Deliverable No. 4.4

Initial release of the patient view and its evaluation

Grant Agreement No.:270089

Deliverable No.: D4.4

Deliverable Name: Initial release of the patient view and its evaluation

Contractual Submission Date: 31/01/2013

Actual Submission Date: 30/01/2013

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:	p-medicine
Project Full Name:	From data sharing and integration via VPH models to personalized medicine
Deliverable No.:	D 4.4
Document name:	Initial release of the patient view and its evaluation
Nature (R, P, D, O)	P
Dissemination Level (PU, PP, RE, CO)	PU
Version:	1
Actual Submission Date:	28/01/2013
Editor: Institution: E-Mail:	Emilio M. Sanfilippo: IFOMIS: emilio.sanfilippo@ifomis.uni-saarland.de

ABSTRACT:

This deliverable outlines the development, test and evaluation of a patient view connected to the semantic framework of p-medicine supplied by the Health Data Ontology Trunk (HDOT) and its modules. In our understanding a patient view is the organization and elaboration of all kinds of data collected and used within the p-medicine project in such a way that patients can easily access relevant or otherwise desired pieces of information, this information being presented together with its bigger conceptual context and enriched by further descriptions of the used concepts in lay terminology. In this process, a fundamental step is the definition of a lay language by which information primarily coded in medical jargon is then presented to patients in an easily understandable and non-technical fashion. For this purpose, we extend HDOT in a dedicated specific patient-oriented module called the Health Data Ontology Trunk for Patient Empowerment (HDOT_PEM). We further describe an evaluation and test procedure for gauging the usefulness of the information we supplied in this module for patients. The way the information is presented to patients can only be properly evaluated once the patient empowerment service developed in WP14 is implemented and tested.

KEYWORD LIST: Patient Empowerment, Ontology, Lay Terminology, Patient View

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n^{o} 270089.

MODIFICATION CONTROL			
Version	Date	Status	Author
0.1	01/11/2012	Draft	Emiio M.Sanfilippo
0.2	15/12/2012	Draft	Luc Schneider, Ulf Schwarz
0.3	23/12/2012	Draft	Ketti Mazzocco
1	28/01/2013	Final	Emilio M.Sanfilippo

List of contributors

- Emiio M.Sanfilippo, IFOMIS
- Luc Schneider, IFOMIS
- Ulf Schwarz, IFOMIS
- Ketti Mazzocco, IEO

Contents

1. Executive Summary

The main purpose of this deliverable consists in outlining the development, test and evaluation of a patient view connected to the semantic framework of p-medicine supplied by the Health Data Ontology Trunk (HDOT) and its modules. In our understanding a patient view is the organization and elaboration of all kinds of data collected and used within the p-medicine project in such a way that patients can easily access relevant or otherwise desired pieces of information, this information being presented together with its bigger conceptual context and enriched by further descriptions of the used concepts in lay terminology. In this process, an important step obviously is the provision of information in a lay language by which information primarily coded in medical jargon is then presented to patients in an easily understandable and non-technical fashion. For this purpose, we extend HDOT in a dedicated specific patient-oriented module called the Health Data Ontology Trunk for Patient Empowerment (HDOT_PEM). We further describe an evaluation and test procedure for the usefulness to patients of the information we supplied in this module. The way it is presented to them can only be properly evaluated once the patient empowerment service developed in WP14 is implemented and tested.

This patient view connected to HDOT is developed in close collaboration with WP2 and WP14, which is responsible for the development of the overall patient empowerment strategy of p-medicine and WP2 which deals with use cases.

2. Introduction

In the framework of personalized medicine patients play a central role at all stages of the health care procedures. Nevertheless, clinical trials, case report forms and the different clinical documentations usually contain medical jargon, so that they might be difficult to understand for those who lack technical knowledge.

The deliverable has four main purposes: 1) the definition of a data schema to serve as knowledge base and of 2) lay definitions, by which expert knowledge can be presented in a more suitable way to patients; 3) the development of a modular ontology under the semantic umbrella of the Health Data Ontology Trunk, by which the data schema can be integrated; 4) the outline of an ontology-driven patient view, i.e. a strategy to exploit the semantic references of the ontology to present biomedical and clinical information to patients.

We consider the ALGA (Mazzocco, at al., 2012) questionnaire and its results as the main knowledge base for the extraction of the data schema. The ALGA, indeed, has been intentionally designed in p-medicine to obtain psychological information about patients concerning their health status, so that by its results we are able to understand how patients react to their diseases and to decision-making procedures.

The lay definitions play a key role in this context, because with their help clinical information and biomedical knowledge can be presented to patients in a more suitable way. However, the development of such definitions is very difficult, because it needs to take into account different aspects of patients' life, e.g. literacy, education, health history, response to educational activities. At the same time, the definitions cannot contradict experts' knowledge, so that there is the need to work closely with both patients and health care providers.

Nevertheless, lay definitions by themselves cannot assure data sharing and integration, so that there is the need to fix semantic references among different people and institutions. By the development of a patient empowerment module within the semantic space of HDOT, respectively HDOT_PEM, we aim at extending the ontology library of p-medicine for the needs of the patient empowerment services. Thus, we argue that in order to present expert-biomedical knowledge to patients, we need not only the definition of a lay language, but also to fix stable references among a pre-defined semantic space.

Inasmuch as the ontology is developed to provide the semantic references for a patient empowerment tool, we define a strategy for the development of an ontology-driven patient view to show the results of the ALGA questionnaire. The patient view consists in a re-arrangement of the ALGA structure according to the hierarchy of the ontology such as to fix semantic references among the community of p-medicine, while at the same time improving laymen understanding.

The deliverable has the following structure: in the third chapter, we define the role of ontology in a patient empowerment scenario; in the fourth chapter, we develop the HDOT_PEM and provide a first glass-box evaluation; in the fifth chapter, we define a strategy for the development of an ontology-driven patient view and provide a black box evaluation for the HDOT_PEM. Moreover, in this chapter we provide more details about the development and evaluation of the lay definitions. The data schema extracted out from the ALGA questionnaire is included in the Annex 1.

3. Patient empowerment and semantic resources

Introduction

The goal of the WP4 is not the definition of a patient empowerment strategy for p-medicine. Rather, we develop a patient view connected to HDOT, i.e. the HDOT Patient Empowerment Module (HDOT_PEM). This new module helps to organize information from the data collection of p-medicine in a way that is more suitable for the patients' understanding of medical complex terms. However, before the development of HDOT_PEM, we need to identify the semantic requirements for a patient empowerment scenario in p-medicine, stating how ontology can help with this.

The semantic needs of p-medicine for a patient empowerment strategy

In the framework of personalized medicine, patients are encouraged to play an active role in decisions about their health care. It is thus important for them to be able to access factual, understandable and appropriate information about their therapeutic choices, if they have to interact responsibly in the decision making process together with their physicians.

According to the context scenarios for usability testing defined in the deliverable D.2.2[1], in a patient empowerment strategy the structure of information should be presented in such a way that patients understand their clinical documentation and are able to make informed choices. The process of understanding the whole data set that the hospital has collected is basically conceived in two ways:

1. Patients shall be able to understand medical statements, as well as legal and ethical considerations. This means that information must be translated into a patient understandable language;
2. The data must be structured and organized in a way that makes it easier to decide for patients what is of interest to them in order to acquire the desired knowledge about their health status and make informed decisions. Therefore, the information should be clearly arranged, so that the user can reach the desired piece of information overlooking the details quickly, collecting important issues and not get distracted from unnecessary information.

[1] (Graph, 2011), paragraph 4.3.2 and 4.3.3

In line with the first point, there is the need to present biomedical information to patients in a lay language that shall be easy to understand. However, this is not a simple task, because there is the need to take into account different aspects of patients' life style. In the last chapter we provide more details about the way we come to lay definitions. We focus mainly on the second point, extending HDOT in a patient-driven module, called the Health Data Ontology Trunk Patient Empowerment Module (HDOT_PEM). In the deliverable D.14.2, *Specification of the linguistic schema* (Mazzocco et al., 2012), we briefly showed how HDOT might help in a patient empowerment scenario. In this deliverable, we explain in more details our approach for the benefits of the project.

From expert to layman terminologies

The problem of organizing data in a way that patients can easily access the desired pieces of information must take into account that clinical information is provided using expert terminologies. However, this is not only a linguistic problem. In many cases, indeed, biomedical lay terminologies are compatible with expert terminologies [cf. (Fisch, K., & O'Connel, 1998)]. For example, the *Alternative Lay Language for Medical Terms in Consent Form*, developed by the University of Iowa, provides a glossary of terms to be used instead of medical jargon[2]. Following this approach, it could be the case that every expert term has a corresponding lay term, or at least an explanation in an easier and not-technical way.

Nevertheless, there are mainly two problems in translating expert terms into lay ones: first of all, the translation would not assure that both the expert and the layman are actually referring to one and the same existing entity, which has specific characteristics and holds specific relations to other entities. This could be the case because laymen might have only a very rough understanding of the biomedical world, while experts know more in details their domain of investigation. Secondly, a mere linguistic translation cannot afford data sharing and integration among different institutions. This is a very well known problem [cf. (de Bono et al., 2011); basically, if we aim at sharing data, we are not interested in terms but in their content, i.e. what the term is about. Therefore, there is the need to use semantic machine-readable languages (e.g. RDFS and OWL), so that the process of sharing and integrating data can be speeded up and automated without the risk of loosing important information.

The translation of medical jargon into lay terms would only be a part of a more general ontology-driven approach to patient empowerment. In an ontology the investigated domain is represented in order to constrain the possible interpretations of the same world among different users, so that they can exploit one and the same semantic description of data for different purposes, even using different labels. In other words, ontologies fix an objective

[2] http://hso.research.uiowa.edu/medical-terms-lay-language

meaning within their semantic space, so that it does not matter whether we use expert or lay terms to name something, because their reference would be exactly the same ontological class.

As an example, let us consider the definition of neoplasia given by the *Glossary of Lay Terms for Use in Informed Consent Forms* developed by the University of Florida[3]:

Neoplasia: Tumor, may be benign or malignant

Although this definition might be very easy to understand for a layman, from the biological point of view one should distinguish the *tumor* (or neoplasm) as an abnormal tissue in a part of the body from its developmental process, i.e. the *neoplasia* [cf. (Venes, 2001)]. The definition above it is only an all-encompassing explanation to avoid medical technicalities, and a physician should be aware of this fundamental distinction. It should be noticed that if such a definition were integrated in a semantic structure to share, annotate or integrate distributed data, it would be probable to produce inconsistencies and misunderstandings. Indeed, by 'neoplasia' some people may mean the tumor, while others could mean the neoplastic process, and of course these two entities have different characteristics (e.g. the tumor has a mass, while a neoplasia has not). However, using an ontological reference, we can avoid this kind of inconsistencies, while at the same time facilitating the understanding of laymen.

For example:

- http://purl.obolibrary.org/obo/DOID_0060072: This URI fixes a stable reference for the representation of 'benign neoplasm'. Even if we use different labels, e.g. 'benign tumor', or 'benign tumour', the reference does not change, as far as we keep the same URI;
- http://purl.obolibrary.org/obo/DOID_162: This URI fixes a stable reference for the representation of 'cancer'. As in the case above, the reference does not change if we use different labels, e.g. 'malignant tumor', or 'malignant tumour';
- http://hdot.googlecode.com/svn/trunk/hdot_pfm.owl#HDOT_PFM_neoplasia: This URI fixes a stable reference for the representation of the abnormal new growth of cells. In this case, too, we can use different labels, without affecting the meaning of the class.

In order to enhance the representation, we could even include in the ontology some owl axiom to indicate the relations between the above-mentioned classes. Using the ontological framework of the Basic Formal Ontology 2.0[4], the following axioms could be stated:

---

[3] http://irb.ufl.edu/glossary.htm#N
[4] http://code.google.com/p/bfo/

- benign tumor *realized in* some neoplasia;
- cancer *realized in* some neoplasia;
- neoplasia *realizes* some (benign tumor or cancer);

Although the meaning of neoplasia provided in the *Glossary* does not match with its meaning in the ontology, the definition can still be included. Indeed, in the example above neoplasia is modeled as the process by which abnormal cells growth, i.e. a process whose outcome is the tumor. Therefore, it still makes sense to include the definition in the ontology, adding a comment that makes clear the distinction between the tumor and its development.

In the next paragraph we explain in more details how ontology can help in a patient empowerment strategy.

The role of ontology in a patient empowerment strategy

The task of understanding ontologies is very problematic for those users who are not familiar with the complexity of the ontological engineering (e.g. clinicians and patients). An owl-ontology class, for example, looks like the following piece of code taken from a module of HDOT. In order to fully understand it, it is necessary a minimal knowledge background about different subfamilies of description logics (DLs) and the Ontology Web Language (OWL):

```
<owl:Class
rdf:about="https://hdot.googlecode.com/svn/trunk/hdot_pem.owl#HDOT_PEM_012"
>
        <rdfs:label xml:lang="en">cognitive aspect category</rdfs:label>
        <rdfs:subClassOf
rdf:resource="https://hdot.googlecode.com/svn/trunk/hdot_pem.owl#HDOT_PEM_0
08"/>
        <rdfs:subClassOf>
            <owl:Restriction>
                <owl:onProperty rdf:resource="&obo;IAO_0000136"/>
                <owl:someValuesFrom>
                    <owl:Class>
                        <owl:unionOf rdf:parseType="Collection">
                            <rdf:Description rdf:about="&obo;GO_0007613"/>
                            <rdf:Description
rdf:about="https://hdot.googlecode.com/svn/trunk/hdot_pem.owl#HDOT_PEM_005"
/>
                            <rdf:Description
rdf:about="https://hdot.googlecode.com/svn/trunk/hdot_pem.owl#HDOT_PEM_017"
/>
                            <rdf:Description
rdf:about="https://hdot.googlecode.com/svn/trunk/hdot_pem.owl#HDOT_PEM_018"
/>
                        </owl:unionOf>
                    </owl:Class>
                </owl:someValuesFrom>
```

```
    </owl:Restriction>
</rdfs:subClassOf>
</owl:Class>
```

If ontologies were presented in such a way to users, it their exploitation in daily contexts would be compromised.

Accordingly to Beck et al. (Beck, et al., 2012), there is currently a big gap between the research about ICT technologies for the biomedical domain and their exploitation in daily contexts and workflows. Several times researches are far away to know what really happens in the daily workflow of hospitals and single physicians, while health care providers are often reluctant to exploit technologies that seem to be very complex to use. We think that this gap is going to increase, if the results of researchers are presented not only to domain experts, but also to laymen (e.g. patients), who in many cases lack a detailed knowledge to grasp the complexity of both biomedicine and computer science applications.

However, while ontologies can be considered as complex logical structures, they are very convenient solutions for distributed data integration. Indeed, they disambiguate the meanings of data among different communities of agents assuring high levels of semantic interoperability.

Our strategy is mainly based on keeping ontology development, evaluation and maintenance at the expert level, while facilitating the exploitation of semantic resources in semantic-driven systems whose structures are explicitly designed to allow lay users to work efficiently with them in a user-friendly way.

It is very probable that from the user perspective an ontology will still be considered as a black-box, but this lies in the very nature of ontological engineering, i.e. that ontologies are for programs, not for people [cf. (Hunter, 2002)]. The main consequence of this view is that an ontology cannot be directly patient (user)-friendly, but it is the whole system in which the ontology is integrated that can be as such, i.e. patient (user)-oriented. A computational ontology by itself cannot achieve anything more than the domain representation in a machine readable format according to domain specifications and ontological engineering standards, while a whole computational system can be designed in such a way to be useful for both experts and laymen.

Nevertheless, it is true that the structure of an ontology can be designed in a way that it facilitates the user-friendliness, for example adding classes definitions, comments, examples and synonyms that might help users.

Therefore, aiming at the development of an ontology that contains biomedical information, which can be easily explored by patients through specific semantic-driven software and applications (e.g. the Interactive Empowerment Service, IEmS), we provide p-medicine with HDOT_PEM. We design, evaluate and continuously extend this module according to the needs of p-medicine, particularly for the benefits of the patient empowerment services. In the module, we define comments, definitions and eventually synonyms and examples of use for all the classes, so that laymen can explore the ontology and understand the information stored in it without being ontology experts.

In particular, we use in HDOT_PEM the following annotations:

- `rdfs:label`. It provides a label for the class in natural language (English). If there are more labels available, we include different ones according to domain experts' standards. For example, the class `http://purl.obolibrary.org/obo/OGMS_0000060` has 'bodily process' and 'organismal process' as labels;

- `IAO_definition`. It provides a definition for the class in natural language (English). According to the principles of the OBO Foundry[5], definitions shall be written using the genus-differentia format: 'A is a B that C', where A is the class we want to define, B is its immediate upper-level class and C is something by which is possible to get the difference between A and B. For example, the class `http://purl.obolibrary.org/obo/OGMS_0000045` (`rdfs:label` 'disorder') has definition: 'A disorder (A) is a material entity (B) which (C) is clinically abnormal and part of an extended organism'. However, this kind of definitions are more understandable for ontology designers rather than for the real users, so we include also alternative definitions to enhance the understanding of the class;
- `HDOT_PM_patientLanguageDefinition`. It provides a definition for the class in natural language (English) and is explicitly thought to facilitate layman's understanding. For example, the class `http://purl.obolibrary.org/obo/OGMS_0000077` (`rdfs:label` 'pathological formation') has `patientLanguageDefinition`: 'Pathological formation is an abnormal grow of tissues in the organism that harms healthy tissues (neoplasm)';
- `Rdfs:comment`. It provides a comment for the class in natural language (English), i.e. it adds more information about the class. For example, the class `http://purl.obolibrary.org/obo/MD_0000049` (`rdfs:label` 'depression') has `rdfs:comment`: 'Depression is considered a heterogeneous condition in which different biologic abnormalities may be present such as a significant weight loss or gain, too little or too much sleep, physical agitation or slowing down, fatigue or loss of energy, feelings of worthlessness or excessive guilt, feelings of hopelessness and helplessness, lowered ability to think, to concentrate or make decisions and recurrent thoughts of death or suicide';
- `IAO_example of usage`. It adds general information about the use of the class, i.e. it specifies which classes can be subsumed under it. For example, the class `http://purl.obolibrary.org/obo/OGMS_0000077` (`rdfs:label` 'pathological formation') has `example of usage`: 'tumour, ulcer'.

The following example shows the pathological formation class under HDOT_PEM in which we apply our patient empowerment strategy:

```
<!-- http://purl.obolibrary.org/obo/OGMS_0000077 -->
```

5 http://obofoundry.org/wiki/index.php/FP_006_textual_definitions

```
    <owl:Class rdf:about="&obo;OGMS_0000077">
        <rdfs:label xml:lang="en">pathological formation</rdfs:label>
        <rdfs:subClassOf
rdf:resource="&trunk;hdot_core.owl#HDOT_CORE_026"/>
        <obo:IAO_definition xml:lang="en">(HDOT_CORE) A pathological
formation is a material pathological entity that results from a
pathological process</obo:IAO_0000115>
        <obo:IAO_definition source
xml:lang="en">http://purl.obolibrary.org/obo/OGMS_0000077</obo:IAO_0000119>
        <obo:IAO_example of use xml:lang="en">tumour,
ulcer</obo:IAO_0000112>
<hdot_pm_patientLanguageDefinition xml:lang="en"> Pathological formation is
an abnormal grow of tissues in the organism that harms healthy tissues
(neoplasm)</hdot_pm_patientLanguageDefinition>
    </owl:Class>
```

In order to grasp the difference between the patient empowerment strategy and the usual way to define a class in ontological engineering, we show exactly the same class without any of the annotations above:

```
<!-- http://purl.obolibrary.org/obo/OGMS_0000077 -->

    <owl:Class rdf:about="&obo;OGMS_0000077">
        <rdfs:label xml:lang="en">pathological formation</rdfs:label>
        <rdfs:subClassOf
rdf:resource="&trunk;hdot_core.owl#HDOT_CORE_026"/>
    </owl:Class>
```

From the point of view of the machine there is no difference between the two pieces of code, which means that the machine reasons exactly in the same way both in the first and in the second case. Indeed, currently there is only one axiom (`<rdfs:subClassOf>`) defining the ‘pathological formation’ class in the HDOT. Nevertheless, from the user point of view there is more detailed information in the first than in second case.

4. The Patient Empowerment Ontology

Introduction

In the following we apply the strategy developed in the previous chapter for the development of HDOT_PEM. We extract a data schema from the ALGA results [cf. (Mazzocco, et al., 2012) and provide classes definitions in a lay language.

The state of HDOT

HDOT results of a complete re-design of the Middle Layer Ontology for Clinical Care[6] (MLOCC), which also has been developed by IFOMIS. The main methodological difference in the development of HDOT is that as many classes as possible are actually re-used from other, well-established ontologies that are part of the OBO Foundry, thus maximizing re-use and compatibility. HDOT is composed of HDOT_CORE, a minimal specification of clinically relevant classes that immediately specify BFO 2.0 leaf nodes, and HDOT_PM, the extension of HDOT_CORE that is aimed at providing a large enough basis for the development of all ontology modules within the p-medicine project.

HDOT_CORE integrates under the same semantic umbrella the most recent version of the Basic Formal Ontology[7] (BFO 2.0), the Information Artifact Ontology[8] (IAO), parts of the Phenotypic Quality Ontology[9] (PATO), the Chemical Elements of Biological Interest (ChEBI)[10] the Foundational Model of Anatomy (FMA)[11], and most of the Ontology for General Medical Science[12] (OGMS). They are contained in the OBOFoundry initiative [3] and are widely used in the biomedical domain for data annotation and integration. HDOT_PM extends this core by including classes from the Ontology of Biomedical Investigations (OBI)[13], the Mental Functioning Ontology (MFO)[14], as well as further classes from the

6 http://www.ifomis.org/chronious/mlocc
7 http://purl.obolibrary.org/obo/bfo.owl
8 http://obofoundry.org/cgi-bin/detail.cgi?id=information_artifact
9 http://obofoundry.org/cgi-bin/detail.cgi?id=quality
10 http://purl.bioontology.org/ontology/CHEBI
11 http://purl.bioontology.org/ontology/FMA
12 http://obofoundry.org/cgi-bin/detail.cgi?id=OGMS
13 http://purl.bioontology.org/ontology/OBI
14 http://code.google.com/p/mental-functioning-ontology/source/browse/trunk/ontology/MF.owl

aforementioned ontologies. The imported classes retain their URI's, except in case they are not compatible with Protégé (as in the case of FMA). Only a few proper HDOT-classes have been added in order to ensure a better integration of the different components.

HDOT (i.e HDOT_CORE and its extension, HDOT_PM) is designed in a modular fashion as a middle-layer ontology in the sense that it specifies upper-level domain independent classes down to the biomedical domain while maintaining at the same time a very general semantic and axiomatic structure that can be further developed and specialized in different modules for different purposes and applications. Therefore a middle-layer ontology can be thought as an ontology whose classes are domain-driven but application-independent, in the sense that they represent general structural characteristics and properties of the domain without specializing them in very specific semantic and axiomatic features. HDOT is intentionally developed, maintained and further extended as a modular ontology, so that we have the possibility to add more information and introduce new subclasses according to users' needs without ever having to alter its overall structure. One core requirement put on HDOT is that it is broad enough to contain all general classes under which all necessary and more specific semantic content can be subsumed in such a way that a meaningful axiomatic relational structure is conferred upon those specific area of contents. Indeed, we can enrich the semantic content of different resources by subsumption under classes of HDOT_PM, e.g. integrating specific parts of biomedical terminologies like ICD-10, the NCI thesaurus or SNOMED-CT, in order to enhance their formal and semantic constraints in a computable way.

HDOT's development as a middle-layer ontology is governed by three main related structural considerations in order to achieve the highest level of semantic interoperability between heterogeneous data sources, maintain a high level of ontological soundness and ensure a high degree of expandability:

- HDOT is designed at a level of generality such that HDOT classes and relations cover all areas of the health-care domain, i.e. there is a meaningful ontologically well-defined HDOT super-class under which all necessary parts and pieces of semantic data descriptions (annotations, metadata) can be directly subsumed or otherwise represented[15]. The semantics of HDOT central body is supposed to change in the further development of the project only in case problems related to HDOT's application to the project itself or clinicians' needs emerge;
- The core ontological structure integrates different modular ontologies at different levels of granularity. Each class is provided with a deep axiomatization, which guarantees to the users' workflow high degrees of semantic representation and

[15] However, it is worth noting that the HDOT is currently under development, therefore it is very probable that the domain coverage is not yet complete.

syntactic reasoning, together with the ability to construct defined classes and composite terms;

- HDOT's modules for specific applications can be obtained by stating further specifications of HDOT classes, i.e. by inserting subclasses under existing HDOT slots (super-classes).

The most important owl-class axioms are governed by the ordering relation of subsumption between classes while most of the other axioms originate from the following four considerations: (1) To provide machine readable and computable class constructions; (2) to provide ontologically sound relations between classes and the corresponding labels to enable the desired reasoning and inference capabilities; (3) to provide the basis for the composition of biomedical complex concepts (e.g. 'blood pressure increase', 'resection of tumour in kidney') by axiomatizing the necessary relations between HDOT's constituents; (4) to include relations which bridge different levels of granularity (e.g. "part of" or "contains").

At the current state, we provide a library of HDOT modules according to the different needs of p-medicine:

- HDOT_CORE: minimal representation of clinical relevant classes;
- HDOT_PM: representation of p-medicine relevant classes;
- HDOT_PFM: representation of oncological relevant classes;
- HDOT_PEM: representation of the data collection of p-medicine in a patient-driven way;
- HDOT_BSDS: representation of classes for biobank access;

The library is accessible under the following web address: http://code.google.com/p/hdot/[16].

Designing the HDOT Patient Empowerment Module

In order to develop the HDOT_PEM, we need to pursue the following steps:

1. Identify clearly the ALGA terms to be integrated in the ontology. This depends on the needs of the IEmS, as well as on the consistency of the representation (e.g. if 'sensory perception of cancer related pain' is included, then there is the need to include information about cancer);

[16] To explore HDOT and its modules using Protégé, you have to open the google code page, go to the page "source" and then to "browse". Click on "trunk" on the right column; open the file you want to explore (hdot_pm.owl for example), open the link "view raw file" on the right side of the page with Protégé.

2. Check the ontological structure of the entity to which the term refers. We need to identify the status of the entity we want to include under the module. Developing a multi-scale ontological representation for data sharing and integration, we use the top-level ontological framework of the BFO 2.0. In this second step a term starts to be considered as a class (type, universal, node, etc.), i.e. a group of entities sharing exactly the same properties;
3. Verify whether the classes to be integrated are already part of the main structure of the HDOT. In this way, we need to face two different scenarios:
    a. The class is part of the HDOT. Thus, we do not create a new owl class but reuse the existing one. Nevertheless, we must be sure that the questionnaire term and the one in the ontology share the same meaning, i.e. the same semantic description in the ontology. If they do not mean the same, then we need to disambiguate them. Moreover, even if the meanings are the same, we might include more information in the ontology according to the questionnaire results;
    b. The term is not part of the HDOT. Thus, we need to include it under its structure by creating a new owl class. However, we have to check whether other ontological resources contain the entities we are going to represent, so that we foster the semantic interoperability between different resources;
4. Check whether the class is part of existing resources out of the HDOT library. We might reuse classes from external ontologies that are widely used among the scientific community and have undergone an evaluation process. It is very probable that we reuse OBO Foundry ontologies[17] and the ones that are developed under the frame of the BFO.

   Therefore, when we create the new owl class under the module, we might have three different possibilities: 1) to develop it as a new class, or 2) to reuse an HDOT class extending it with more useful information, or 3) to import it from external resources and adapting it to the needs of p-medicine.
5. Provide an owl axiomatic description for the new class. Once we subsume the owl class in the module, we might provide owl axioms to enhance the expressivity of the representation;
6. Provide class labels, comments and definitions in a lay language. In this way, we aim at hiding the complexity of the ontology to the final user, improving at the same time the user's understanding of its structure.

[17] http://obofoundry.org

As final development steps, we need to evaluate the module and define an integration strategy for the IEmS.

In the Appendix 1 at the end of this document we attach the data schema extracted from the ALGA questionnaire.

The HDOT Patient Empowerment Module

The Health Data Ontology Trunk Patient Empowerment Module (HDOT_PEM) is a modular ontology, encoded in OWL 2 (description logic expressivity: SIQ) and developed using the ontology editor Protégé. It is accessible on the following web address: https://code.google.com/p/hdot/.

HDOT_PEM imports the basic ontologies of p-medicine, i.e. HDOT_CORE, HDOT_PM and HDOT_PFM. The modules are bound together and made interoperable with each other using the construct `rdfs:SubclassOf`, so that the classes constituting a module are always the subclasses of a more general module. The integration of the different modules of HDOT is a necessary step for the purposes of HDOT_PEM, because they represent the data collection of p-medicine, which we want to represent in a patient friendly way, according to the patient empowerment strategy of the project.

The ALGA questionnaire results are integrated in the ontology following the specifications given in the data schema (see Appendix 1). However, the questionnaire basically concerns psychological information about patients, and no module of HDOT was originally conceived for such a representational purpose. Therefore, we import under HDOT_PEM the main classes of the Mental Functioning Ontology[18] (MF), the Mental Disease Ontology[19] (MD) and the Emotion Ontology[20] (MFOEM). We decide to use these resources, because: 1) their structure is HDOT-compatible, i.e. their building blocks are the same of HDOT (e.g. BFO 2.0 and OGMS[21]); 2) they provide the main general classes for the representation of psychological information.

In the first release only few owl-classes (almost 40) have an HDOT_PEM URI, while different ones have been integrated from external resources using their original URIs. In this way, by reusing already existing resources, we foster the interoperability between different semantic

[18] http://code.google.com/p/mental-functioning-ontology/
[19]http://code.google.com/p/mental-functioning-ontology/source/browse/trunk/ontology/MD.owl
[20] http://code.google.com/p/emotion-ontology/
[21] http://obofoundry.org/cgi-bin/detail.cgi?id=OGMS

terminologies and ontologies. The main relation stated in HDOT_PEM between classes is `rdfs:SubclassOf`, which is the fundamental one to create an hierarchy of terms associated with their meanings. In some case, we include also owl axioms where we need to bind together the classes in more details. However, it is possible that the number of classes having a proper HDOT_PEM URI increases, once more details about the use of the module are available.

The following picture (figure 1) shows a branch of HDOT_PEM at the level of the BFO upper level class (i.e. domain independent) ‘continuant’. Particularly, it shows the sub-branch of the BFO ‘disposition’, which is specified in more details according to the ALGA questionnaire. The HDOT_CORE class ‘ability’ is specified in HDOT_PM in ‘cognitive ability’, ‘social ability’ and ‘physical ability’. HDOT_PEM extends the first one in ‘ability to learn’, ‘ability to make decisions’, ‘ability to plan’, ‘ability to process information’ and ‘ability to reason’; the second one in ‘ability to interact with other’, The OGMS class ‘disease’ has different subclasses, taken from the Ontology for Human Diseases (DOID). Following Ceuster and Smith (Ceusters & Smith, 2010) and the ontological representation put forward by the MD, we specify the class ‘disease of mental health’ in ‘depression disease’ and ‘cancer-related depression disease’.

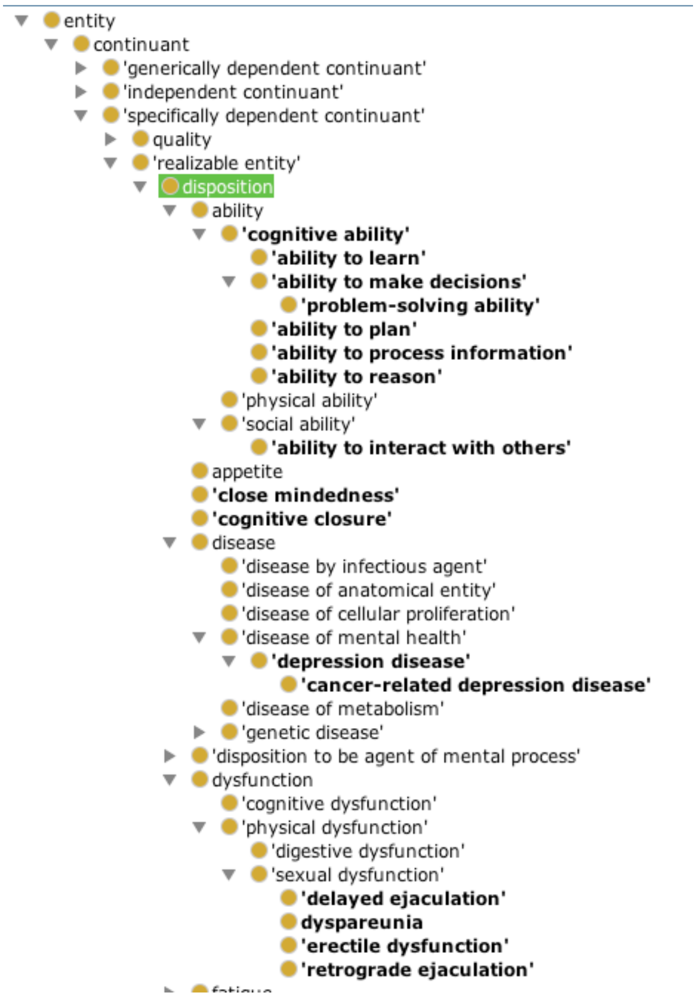


Figure 1 – The HDOT_PEM at the level of the BFO ‘continuant’/’disposition’

The following picture (figure 2) shows HDOT_PEM at the level of the BFO class ‘occurent’. Extending the hierarchy of the BFO with the OGMS, the MFO and the MFOEM, we subsume under the OGMS class ‘bodily process’ the MFO class ‘mental process’ together with its subclasses, i.e. ‘appraisal process’ and ‘cognitive process’. The last one is extended in ‘emotion process’ using the specifications given by the MFOEM. Recalling the representation of the diseases and following the framework of the MD, we import the class ‘pathological mental process’ and its subclass ‘depression mood episode’.

Figure 2 – The HDOT_PEM at the level of the BFO ‘occurent’

According to the ontology-driven patient empowerment strategy described in the previous chapter, we provide comments and patient language definitions for many classes in the ontology. In the following, we show as example the class ‘depression’ and its annotations:

```
<!-- http://purl.obolibrary.org/obo/MD_0000049 -->

    <owl:Class rdf:about="&obo;MD_0000049">
        <rdfs:label xml:lang="en">depression disease</rdfs:label>
        <rdfs:label xml:lang="en">depression</rdfs:label>
        <rdfs:subClassOf rdf:resource="&obo;DOID_150"/>
        <obo:IAO_0000115 xml:lang="en">Depression is a disease of mental
health that is realized in a depressed mood episode</obo:IAO_0000115>
<obo:IAO_0000115 patientLanguageDefinition xml:lang="en">Depression is a
mental disease characterized by an all-encompassing low mood accompanied by
```

```
low self-esteem, and by loss of interest or pleasure in living
</obo:IAO_0000115>

        <hdot_pm:patientLanguageDefinition xml:lang="en">Depression is a
long lasting mood change, characterized by mild to deep sadness, reduced
interest or pleasure in activities, social withdrawal with associated low
self-esteem and fatigue


source: p-medicine, D.4.4, Initial release of the patient view and its
evaluation</hdot_pm:patientLanguageDefinition>

<rdfs:comment xml:lang="en">Depression is defined as a psychophysical
condition associated with loss events (e.g., loss through death of a loved
one; loss of one&apos;s physical health), in which something or someone
that is valued is lost and perceived as hard or impossible to regain.
Depression is considered a heterogeneous condition in which different
biologic abnormalities may be present such as a significant weight loss or
gain, too little or too much sleep, physical agitation or slowing down,
fatigue or loss of energy, feelings of worthlessness or excessive guilt,
feelings of hopelessness and helplessness, lowered ability to think, to
concentrate or make decisions and recurrent thoughts of death or suicide.

source: p-medicine, Deliverable 14.2, Specification of linguistic
schema</rdfs:comment>

    </owl:Class>
```

It is worth remembering that the main advantages of including the results of the ALGA questionnaire in an HDOT-module perspective:

- The improved communication among the institutions involved in p-medicine. Using ontological references, the meanings of complex terms are disambiguated among the community of p-medicine. Moreover, everyone using p-medicine architecture can share the same semantic data descriptions avoiding inconsistencies and misunderstandings;

- The automated possibility to share and integrate data. The axiomatization of the ALGA results in OWL aims at making the data machine-understandable with high degrees of semantic precision and expressiveness. Structuring the information using the linguistic standards put forward by the W3C, we facilitate the automatic process of data sharing and integration among different institutions;

- The understanding of complex expert terminologies from lay people. Structuring the information in the ontology, we organize the data so that people can easily access the desired pieces of information, overlooking the rest. At the same time, the ontology provides class definitions, synonyms and comments with the purpose of facilitating the understanding of complex biomedical terms for laymen.

However, as we stress in the 3rd chapter, if we aim at making patients more involved in the health care and decisions making processes, then the ontology cannot do the job by itself. Rather, it should be considered as a part of a more complex ontology-driven user-friendly software architecture, by which patients access their data using the semantic references given by the ontology.

Glass box evaluation of the HDOT_PEM

In the deliverable 4.1, Domain quality checked ontologies and initial release of HDOT (Sanfilippo, Strößner, & Schwarz, 2012), we described the details of the ontology evaluation process (5th chapter) and provided a methodology to evaluate ontologies and terminologies (7th chapter). We applied our methodology to the evaluation of different resources and the table below (table 1) shows the results of the evaluation process

Terminologies/ Ontologies	FT	RDS	ILO	ISC	CO	SEM	LI
ICD-10	1	0	1	1	1	1	1b
SNOMED CT	0	0	1	1	1	2	3
LOINC	0	Out*	1	1	1	0	1b
DICOM	0	0	1	1	1	0	1b
FMA	1	1	1	1	1	3	3
GO	1	1	1	1	1	3	3
HL7 RIM v3	1	0	1	1	1	0	1b
openEHR archetypes	0	1	1	1	1	3	3
MedDRA	1	0	1	0	1	0	1a
CDISC	1	0	1	1	1	0	1b
UMLS	1	1	1	1	1	2	2b
DRG-G	1	Out*	1	1	1	0	1b
NTCI	1	1	0	1	1	2	2b-ILO
ATC/DDD	1	0	1	0	1	0	1a
ICF	1	0	1	1	1	1	2a
ICNP	1	0	1	1	1	1	2a

Table 1 – Evaluation of semantic terminologies and ontologies

In the following table (table 2), we apply the same methodology for the glass box evaluation of HDOT_PEM.

Resource	FT	RDS	ILO	ISC	CO	SEM	LI
HDOT_PEM	1	1	-	1	-	3	3

Table 2 – Glass box evaluation of the HDOT_PEM

Currently, it is not possible to evaluate HDOT_PEM according to the specifications given by ILO and CO. An inter-linguistic framework (ILO) is not provided within the framework of p-medicine, and it will be possible to take into account CO, once the structure of the IEmS is specified. However, in the next chapter we show a preliminary strategy for the integration of the ontology in a software architecture, specifically to show the ALGA results in a graphical user interface (GUI).

In the following, the legend of the table:

1. FT: Free Text: whether the terminology/ontology provides human readable text together with computer readable format;

   0 – No, 1 – Yes

2. RDS: Rigid Domain Specification: whether there is a clear distinction about what a domain is according to different levels of granularity, and what experts say about that domain by medical statements, reports, clinical material, etc… (e.g. use of ambiguous terms like “NOS” in ICD-10)

   0 – No, 1 – Yes

3. ILO: Inter-linguistic operability: Can the use of the terminology help to overcome linguistic barriers? This criterion covers “Multi-linguistic frame” and “Use of ID terms”.

   0 – No, 1 – Yes

4. ISC: Inter-standard connection: Are there any efforts to connect the standard to other standards?

   0 – No, 1 – Yes

5. CO: Conventionality: Is the terminology widely used? Are there any norms to use the terminology? There seems to be some advantage for interoperability if (almost) everyone uses the same standards.

   0 – No, 1 – Yes

6. SEM: Semantics: There is no semantic interoperability if there is no semantics. Does the standard provide any semantic information? Is the information provided in a machine readable way? Is the semantic coherent, consistent and is the information true?

0 – No formal semantics, 1 – Few formal definitions/relations, hierarchical structure 2 – Ontology-like but semantically incomplete (case of ambiguities, see Rigid Domain Specification) 3 – Ontology

7. LI: Levels of interoperability:
   - Level 0: no interoperability
   - Level 1: syntactic interoperability
     - 1a: inter-linguistic operability
     - 1b: inter-linguistic + inter-standard or/and inter-linguistic + high conventionality
   - Level 2: semantic interoperability: syntactic interoperability (1b) and
     - 2a: only few machine readable definitions and relations and some semantic information
     - 2b: much semantic information, a lot of ontology-like features but not complete (or/and some incoherent definitions)
   - Level 3: ontology

If a general higher level is reached but one criterion for a lower level is not met the higher level is given with an indication of the missing criteria. E.g. an ontology with inter-standard operability but without inter-linguistic interoperability is graded as 3-ILO

5. The Ontology View of the ALGA

Introduction

In the following chapter we show a strategy for the use and implementation of HDOT_PEM in an ontology-driven system, so that the data collection of p-medicine can be presented in a user-friendly way using semantic references. Our strategy basically consists in linking the ALGA to the ontology, so that the results of the questionnaire can be easily and quickly retrieved using the semantic references in HDOT_PEM. In this approach, the ontology remains a black box from the users' perspective. We aim at giving them the possibility to work with ontology-driven systems without the necessity to fully understand the complex structure of the ontology behind. Through this procedure we test the usefulness of HDOT_PEM for the benefits of p-medicine.

The patient view of the ALGA according to the HDOT_PEM

We aim at creating an ontology view that can be used to facilitate the work of the software engineers in the development of a graphic user interface (GUI) for different applications, in this case to show the results of the ALGA. Our strategy consists in showing the main categories of the questionnaire and let the user decide which category she wants to browse according to her needs.

Therefore, let us recall the main structure of the ALGA:

1) Perceived health state:
    a. Global Self-Rated Health;
    b. Fatigue;
    c. Pain;
    d. Physical abilities;
    e. Appetite.

2) Psycho-social aspects:
    a. Social abilities;
    b. Sexual problem;
    c. Body Image;

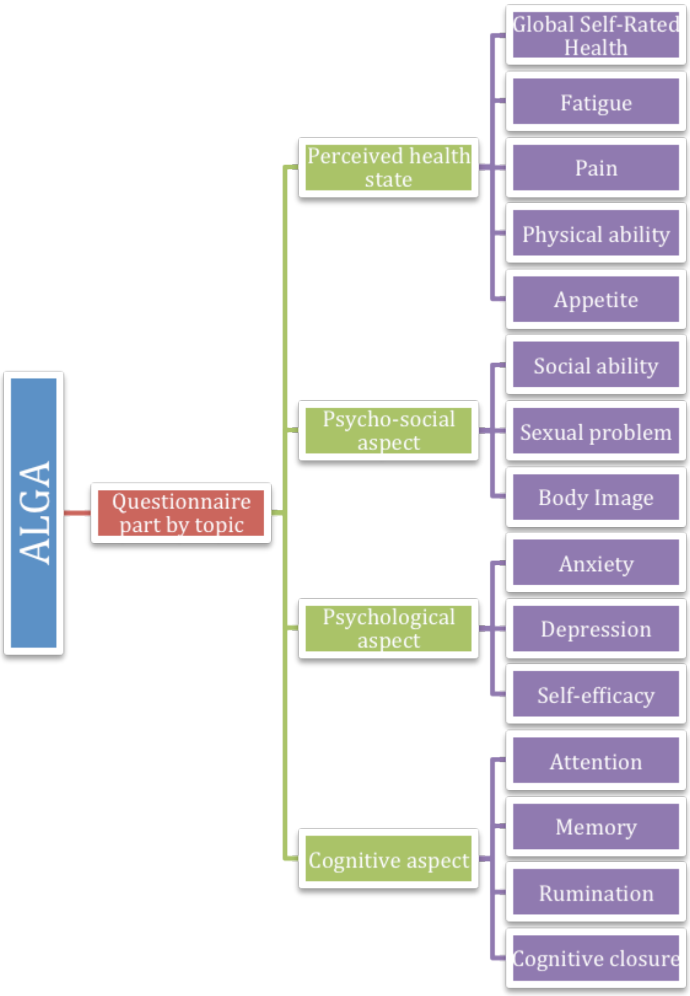
3) Psychological aspects:
    a. Anxiety;
    b. Depression;
    c. Self-efficacy;

4) Cognitive aspects:
    a. Attention;
    b. Memory;
    c. Rumination;
    d. Cognitive closure

We need to re-structure the ALGA according to the hierarchy of HDOT_PEM. The restructuring process is fully documented in the data schema in the Appendix 1 of this deliverable.

As an example let us consider 'physical ability', which belongs to the first category of the ALGA (perceived health state), because it refers to the way in which patients perceive their physical skills. However, from an ontological point of view it is not possible to state a subsumption relation (`rdfs:SubclassOf`) between the classes 'perceived health state' and 'physical ability', because not all particular instances of physical abilities are instances of perceived health state. Moreover, it seems possible to consider one's physical abilities without taking into account what she actually thinks about them. This point of view reflects also the clinical reality: when a physician evaluates her patients' physical abilities, she does not consider what they think about their skills, but the abilities themselves. Therefore, in HDOT_PEM 'physical ability' refers to one's physical skills regardless of one's mental attitudes. Then, we create a new owl class 'appraisal of physical abilities', which is the subjective perception of one's own physical abilities, i.e. it is relative to how the patient evaluates her physical skills.

The following picture (figure 3) shows the view of the ALGA according to the data structure of HDOT_PEM.



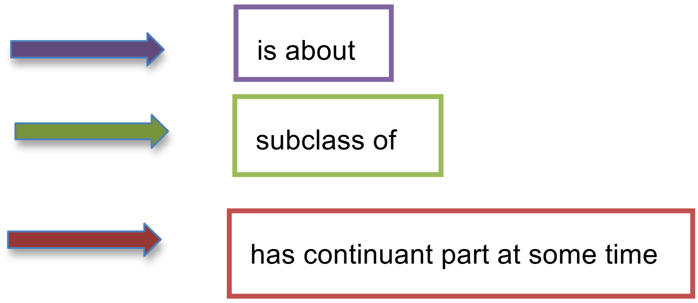


Figure 3 – ALGA view using the HDOT_PEM

The above view can be implemented in a GUI, showing the ALGA to the users and having the ontology as semantic reference. Despite the fact that its data structure is the same one of the ALGA, the terms are structured according to the semantics of HDOT_PEM. The categories of the questionnaire in the violet boxes belong to different categories of the ontology and are related to the categories in the green boxes using a specific axiom. The development of the GUI indeed has to take into account the logical structure of the ontology, specifically, the relations and the axioms between the different classes. In the picture above, the terms of the ALGA are related to each other with three different axiomatic relations: 'has continuant part at some time', 'subclass of' and 'is about'. The first one is useful to relate 'ALGA' to 'questionnaire part by topic'; the second one is useful to relate 'questionnaire part by topic' to 'cognitive aspect', 'perceived health state', 'psycho-social aspect' and 'psychological aspect'; the third one is useful to relate the main categories of the questionnaire to their different topics specified in the ontology. Since the user must be able to access the information stored in the ontology, each topic of the questionnaire has to be related with its subclasses using the 'subclass of' relation and the different axioms defined in the ontology. Therefore, the developer of the GUI should pay attention to the different axioms defined in the ontology for each class and make a decision about which of them should be included in the GUI according to her needs.

In order to understand the difference between the ontology-driven patient view of the ALGA (figure 3) and the real structure of HDOT_PEM, we show the subclasses of the class 'questionnaire part by topic' and their related owl axioms. These pictures are taken from HDOT_PEM using the ontology editor Protégé.

Figure 4 – 'Cognitive aspect' and its related topics

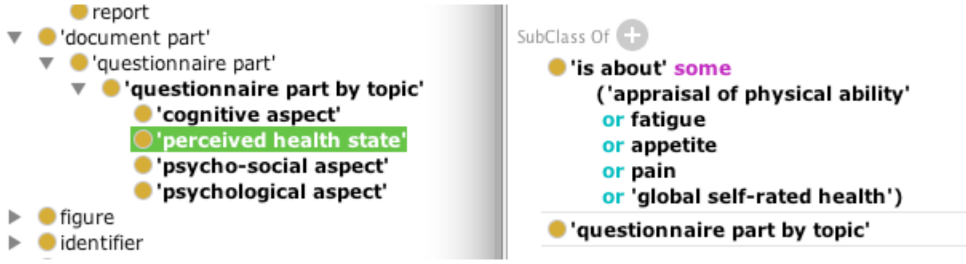


Figure 5 – 'Perceived health state' and its topics

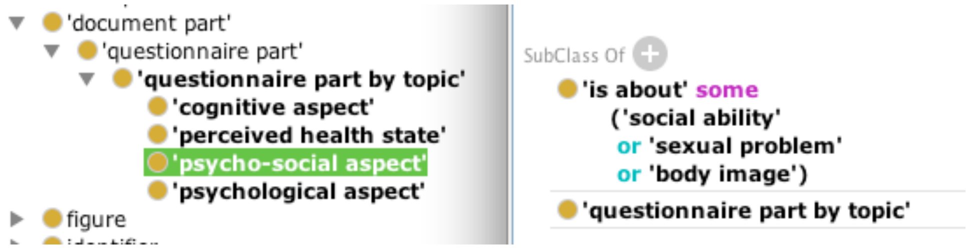


Figure 6 – 'Psycho-social aspect' and its topics

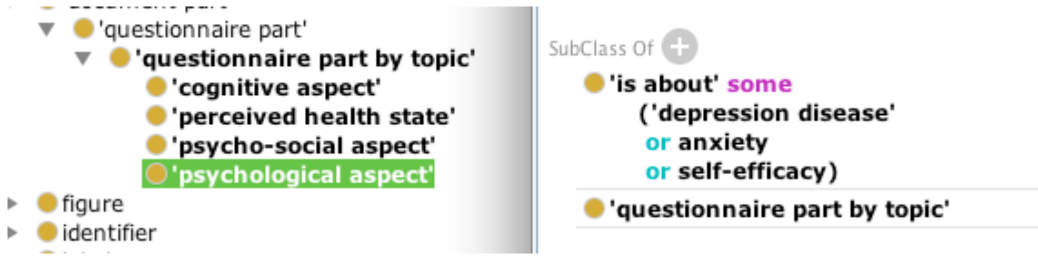


Figure 7 – 'Psychological aspect' and its topics

From these pictures it is clear that the data structure of the GUI proposed in the figure 3 is not the same structure of the ontology. Indeed, in the GUI we restructure the data so that they can be easily accessed and retrieved by the users.

As a use case, let us suppose that the user is interested in acquiring knowledge about the ALGA category 'psycho-social aspect'. First of all, she gets the information provided in the ontology for this class, i.e. the annotations defined for the class; then, she can decide which subcategory of the 'psycho-social aspect' she is interested in, e.g. the category 'sexual problem'. According to the specifications given in the ontology, she can get more information about delayed ejaculation, erectile dysfunction, retrograde ejaculation and dyspareunia. Moreover, if the different axioms of the ontology are included in the structure of the GUI, then the user can access even more information, e.g. 'sensory perception of pain during sexual intercourse' related to 'dyspareunia'. In this way, the user is able to access the structure of the ontology and the information stored in it without being aware of its logical complexity.

Preliminary black box evaluation of HDOT_PEM

In the deliverable D.4.1 (Sanfilippo et al., 2012) we defined how an ontology can be evaluated *in sensu* black box according to the specifications given by Gangemi et al. (Gangemi et al., 2005). Basically, this process consists in evaluating the suitability and the usefulness of the ontology to the necessities of the daily work. Indeed, if an ontology is developed for a specific purpose (i.e. to improve the semantic interoperability between different databases), it is fundamental that it can be successfully applied to the solutions of specific problems at hand.

At the current state, it is not possible to provide a definitive evaluation of HDOT_PEM from this perspective, because the module has not been integrated in a software architecture yet. However, in the previous paragraph we define a strategy for its integration in a GUI to show the ALGA results. Therefore, we define in the following a preliminary black box evaluation of the module according to the structure of the GUI previously defined, the ALGA questionnaire and the interaction of the module with the IEmS.

According to the strategy defined in the D.4.1, a black box evaluation consists in two main components: 1) suitability for semantic interoperability and 2) real context usability. The first point regards the possibility that users have to cope with the semantic disorganization in the communication between different databases among the same entities by using the ontology. The second one is a more general requirement that regards the response of the ontology to the users' needs. Taking for granted the patient empowerment scenario, the usability can be defined in the terms of the possibility that patients have to easily access, retrieve and understand their clinical data by using the data structure provided in the ontology. Moreover, within a patient empowerment framework, we need to evaluate 3) the patient-orientation of the ontology, i.e. whether the information stored in its structure can be easily accessed and retrieved by patients. Lastly, we take into account 4) the interaction between the ontology and the IEmS, i.e. whether the ontology is spendable for the development of a tool for patient empowerment. The following picture (figure 8) sumps up the black box evaluation cycle for HDOT_PEM.

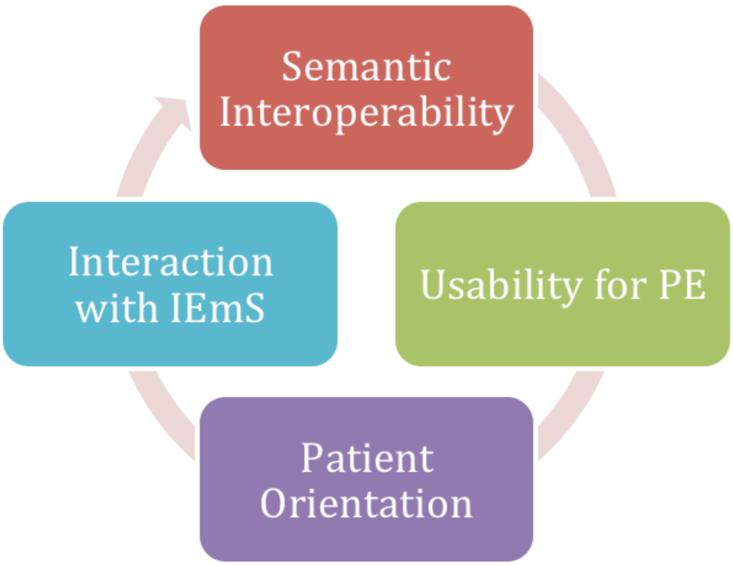


Figure 8 – Black-box evaluation cycle for the HDOT_PEM

Black box evaluation for HDOT_PEM:

- Semantic interoperability: HDOT_PEM is encoded in OWL 2 and integrates under the same semantic umbrella different external ontologies, most of which are part of the OBO Foundry initiative. Its backbone taxonomy is built using the subsumption relation between classes of different generality, and in some case other owl axioms are included. Moreover, its structure consists in a domain- and task-driven specification of an upper-level ontological framework (i.e. the one defined by BFO 2.0), by which it is able to disambiguate complex concepts and to foster the interoperability with other semantic resources. Therefore, we are confident that high levels of interoperability can be achieved by its exploitation in database annotation;

- Usability for patient empowerment: HDOT_PEM includes different class annotations (definitions, comments, examples of usage, link to external resources, patient language definitions), so that its semantic specifications can be easily understood by laymen and not only by ontology designers. In the GUI previously defined users have the possibility to browse their data, together with the annotations defined in the ontology. Moreover, since the ontology links different terms to each other, users have the possibility to access all the desired pieces of information, without being aware of the logical structure of the ontology. Therefore, we are confident that the ontology can be successfully exploited for the benefits of the patient empowerment;

- Patient-orientation: In the deliverable D14.1 (McVie et al., 2012), it is documented that 163 patients (51% women and 49 % men) with different education backgrounds were involved in the ALGA survey with the purpose of understanding their level of health literacy and therefore the modalities the professionals should use to interact with them. Moreover, the deliverable D2.5 (Pravettoni at al., 2012) documents the patient-orientation of the ALGA, i.e. the idea that its questions have been formulated in such a way that patients feel comfortabe in answering them[22]. In the development of HDOT_PEM and the ontological patient view of the questionnaire, we model the ALGA according to the principles of ontology design, while keeping at the same time its patient-orientation. Particularly, we keep the same term wording, so that if patients are able to understand the ALGA, they will be able to understand its ontological representation. Moreover, we add lay definitions to different classes in the ontology, so that the medical technicalities are put aside and information is presented in a lay language. These definitions have been developed at the IEO and in the next paragraph we provide more information about their development and evaluation. Therefore, we are confident that the ontology and its patient view are reliable for a patient-friendly software application;

[22] Cf. D.2.5 (Pravettoni, Gorini, & Zasada, 2012), paragraph 3.3.2

- Interaction with the IEmS: The categories of the ALGA play a central role in the definition of the users' profiles for the IEmS. Inasmuch as HDOT_PEM contains the ontological representation of the ALGA, its semantic references can be used as the main linguistic component of the IEmS, and in the in the previous paragraph we show how the ontology can be used for the development of a GUI. We are confident that the structure of the ontology can be easily exploited to the IEmS, although more users' requirements must be provided, so that we can provide the best ontological solution for the benefits of the IEmS.

It is worth noting that this evaluation process is just a preliminary one and a complete evaluation shall be defined once the module is integrated in a software component to achieve specific goals.

Development and evaluation of the lay definitions in the linguistic schema

In order to develop and evaluate the definitions of terms that will be included in the HDOT library of ontologies[23] we decided to start from definitions made by lay people and later ask few doctors to evaluate those lay definitions.

One possible way to proceed was to start from definitions by experts and to ask lay people how understandable they were. The main advantage of this procedure is that we are certain that the provided definition is correct. A possible disadvantage could be that starting from technical language we stuck too much to technical wording without being able to really use lay language.

An alternative way to proceed was to ask lay people to give a definition of each term to be included in the linguistic schema. This way would have allowed to actually use lay language.

We decided for this second procedure, with the awareness that definitions could be wrong and that we had to check them with professionals. Six people that don't work in psychological or medical field were contacted separately and asked to give their definitions of each presented word. After that, two physicians checked independently those definitions, in order to highlight possible errors and to indicate the most adequate definitions given by lay people. When there was no consistency between the two doctors' judgment we decided for the most frequent correct wording of the 6 lay people.

[23] https://code.google.com/p/hdot/source/browse/#svn%2Ftrunk

6. Conclusion

In the deliverable we stress that one of the main purposes of a patient empowerment strategy is to help patients to understand the clinical information about their health status. Indeed, patients are more willing to actively interact with the health care providers and more responsive in the decision making processes concerning their disease(s), when clinical information is presented in a more understandable way.

In order to improve patients understanding, there is the need to avoid medical jargon as far as possible. Information shall be presented with the use of an empirical-tested lay language, by which patients feel comfortable in understanding what it is meant by the health care providers and are not annoyed with medical technicalities. However, a lay language shall simplify the interaction between patients and medical experts, while being at the same time consistent with the medical knowledge nowadays at hand. Therefore, there is the need to fix objective references for different terms labels and explanations, so that information can be consistently shared and presented among different people and institutions.

According to the need of a lay language for medical terms, we extract from the ALGA questionnaire a data schema and provide lay definitions for each term. These definitions have been developed and tested in real-world contexts, together with experts, who checked their quality according to their biomedical knowledge, and patients, who checked how easily they understood them.

The data schema has served as knowledge base for the development of a modular ontology, HDOT_PEM, which enriches the ontology library of p-medicine. The main purpose of this new module is to present the data collection of p-medicine in a patient-oriented way. Indeed, we include the lay definitions in its structure, so that information is presented in a patient friendly manner, while at the same time it maintains a stable reference among the semantic space of HDOT.

Moreover, in order to evaluate HDOT_PEM, we define a strategy for the development of an ontology-driven patient view, i.e. a way to exploit the semantic references of the ontology in a GUI. We aim at presenting the structure of the ALGA by the use of the ontology, so that its structure can be integrated and exploited for the needs of the patient empowerment services, i.e. the IEmS.

However, it is worth noting that while ontologies can be used in patient empowerment scenarios to facilitate the laymen understanding of biomedical terminologies and clinical data, they cannot achieve such purposes by themselves. Ontologies 1) disambiguate the meaning of complex medical terms between experts and laymen; 2) state clearly whether two different terms have the same semantic reference; 3) can be further developed and reused for different applications, while maintaining their main backbone structure; 4) can be enriched using class synonyms and definitions without sacrificing the consistency of the ontological representation. Nevertheless, ontologies cannot make patients fully understand data and in our opinion this strictly depends on the use of tools explicitly designed to enhance patients understanding, as well as educational and dissemination activities.

Bibliography

Beck, T., Gollapudi, S., Brunak, S., Graf, N., Lemke, H., Dash, D., et al. (2012). Knowledge Engineering for Health: A New Discipline Required to Bridge the "ICT Gap" Between Research and Healthcare. *Human Mutation , 33* (5), 797-802.

Ceusters, W., & Smith, B. (2010). Foundations for a realist ontology of mental disease. *Journal of Biomedical Semantics , 1* (10).

de Bono, B., Hoehndorf, R., Wimlaratne, S., Gkoutos, G., & Grenon, P. (2011). The RICORDO approach to semantic interoperability for biomedical data and models: strategy, standards and solutions. *BMC Research Notes , 4* (313).

Fisch, K., M., & O'Connel, C. A. (1998). Creation of a dictionary of medical and lay terms for use in the preparation of patient information leaflets. *Drug Information Journal , 32*, 533-538.

Gangemi, A., Catenacci, C., Ciaramita, M., & Lehmann, J. (2005). A theoretical framework for ontology evaluation and validation. *Proceedings of SWAP2005 .* Spring.

Graph, N. (2011). D.2.2, Definition on scenarios and use cases and report on scenario based user needs and requirements. p-medicine.

Hunter, L. (2002). Ontologies for programs, not people. *Genome Biology , 3* (6), 1006.1-1006.2.

Mazzocco, K., Pravettoni, G., Gorini, A., Gordon, M., Schwarz, U., & Sanfilippo, E. M. (2012). *D14.2 Specification of the linguistic schema.*

McVie, G., Pravettoni, G., Gorini, A., & Mazzocco, K. (2012). D14.1, Specification of patient involvement demands in the p-medicine scenarios.

Pravettoni, G., Gorini, A., & Zasada, S. (2012). D2.5, Specification of tools and services supporting patient empowerment.

Sanfilippo, E. M., Strößner, C., & Schwarz, U. (2012). D4.1 List of domain quality-checked ontologies and initial release of HODT and its Glass Box Evaluation.

University of Iowa. (n.d.). Medical Terms in Lay Language. Alternative Lay Language for Medical Terms in Consent Forms .

Venes, D. (Ed.). (2001). *Taber's cyclopedic medical dictionary.* Philadelphia: F.A. Davis Company.

Appendix 1- Data schema for ontological purposes

In the following table we extend and modify the data schema prototype provided in the deliverable D14.2, organizing the information in such a way that it can be spendable for the development of HDOT_PEM.

The schema has the following structure:

a. ALGA term: the term taken from the ALGA;
b. HDOT class: the class in the ontology corresponding to the term in (1).
c. HDOT parent class: HDOT class subsuming the class in (2);

The classes in (2) and (3) can refer to the different modules of HDOT (HDOT_CORE, HDOT_PM, HDOT_PFM, HDOT_PEM) as well as to different external ontologies, e.g. GO, OGMS, PATO, MF, MFOEM, etc.

d. Location: the location of the new class in the library of HDOT, i.e. in which module the class can be found;
e. Class definition: the class definition according to the structure of HDOT;
f. Class definition in a patient language: a class definition in a lay language;

In this context a lay language is considered as a subset of natural language (English in this case) by which complex biomedical meanings are presented to laymen in a non-technical manner as far as possible.

Appendix 2 - Abbreviations and acronyms

BFO	Basic Formal Ontology
DL	Description Logic
DOID	Human Disease Ontology
GUI	Graphical User Interface
HDOT	Health Data Ontology Trunk
HDOT_PM	Health Data Ontology Trunk p-medicine Module
HDOT_PFM	Health Data Ontology Trunk Pathological Formation Module
HDOT_PEM	Health Data Ontology Trunk Patient Empowerment Module
IAO	Information Artifact Ontology
IEmS	Interactive Empowerment Service
MD	Mental Disease Ontology
MFO	Mental Functioning Ontology
MFOEM	Emotion Ontology
OGMS	Ontology for General and Medical Science
OWL	Ontology Web Language
PATO	Phenotypic Quality Ontology
RDFS	Resource Description Framework Schema
URI	Uniform Resource Identifier
W3C	World Wide Web Consortium

ALGA term	Class under HDOT	Parent class in HDOT	Location	Class definition	Class definition in patient language
Ability (physical)	HDOT_PM: Physical ability	HDOT_CORE: ability	HDOT_PM	Physical ability is an ability that enables one to perform some physical act	Physical ability is the ability to do something efficiently with the body
Ability (reasoning)	HDOT_PEM: Ability to reason	HDOT_CORE: Ability	HDOT_PEM	Ability to reason in an ability that is realized in a reasoning process	Reasoning ability is the ability to use information to reflect on a problem
Ability (problem-solving)	HDOT_PEM: Problem-solving ability	HDOT_CORE: ability	HDOT_PEM	A problem-solving ability is an ability that enables one to find solutions for specific problems	Problem-solving ability is the ability to find a non immediate solution to a specific actual situation
(social) Ability	HDOT_PM: Social ability	HDOT_CORE: ability	HDOT_PM	Social ability is an ability that facilitates interaction and communication with other people	Social ability is the ability to create and keep up good relationships
Anemia	HDOT_PM: Anemia	HDOT_CORE: disorder	HDOT_PM	Anemia is a disorder in which the body does not have enough healthy red blood cells	Anemia is a reduction of red blood cells

Anger	MFOEM: Anger	MFOEM: emotion process	HDOT_PEM	Anger is a negative emotion characterized by feelings of unpleasantness and high arousal in the form of antagonistic feelings and action tendencies	Anger is an emotion a person feels toward a situation or a person that can lead to aggressive or violent behaviors
Anxiety	MFOEM: Anxiety	MFOEM: emotion process	HDOT_PEM	Anxiety is a negative emotion provoked by the prospect of future threats	Anxiety is an emotional state felt in uncertain situations that could lead to negative consequences
Appetite	HDOT_PM: Appetite	HDOT_CORE: disposition	HDOT_PM	Appetite is a disposition that specifies the desire for food	Appetite is a desire to eat caused by a light feeling of empty stomach
(Loss of) Appetite	HDOT_PEM: Loss of appetite	OGMS: Bodily process	HDOT_PEM	Loss of appetite is a bodily process that is the decreased sensation of appetite.	Loss of appetite is the feeling of full stomach that leads not to eat
(Loss of) Appetite cancer-related	HDOT_PEM: Cancer-related loss of appetite	OGMS: Bodily process	HDOT_PEM	Cancer-related loss of appetite is a loss of appetite that is caused by an instance of cancer occurring in the organism	Cancer-related loss of appetite is the feeling of full stomach that leads not to eat caused by a cancer in the organism
Attention	MF: Attention	MF: cognitive process	HDOT_PM	Attention is the act or process of concentrating, focusing the attention and mental energy on a single object to the	Attention is the ability to concentrate on something for a relative long time

				exclusion of others.	
(self-efficacy) Belief	HDOT_PEM: Appraisal of self efficacy	MF: cognitive representation	HDOT_PEM	Self-efficacy appraisal is a cognitive representation that represents a judgment about one's own ability to complete autonomously tasks and reach goals	Self-efficacy belief is the belief to be autonomous and independent from others and to be useful to herself
Belief about the body	HDOT_PEM: Appraisal of body	MF: cognitive representation	HDOT_PEM	Appraisal of body is a cognitive representation which represents a judgment about one's own body	Body image is how a person perceives her own body
Body image after surgery	HDOT_PEM: Appraisal of body after surgery	MF: Cognitive representation	HDOT_PEM	Appraisal of body after surgery is one's appraisal of her body after a surgical procedure	Body image after surgery is a different way to perceive her own body after surgery
Cancer	DOID: Cancer	OGMS: Disease	HDOT_PFM	A disease of cellular proliferation that is malignant and primary, characterized by uncontrolled cellular proliferation, local cell invasion and metastasis.	Cancer is a destructive disease caused by a non normal grow of body cells

Cancer of breast	DOID: Breast cancer	OGMS: Disease	HDOT_PFM	A thoracic cancer that originates in the mammary gland.	Breast cancer is a destructive disease caused by a non normal grow of body cells in breast tissues
Cancer of lung	DOID: Lung cancer	OGMS: Disease	HDOT_PFM	A respiratory system cancer that is located in the lung.	Lung cancer is a destructive disease caused by a non normal grow of body cells in the lungs
Carcinoma	DOID: Carcinoma	OGMS: Disease	HDOT_PFM	A cell type cancer that has_material_basis_in abnormally proliferating cells derived_from epithelial cells.	Carcinoma is a malignant disease caused by a non normal grow on the tissue that covers the body and its organs
Carcinoma of breast	DOID: Breast carcinoma	OGMS: Disease	HDOT_PFM	A breast cancer that derives_from breast tissue.	Breast carcinoma is a malignant disease caused by a non normal grow in the breast tissue
Carcinoma of lung	DOID: Lung carcinoma	OGMS: Disease	HDOT_PFM	A lung cancer that is located_in the lungs and has_symptom cough and has_symptom chest discomfort or pain and has_symptom weight loss and has_symptom hemoptysis.	Lung carcinoma is a malignant disease caused by a non normal grow in the lung tissue

Chemotherapy	HDOT_PFM: Chemotherapy	OGMS: Treatment	HDOT_PFM	Chemotherapy is the therapeutic process of cancer with an antineoplastic drug or with a combination of such drugs into a standardized treatment regimen	Chemotherapy is a chemical therapy used to destroy in a selective manner the pathological appearance
Close mindedness	HDOT_PEM: close mindedness	BFO: disposition	HDOT_PEM	Close mindedness is a disposition of some extended organism that is realized in a cognitive process by which someone is intolerant of the beliefs and opinions of other people	Close mindedness is the inability to consider other possibilities, believes and points of view
Cognitive closure	HDOT_PEM: cognitive closure	BFO: disposition	HDOT_PEM	Cognitive closure is a disposition of some extended organism that is realized in a cognitive process, which indicates one's desire for definite knowledge on some issue and the eschewal of confusion and ambiguity	Cognitive closure is the condition in which a person looks for an answer whatever it is, that is preferable to uncertainty and ambiguity

Depression	MD: Depression	DOID: disease of mental health	HDOT_PEM	Depression is a disease of mental health that is realized in a depressed mood episode	Depression is a long lasting mood change, characterized by mild to deep sadness, reduced interest or pleasure in activities, social withdrawal with associated low self-esteem and fatigue
Depression (cancer-related)	HDOT_PEM: Cancer-related depression disease	DOID: disease of mental health	HDOT_PEM	Cancer related depression is a disease of mental health that is realized in a depressed mood episode because of an instance of cancer in the organism	Cancer related depression is a long lasting negative mental state, characterized by sad mood and apathy caused by the presence of cancer
Distress (psychological)	HDOT_PEM: psychological distress	MF: cognitive process	HDOT_PEM	Psychological distress is a cognitive process that refers to the inability of someone to adapt completely to stressors and shows maladaptive behaviors	Distress is a condition in which a person has difficulties in coping with a long-lasting stressful and frustrating situation
Dysfunction (cognitive)	HDOT_PM: Cognitive dysfunction	HDOT_CORE: dysfunction	HDOT_PM	Cognitive dysfunction is a dysfunction of cognitive processes	Cognitive dysfunction is a reduced ability of learning, thinking and reasoning

Dysfunction (sexual)	HDOT_PM: Sexual dysfunction	HDOT_CORE: dysfunction	HDOT_PM	Sexual dysfunction is a physical dysfunction of the sexual apparatus	Sexual dysfunction is a condition in which sexual apparatus does not work in a proper way
Dysfunction (erectile)	HDOT_PEM: Erectile dysfunction	HDOT_PM sexual dysfunction	HDOT_PEM	Erectile dysfunction is a sexual dysfunction that is characterized by the inability to develop or maintain an erection of the penis during sexual performance	Erectile dysfunction is a difficulty for the male to have and maintain penis erection during sexual performance
Ejaculation (going backward into the bladder)	HDOT_PEM: Retrograde ejaculation	HDOT_PM: sexual dysfunction	HDOT_PEM	Retrograde ejaculation is a sexual dysfunction that occurs when semen, which would normally be ejaculated via the urethra, is redirected to the urinary bladder	Retrograde ejaculation is a male condition in which semen goes in the bladder instead to go out from the penis
Fatigue (cancer-related)	HDOT_PEM: Cancer related fatigue	BFO: disposition	HDOT_PEM	Cancer-related fatigue is a disposition of some extended organism due to an instance of cancer	Cancer related fatigue is a mental and/or physical feeling of tiredness and lack of energy that prevents to conduct normal daily activities

Fear	MFOEM: Fear	MFOEM: emotion process	HDOT_PEM	Fear is an emotion process that motivates attempts to cope with events that provide threats to the survival of well-being of organisms	Fear is a negative emotion felt in presence of threatening and dangerous that motivates the person to fight or to escape.
Feeling depressed	HDOT_PEM: feeling depressed	MF: cognitive representation	HDOT_PEM	Feeling depressed is a cognitive representation that specifically depends on some extended organism who participates in a depressed mood episode	Feeling depressed is condition in which a person presents a long lasting mood change, characterized by mild to deep sadness, reduced interest or pleasure in activities, social withdrawal with associated low self-esteem and fatigue
Feeling guilty	HDOT_PEM: feeling guilty	MF: cognitive representation	HDOT_PEM	Feeling guilty is a cognitive representation that specifically depends on some extended organism and refers to the subjective emotional feeling to have a sense of personal responsibility	Feeling guilty is a condition in which a person feels responsible for something bad and adverse that happened after a her choice or action which negative consequences are unintentional

Feeling hopeless	HDOT_PEM: feeling hopeless	MF: cognitive representation	HDOT_PEM	Feeling hopeless is a cognitive representation that specifically depends on some extended organism and refers to a condition of being without hope because there seems to be no possibility of comfort or success	Feeling hopeless is a condition in which a person feels as she has no possibilities to reach a positive outcome
Feeling helpless	HDOT_PEM: feeling helpless	MF: cognitive representation	HDOT_PEM	Feeling helpless is a cognitive representation that specifically depends on some extended organism and refers to the subjective emotional feeling of being without help	Feeling helpless is a condition in which a person perceives to be without support or protection
Feeling nervous	MFOEM: feeling nervous	MF: cognitive representation	HDOT_PEM	Feeling nervous is a cognitive representation that specifically depends on some extended organism and refers to the subjective feeling of being not at ease, of being anxious or agitated.	Feeling nervous is a condition in which a person feels impatient, tens and anxious, not confortable and calm
Feeling stressed	HDOT_PFM: feeling stressed	MF: cognitive representation	HDOT_PEM	Feeling stressed is a cognitive representation that specifically depends on some extended organism and refers to the feeling of	Feeling stressed is a condition in which a person perceives she has no more resources to cope with events and situations

				being tensed resulting from factors that tend to alter an existent equilibrium	
Feeling weak	MFOEM: Feeling weak	MF: cognitive representation	HDOT_PEM	Feeling weak is a cognitive representation that specifically depends on some extended organism and refers to the subjective feeling of weakness, lack of energy and/or capability, faintness, absence of strength.	Feeling weak is a physical and psychological condition in which a person feels that she has no energy to cope with events or to engage in daily activities
Grief	MFOEM: Grief	MFOEM: emotion process	HDOT_PEM	Grief is an intense negative emotion following a bereavement, i.e. the loss of a significant person through that person's death	Grief is an intense long lasting negative emotion caused by a subjective important loss
Inability to ejaculate	HDOT_PEM: Delayed ejaculation	HDOT: sexual dysfunction	HDOT_PEM	Delayed ejaculation is a medical condition in which a male cannot ejaculate, either during intercourse or by manual stimulation with a partner.	Inability to ejaculate is a male difficulty in emitting the semen during otherwise normal sexual activity

Irritation	MFOEM: Irritation	MFOEM: emotion process	HDOT_PEM	An unpleasant emotion closely related to anger but lower in intensity and without the moral dimension of blame and seriousness that is implicated in anger.	Irritation is an emotional reaction in which it is difficult for the person to accept situations or events that are usually accepted and it is associated with emotions of anger and annoyance
Leukemia	DOID: Leukemia	OGMS: Disease	HDOT_PFM	A cancer that affects the blood or bone marrow characterized by an abnormal proliferation of blood cells	Leukemia is a malignant disease in which blood cells grew in an abnormal way causing a decrease in the immune system
Loss of desire for sexual activity	HDOT_PEM: Sexual desire loss	OGMS: Bodily process	HDOT_PEM	Sexual desire loss is bodily process by which one loses interest in sexual intercourses	Loss of sexual desire is decrease in the desire to have sexual intercourses
Memory	GO: Memory	MF: cognitive process	HDOT_PM	Memory is a cognitive process that is involved in the mental information processing system that receives, modifies, stores and retrieves	Memory is a person ability to store, hold and retrieve information about people, events and situations in general

				informational stimuli	
(Benign) Neoplasm	DOID: Benign neoplasm	OGMS: disease	HDOT_PFM	A disease of cellular proliferation that results in abnormal growths in the body which lacks the ability to metastasize.	Benign Neoplasm is an abnormal grow of tissues in the organism that harms healthy tissues, where the grow is localized and do not spread over the rest of the body
(Cancer-related) Pain	HDOT_PEM: Sensory perception of cancer-related pain	GO: sensory perception of pain	HDOT_PEM	Sensory perception of cancer-related pain is a sensory perception of pain that is caused by an instance of cancer in the organism	Cancer related pain is a sensation of subjective physical sufferance caused by a cancer harming the organism
Pain during sexual intercourse	HDOT_PEM: Sensory perception of pain during sexual intercourse	GO: sensory perception of pain	HDOT_PEM	Sensory perception of pain during sexual intercourse is a pain that is experienced during sexual intercourses	Pain perception during sexual intercourse is a sensation of subjective physical sufferance on sexual apparatus during sexual intercourse
Pain perception	GO: Sensory perception of pain	MF: cognitive process	HDOT_PM	The series of events required for an organism to receive a painful stimulus, convert it to a molecular signal, and recognize and characterize the	Pain perception is a sensation of subjective physical sufferance in presence of something that is harming the organism or

				signal.	reducing its well being
Quality of life	HDOT_PM: Quality of life	PATO: organismal quality	HDOT_PM	The subjective measurement of an individual's sense of well-being and ability to enjoy life.	Quality of life is a person's subjective perception on how his/her life condition is in a specific moment and a specific culture
Participant of questionnaire	HDOT_PM: Questionnaire participant	OMRSE: role in human social processes	HDOT_PM	Questionnaire participant is a role that indicates those people who participate in a survey	Questionnaire participant is an informed person who volunteers to answer a set of questions
Pathological formation	OGMS: Pathological formation	HDOT_CORE: material pathological entity	HDOT_PFM	A non-canonical part of an anatomical structure that results from a pathological process	Pathological formation is an abnormal grow of tissues in the organism that harms healthy tissues (neoplasm)
Risk of mortality	HDOT_PM: Mortality risk	PATO: population quality	HDOT_PM	Mortality risk is a population quality that indicates the risk that death will occur in a population	Mortality risk is an information about the probability to dye
Rumination	HDOT_PEM: Rumination	MF: cognitive process	HDOT_PEM	Rumination is a cognitive process characterized by thinking deeply about something	Rumination is a person's tendency to iteratively, constantly and negatively think to his/her problems and emotional

					sufferance without be able to stop
Sadness Toxicity (chemotherapy-induced)	MFOEM: Sadness HDOT_PFM: Chemo-therapy induced toxicity	MFOEM: emotion process HDOT_PM: Chemical toxicity	HDOT_PEM HDOT_PFM	Sadness is a negative emotion felt when an event is appraised as unpleasant, obstructive to one's goals and concerns, and one feels unable to cope with it or modify it Chemotherapy induced toxicity is the toxicity in an organism due to a chemotherapeutical process	Sadness is a negative emotion caused by characterized by a feeling of melancholia and discouragement for something that the person is not able to reach or have Chemotherapy induced toxicity is the level of poisoning in the organism due to chemotherapy