler, J, Kumar, A.: On the Application of Formal Principles to Life Science Data: a Case Study in the Gene Ontology, Data Integration in the Life Sciences, Lecture Notes in Computer Science, Vol.2994, 2004, pp. 79-94

[24] Staab, S., Studer, R.: Handbook on Ontologies, Springer-Verlag 2004, http://books.google.gr/books?id=0Elgz95mM8QC.

[25]Uschold, M.: Building Ontologies: Towards a unified methodology. InI. Watson (Ed.), 16th Annual Conference of the British Computer Society Specialist Group on Expert Systems, Cambridge, UK,1996.

[26]Uschold, M., Gruninger, M.: Ontologies: Principles, methods and applications. The Knowledge Engineering Review, 11(2) ,1996.

[27]Uschold M., King, M.: Towards a Methodology for Building Ontologies, Workshop on Basic Ontological Issues in Knowledge Sharing, held in conduction with IJCAI-95,1995.

[28]Schreiber, A. Th., Terpstra P.: Sisyohus-VT: A CommonKADS solution. Technical Report, ESPRIT Project 8145 KACTUS, University of Amsterdam, 1995. Submitted for publication.

[29]Schreiber, A. Th., Wielinga, B. J., Jansweijer, W. H.: The KACTUS view on the 'O' word. Technical Report, ESPRIT Project 8145 KACTUS, University of Amsterdam, 1995.

Appendix 1 – Abbreviations and acronyms

ACGT	Advancing Clinic-Genomic Trials on Cancer
ACGT MO	Advancing Clinical-Genomic Trials on Cancer Master Ontology
ACGT SM	Advancing Clinical-Genomic Trials on Cancer Semantic Mediation Layer
AEO	Adverse Event Ontology
AERO	Adverse Event Reporting Ontology
At	Anatomy Taxonomy
ASA	Anatomical Structural Abstractin
ATA	Anatomical Transformation Abstraction
ATC	Anatomical Therapeutic Chemical
BCO	BioCaster Ontology
BFO	Basic Formal Ontology
CCC	Clinical Care Classification System
CCD	Continuity of Care Document
CCR	Continuity of Care Record
CDM	Common Data Model
CHEBI	Chemical Entities of Biological Interest
CIM	Common Information Model
CL Ontology	Cell Type Ontology
CT	Clinical Trial
DDD	Defined Daily Dose
DO	Disease Ontology

DOW	Description of Work
EFO	Experimental Factor Ontology
EHR	Electronic Health Record
EMAP	Edinburgh Mouse Atlas
FHHO	Family Health History Ontology
FMA	Foundational Model of Anatomy
GO	Gene Ontology
HDOT	Health Data Ontology Trunk
HIS	Hospital Information system
HL7	Health Level 7
IAO	Information Artifact Ontology
ICD	International Classification of Diseases
ICF	International Classification of Functioning
ICT	Information & Communication Technology
IFOMIS	Institute for Formal Ontology and Medical Information Science
IGS	Institute for Genome Sciences
LOINC	Logical Observation Identifiers Names and Codes
LODD	Linking Open Drug Data
LUI	Lexical Unique Identifier
MA	Mouse Anatomical
MeSH	Medical Subject Headings
MK	Metaknowledge
MLOCC	Middle Layer Ontology for Clinical Care

MO	Master Ontology
MRI	Magnetic resonance imaging
NIF	Neuroscience Information Framework Project
NIFSTD	Neuroscience Information Framework Standardized Ontology
NLM	US National Library of Medicine
NNEW	Nextgen Network Enabled Weather
OBI	Ontology for Biomedical Investigation
OBO	Open Biomedical Ontologies
OGMS	Ontology for General Medical Science
OMRSE	Ontology of Medically Relevant Social Entities
OPB	Ontology for Physics in Biology
ORF	Original Release Format
OWL	Ontology Web Language
p-Medicine	Personalized Medicine
PATO	Phenotypic Quality Ontology
PHR	Personal Health Record
PRO	Protein Ontology
RICORDO	Researching Interoperability using Core Reference Datasets and Ontologies for the Virtual Physiological Human
RO	Relational Ontology
RRF	Rich Release Format
SNOMED CT	Systematized Nomenclature of Medicine Clinical Terms
SNOMED RT	Systematized Nomenclature of Medicine Reference Terminology
SOA	Service Oriented Architecture

TMKB	Translational Medicine Knowledge Base
TMO	Translational Medicine Ontology
TOVE	TOronto Virtual Enterprise methodology
UMLS	Unified Medical Language System
UO	Units Ontology
VPH-NoE	Virtual Physiological Hyman Network of Excellence
WHO	World Health Organisation
WP	Work Package