Grant Agreement No: 611366

PREventive Care Infrastructure based on Ubiquitous Sensing

Instrument: Collaborative Project

Seventh Framework Programme (FP7) Call FP7-2013-10

D4.4 Data fusions, analysis and semantic quality solutions

Due date of deliverable: **01/02/2016**

Actual submission date: **01/02/2016**

Start date of project: November 1st 2013
Duration: 36 months
Project Manager: Professor Jörg Ott
Revision: 1.0

Nature:	R
Dissemination Level:	PU
Version:	1.0
Date:	01.02.2016
WP number and title:	WP4: System Sensors and feedbacks tools
Deliverable leader	Institut Mines-Telecom (IMT)
Authors	Philippe Tanguy (IMT), Christophe Lohr (IMT), Daniela Ramsauer (UNIVIE), Peter Kalchgruber (UNIVIE), Tero Myllymäki (Firstbeat), Todor Ginchev (AALTO), Edward Mutafungwa (AALTO), José Costa (AALTO)
Status:	Final

Document History

Date	Version	Status	Change
10.11.2015	0.1	Draft	Initial ToC
19.12.2015	0.2	Draft	Revised ToC for partner comments
18.01.2016	0.3	Draft	Partner contributions
29.01.2016	0.4	Draft	First Draft of full document
29.01.2016	0.5	Draft	Draft for QEG
01.02.2016		Draft	Included QEG Comments & Suggestions
01.01.2016	1.0	Final	Final amendments

Peer Review History

Date	Version	Reviewed By
01.01.2016	1.0	QEG

Abstract

This document represents the data analysis, data fusion and semantic quality deliverable. Its main objectives are to describe the methods employed to accomplish the data fusions, analysis and semantic quality which are solutions composing the Virtual Individual Model (VIM). A controlled vocabulary has been built with actors of the PRECIOUS project. It provides a common understanding of the data to all partners (users, developers, experts, and health staff), establishes relationships to existing projects and data sources focused on e-health, allows the monitoring and maintaining of quality issues of the vocabulary, harmonizes data from different sensors and input providers, and the usage of a standardized data model for the entire project. Then, a description per domain-knowledge have been proposed to detail the semantic data extraction from low level context such as a semantic analysis of textual data, a food analysis, a heart rate processing, an ambient sensors analysis and a mobile phone sensor data analysis.

List of Acronyms

Abbreviation	Meaning
AUC	Area Under Curve
API	Application Programming Interface
BCT	Behaviour Change Technique
BN	Bayesian network
CVDC	Controlled Vocabulary Development Cycle
DM	Data Models
EPOC	Excess post-exercise oxygen consumption
HITs	Human Intelligence Tasks
HL7	Health Level 7 international
HRV	Heart rate variability
IEQ	Indoor Environmental Quality
LIWC	Linguistic Inquiry and Word Count
LOINC	Logical Obsrvation Identifiers Names and Codes
LR	Logistic Regression
MESH	Medical Subject Heading
PLB	Psycholinguistic based
ROC	Receiver Operator Curve
SNOMEDCT	Systematized Nomenclature of Medicine
SVM	Support Vector Machine
UCA	User Context Awareness

URI	Uniform Resource Identifier
VIM	Virtual Individual Model
WHO	World Health Organization
xAAL	Home Automation Protocol developed at Telecom Bretagne

Executive Summary

The PRECIOUS system aims to promote healthy lifestyles, based on three main components: 1) transparent sensors for monitoring user context parameters and health indicators such as food intake, sleep, stress and physical activity 2) the development of a virtual individual model (VIM) representing users' variables and different parameters collected (both directly from the user and with sensors) for inferring health risks and desired behaviour changes, and 3) application of a motivational service design framework combined with gamification principles to trigger, monitor and sustain mid-to-long term behaviour change.

This document presents the data analysis, data fusion and semantic quality deliverable. To achieve the VIM, it provides a detailed description of data analysis methods to produce high level semantic data representing users' variables. A controlled vocabulary in form of a thesaurus has been built taking into account a common understanding of the data to all partners (users, developers, experts, health staff). It allows the monitoring and maintaining of quality issues of the vocabulary, harmonizes data from different sensors and input providers, and the usage of a standardized data model for the entire project.

In the section 3, the semantic interoperability using a controlled vocabulary is described. It provides a common understanding of the data for users, developers, experts and health staff. The controlled vocabulary is publicly available hosted on a server of the University of Vienna.

In the section 4, a semantic analysis of textual data is presented. It analyse the person's textual communication in social media to extract related mood information. The work shows that a semantic analysis of textual social media can be used to reduce the total sparsity of information and uncertainty of the mood identification process.

In the section 5, the food intake analysis is proposed with explanations of how the nutrient levels for a particular food amount or portion have been produced.

In the section 6, a detailed description of the heart rate analysis is reported. It is explained how the results of heart rate sensor data analysis can be visualized to the user in a form of graphs, bars, points and so on. It is related to the user physiological status regarding periods of stress, recovery, and physical activity in terms of duration and intensity.

In the section 7, the processing of raw mobile phone sensor data for physical activity monitoring and characterisation is described with a focus on the use of 3-axis accelerometer sensors integrated in smartphones.

In the section 8, ambient sensor data analysis are reported. In details, it is described how indoor environment quality variables, related to European norms or guidelines, have been combined and produced to achieve a high level of semantic. High level information, e.g. thermal comfort, could have an impact on the following risk factors: stress and sleep.

Table of Contents

1. Background and objectives

1.1. Background

The PRECIOUS project targets to develop a preventive care system to promote healthy lifestyles with specific focus on the following risk factors: environmental, socio-psychological and physiological. Which one are linked to the Type 2 Diabetes and Cardiovascular diseases.
Therefore, the PRECIOUS project aims to provide new healthcare solutions composed by:

- a transparent sensors/actuators layer allowing to gather seamlessly the user context (health & ambient data) to identify risk factors;
- a virtual individual model (VIM) representing users' variables;
- mobile applications from a motivational service based on gamification theories and motivational strategies to maintain behavioural change

1.2. Objectives

Previous research and previous deliverables within PRECIOUS indicate that lots of different information is necessary to derive divers health-related information about the user. Within this project a VIM (cf. section 2 deliverable 4.4 or more detailed in deliverable 3.2[1]) was created to define the attributes and their semantics representing a user in PRECIOUS. Semantics is derived from the Greek word sēmantikós, is also translated as 'significant' and is the study of meaning. In context of computer science Semantics is the assignment of human-understandable meaning to values computed (cf. Figure 1).

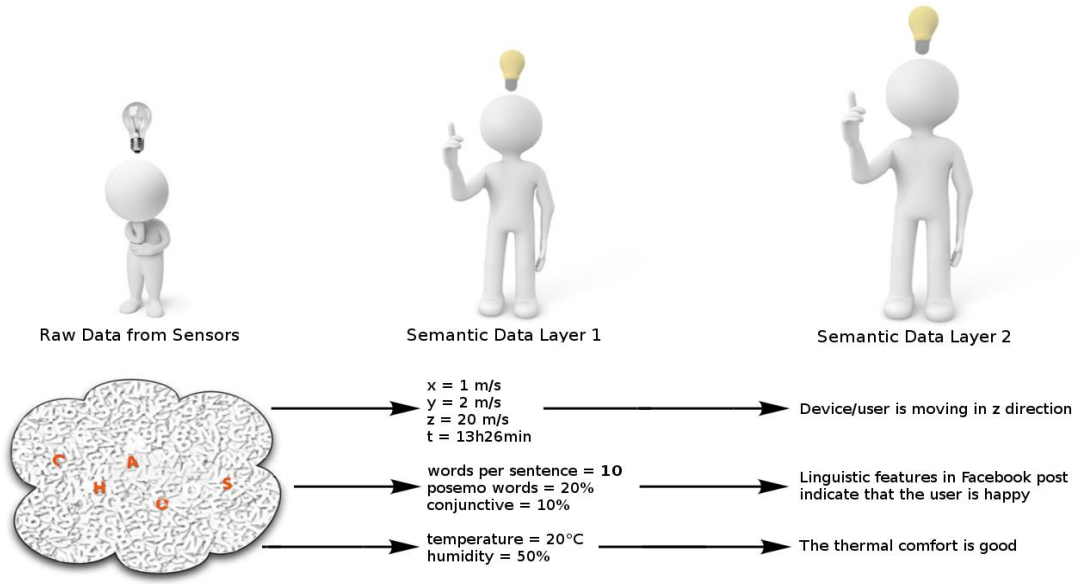

Figure 1: How to get semantic information from raw sensor data

[1] http://www.thepreciousproject.eu/wp-content/uploads/2013/12/D3-2-BehavioralRepresentationandVIM-FinalVersion-docx.pdf

Lots of data from different sensor is collected within PRECIOUS, which can be assigned to the attributes of the Virtual Individual Model. To map the raw sensor data to the attributes used in the VIM it is necessary to process and label the data in a semantic qualitative way. Because users need a meaningful interpretation of the data to understand the feedback of the PRECIOUS system.

Hereinafter examples are presented how the raw data from sensors is processed to obtain semantic information, which has a meaning for the users.

An example for semantic data analysis in PRECIOUS is the processing of social media messages to return information about the user's mood. In this case the sensor data is textual and the values obtained after the first processing of the raw data are linguistic and textual features of social media messages. In a second step these features are used to assign semantic information referring to the writer's mood to social media messages (cf. section 4).

Another example is raw accelerometer data that contains three decimal numbers describing the acceleration in either x-, y- or z-dimension and a time value. For non-expert users these values don't have a meaning per se, but experts can do further processing of this data to return semantic information to the user. This semantic information could be for example, that the user is not moving, walking or running at a particular time (cf. section 7).
But some information the user might need is not only dependent on the data of one sensor but on multiple. A further semantic layer is necessary to process the sensor data and give meaningful information to the users. An example can be found in section 8 where Thermal Comfort is depended from thermometer sensor data as well as from hygrometer data. The first layer of semantic interpretation is the knowledge that a thermometer returns the temperature in °C and the hygrometer returns the humidity of the air in %. The second semantic layer is the interpretation of the combination of these two values to get the semantic information about Thermal Comfort (cf. section 8).

During a PRECIOUS vocabulary collection phase, the PRECIOUS partners described (see D4.1 section 9.2) the raw data related to their domain-knowledge. It allows to have a common definition and collection of concept related to raw data gathered link to the user context @Home, @Work, @Mobility.

The present deliverable has for objectives to describe the methods employed to accomplish the data fusions, analysis and semantic quality which are solutions composing the VIM. The document is organized in several main sections related to each domain-knowledge experts: semantic analysis of textual data, food analysis, heart rate processing, ambient sensors analysis, mobile phone sensor data analysis.

2. Overview of the Virtual Individual Model (VIM)

The high-level diagram of the PRECIOUS system is depicted Figure 2. The heart of the system is the VIM component whose variables are stored in the backend database of the PRECIOUS cloud server. It collects and processes data from different sensors: ambient (temperature, humidity, etc.), Firstbeat heart rate variability (HRV) sensor, wearable devices (smart watch), mobile applications, etc.

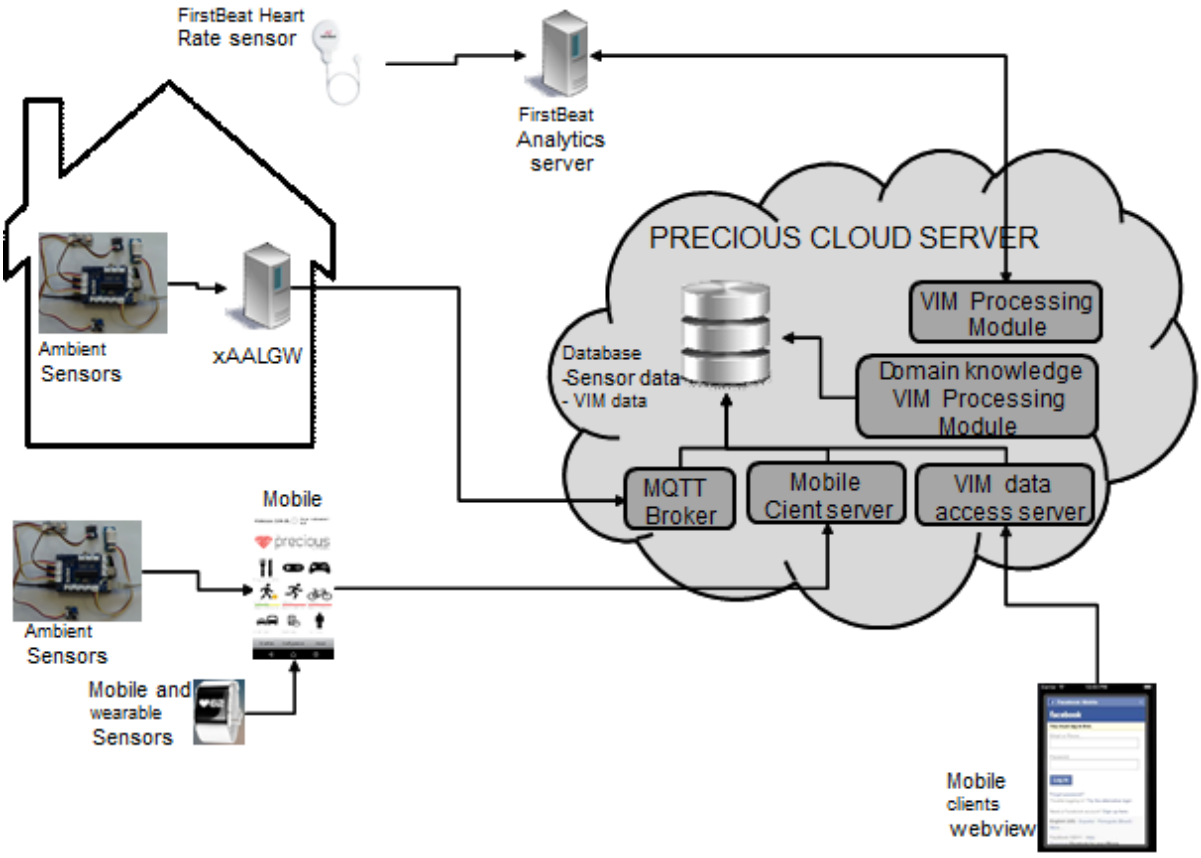

Figure 2: High-level diagram of PRECIOUS system implementation

The VIM variables are precisely described in the deliverable D3.2 - Final report on behavioral representation and virtual individual modelling. The present document will describe the data analysis methods to obtain those variables.

3. Semantic Interoperability Using Controlled Vocabulary

3.1. Background of Controlled Vocabularies

People can't share knowledge if they don't speak a common language (Davenport et al. 1998).

It is important to share the same meaning and interpretation of words, to enable the creation of a system, build upon it. The extension of the Web with semantics also allows applications to share knowledge and requires therefore a common language. Controlled vocabularies are one way to help people and applications finding a common language. Controlled vocabularies are "a standardized, restricted set of defined terms designed to reduce ambiguity in describing a concept" [9].

In the context of PRECIOUS it is very important to talk about the same values especially in the field of sensors. For example, one sensor expert might refer to the heart rate, as a value having a number of beats per minute, calculated from the power flow of electrodes, while another expert says that the heart rate is calculated from various pulse sensors. Both input variables measure the identical source of data, since the pulse is also a result of the heart rate. Will the developers, medical staff or the users understand the difference? Without explicitly defined standard terms, the creation of a single data model and development of a virtual individual model are difficult since the system engineers require the knowledge of all domains.

Figure 3- Overview of partner collaboration, dependencies and domain knowledge hosts

The ANSI/NISO Z39.10 standard describes the use of controlled vocabulary as the improvement of “the effectiveness of information storage and retrieval systems…” [20]. The PRECIOUS project consists of seven participating partners and therefore is absolutely suitable for use controlled vocabularies. The coordination and semantic distinction of the terminology could happen at many points of contact in the PRECIOUS project. Figure 3 gives an impression about overlapping work areas of all project contributors.

The need for a controlled vocabulary is not only to help individual experts of the various domains of the project to find a common understanding of their subject-specific terms. In order to develop data analysis, data fusion and data processing algorithms and rules based on basic input data and basic sensor data, a conceptual semantic understanding of the data input sources and a common understanding and interpretation of the data is indispensable. Since the domain knowledge is divided among all partners located at different physical places a special development process for the creation and the maintenance of the vocabulary (see section 3.2) was needed.

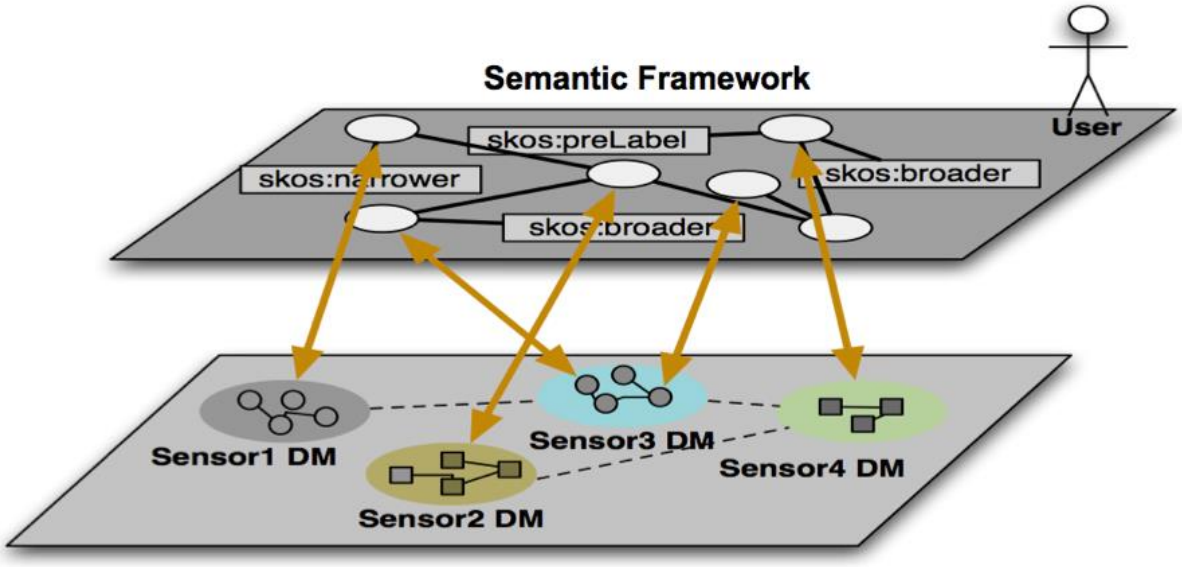

Figure 4 - Semantic Layer of PRECIOUS

In order to get a clear understanding we want to distinguish controlled vocabularies, taxonomies and ontologies at this point. A controlled vocabulary is a consistent list of terms. It has no structure, no relationships, but synonyms, allows indexing and allows the avoidance of homographs. A taxonomy is a special type of a controlled vocabulary. Taxonomies are hierarchically and enable terms to have relations such as broader, narrower and related. Thesauri are another special type of a controlled vocabularies. A thesaurus does not only have hierarchical structures. All terms in a thesaurus have relationships to other terms and allows further the description of the terms in much more detail. Besides hierarchical terms, associatives, equivalents, scope notes, explanations and examples can be annotated with concepts. For PRECIOUS we decided to use a controlled vocabulary in form of a thesaurus, to allow the integration and mapping of the most real world parameters and explanations into the vocabulary.

“A controlled vocabulary is a way to insert an interpretive layer of semantics between the term entered by the user and the underlying database to better represent the original intention of the terms of the user” (Leise 2002) Figure 4 shows one of the top layer views of the controlled vocabulary in the PRECIOUS project. The overlay of the semantic layer above all partly proprietary input data models (DM) from sensors, allows a common global understanding independently of the type, producer and language of the input data from the sensors.

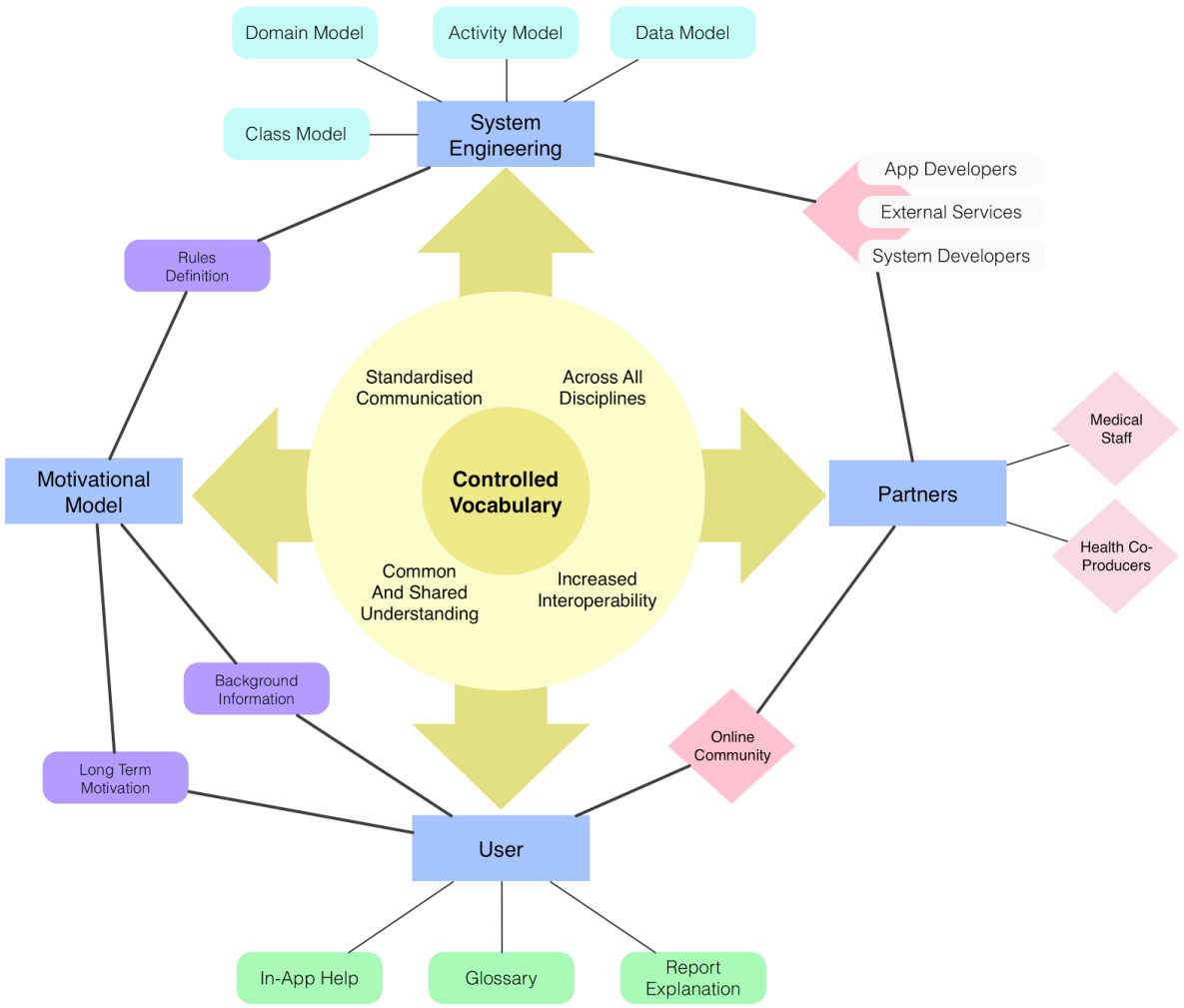

Figure 5 - Global picture of Controlled Vocabulary in context of PRECIOUS

The controlled vocabulary has connections to many actors of the PRECIOUS project. Figure 5 gives an overview of the controlled vocabulary in the PRECIOUS system. The basic interaction blocks of the controlled vocabulary are located in the middle of the Figure 5. Standardized communication: Talking about the same things, especially when it comes to interaction between different experts of various domains is a key issue in projects of this size of diversified domains.

Across all disciplines: The use of state of the art technologies, not only provides smooth exchange of semantic information between partners but also a standardized exchange of data from and to stakeholders and external organizations.

Common and shared understanding: The biggest outcome of the use of the controlled vocabulary is the achievement of a common global understanding of all used terms, concepts and technical vocabulary.

Increased operability: An increasing number of organizations publish their datasets as Linked Data and recognize benefits of increased interoperability between data sources and discovery of additional data [10]. The publication of the PRECIOUS controlled vocabulary allows similar experiences.

Furthermore four main actors who directly profit are highlighted: System engineers need a global picture of all needed variables and concepts that are used for the development of the database. In addition the relation between concepts need to be known by the data model designer in order to develop a data model adjusted to the requirements. Thus system engineers can build the VIM (see Deliverable 3.1, 3.2) based on information knowledge of the controlled vocabulary. It also enables internal and external app developers to build their applications based on that knowledge. Under the item Partners both internal project partners and external associated persons like medical staff and health co-producers are summarized. Users of the controlled vocabulary can on the one hand be active contributors to the vocabulary via online communities and on the other hand gain insights in the domain specific knowledge. The medical staff tends to use a very specific kind of language that is hardly understandable for patients: In-App help, glossary and report explanations interlinked with the controlled vocabulary can help the user gain insights in domain specific terms.

The development of the Motivational Model (see Deliverable 3.4) can be supported by the understanding of the reasons why a user need to change his behaviour, and allows a long-term motivation for the whole program. Rules Definition is situated between Motivational Model and System Engineering where such motivational rules must be computed and developed.

To summarize the whole picture, the PRECIOUS vocabulary provides a common understanding of the data to all partners (users, developers, experts, health staff), establish relationships to existing projects and data sources focused on e-health, allows the monitoring and maintaining of quality issues of the vocabulary, harmonize data from different sensors and input providers, and the usage of a standardized data model for the entire project.

3.2. Development of the controlled vocabulary

The development of the controlled vocabulary required a preciously overview of all possible data, domain terms and output over all domains in advance. There exist different controlled vocabulary life cycle models e.g. waterfall model, evolving model, evolutionary prototyping model, rapid prototyping, spiral model and many more [22]. Due to the the facts, that (1)

PRECIOUS consists of a large group of developers and partners having different roles and profiles, (2) PRECIOUS involve several different domains that were not fully transparent to all partners, (3) at the beginning it was not clear if the requirements are completely covered or may change during the development of the vocabulary, we decided to develop the vocabulary based on the Iterative-Incremental Ontology Network Life Cycle Model. This development model is organized in a set of iterations handled by short mini-projects with a fixed duration. Figure 5 shows the basic activity points of the Iterative-Incremental Ontology Network Life Cycle Model.

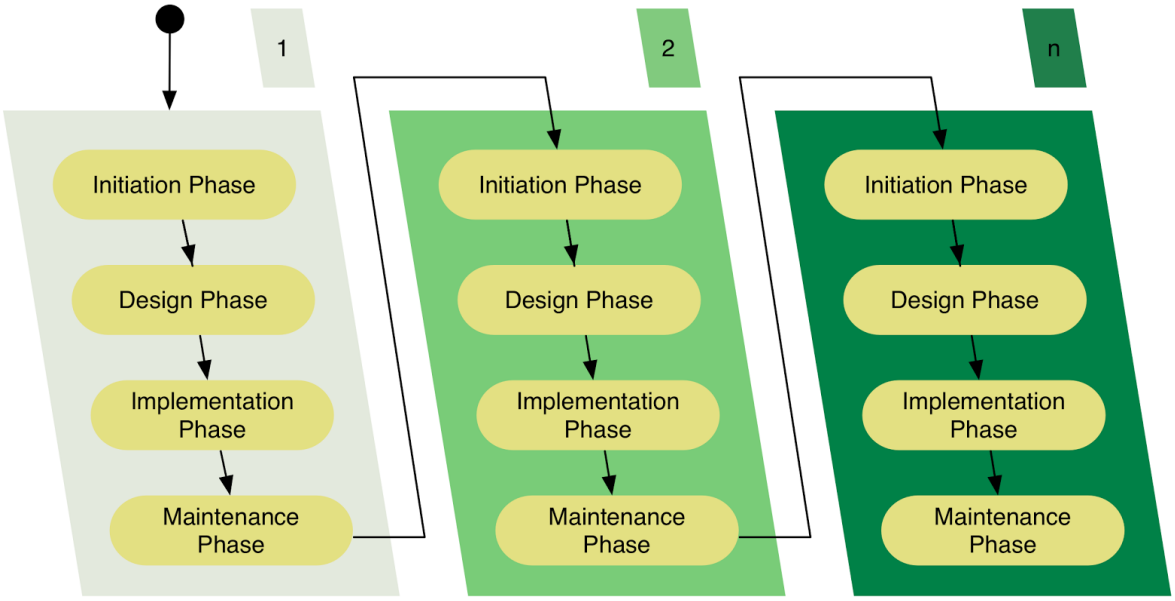

Figure 6 - Iterative-Incremental Ontology Network Life Cycle Model (Suarez-Figueroa 2012)

The controlled vocabulary life cycle described above defines the specific sequence of activities that the contributors of the controlled vocabulary need to carry out. Overall a global development model is needed, to coordinate the creation process of the vocabulary. Therefore we worked out the controlled vocabulary development cycle (CVDC). Figure 6 illustrates the creation and usage flow of the CVDC. Due to the number of partners and disciplines a collaborative approach for the model was indispensable.

The first step of the CVDC, illustrated at the top of the figure, shows the contribution of concepts of the experts of all domains. Each partner attend here at least one iteration of the Iterative-Incremental Ontology Network Life Cycle. The contribution time slots of all partners were set up in a row, with small breaks to conclude implementation and maintenance after each partner's contribution. In the next step of the cycle the controlled vocabulary is moderated. Various aspects of quality of the vocabulary (label issues, structural issues and linking issues) are analysed and fixed (see chapter 3.3). The validation by the domain experts is the final step of the CVDC before new concepts can be disseminated and conflated in the controlled vocabulary. Final proofs after moderation are an essential step to prove that the moderation work did not mix up any concept and maintain a good quality of the vocabulary.

The second flow, to the Usage in the middle of the CVDC is called Usage flow. The process to open closed data silos to the public, making data accessible for all, is a global process and can also be seen through the upcomming of multiple open government data portals and applications. Furthermore many data provider are going the step from the Web of Documents to a Web of Data, whereas all data are semantically distinguishable and identified resources holding various representations addressable via URIs. This allows interconnection and cross-linking with other related projects and vocabularies. As already described above, users are also main recipients of the controlled vocabulary. Users (e.g. Patient Data Requester, Web communities, App Developer) can not only use the vocabulary (Usage flow). Since the vocabulary is open and accessible to all users, they can also contribute to the vocabulary.

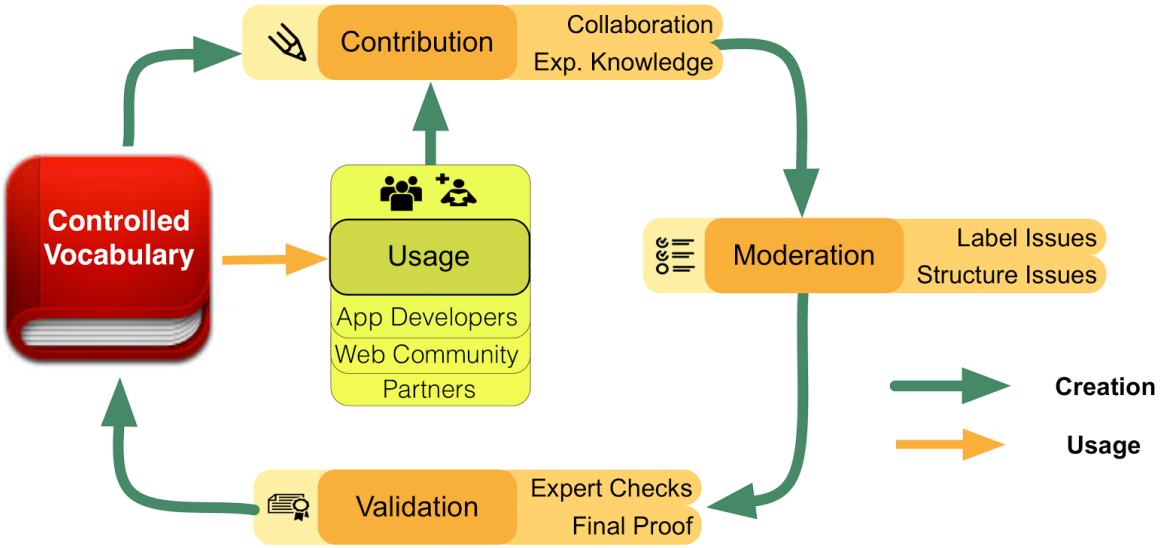

Figure 7 - Controlled Vocabulary Development Cycle (CVDC)

3.3. Quality of Controlled Vocabulary

Numerous studies has been conducted about data quality assessment in information systems. [8] and [12] defined a number of quality dimensions such as accuracy, completeness and consistency. [23] characterize the quality of data onto a high degree need to be measured in the context of the user. The impact of the data quality for the use of the data, for the user is therefore one big indicator for data quality. As discussed above, various projects nowadays migrate their data from proprietary organization formats to Linked Data which enables integration and interoperability with other online resources [15]. For the publication of Linked Data best practices and a list of common shortcomings and mistakes have been identified [10, 13]. But this findings did not provide clarification in the question of quality of controlled vocabularies. Although many standards (ANSI/NISO Z39.19-2005, ISO/DIS 25964-1), tutorials [21] and guidelines [7] exist and this work also going into question of data quality assurance, their definition like “inclusion of all needed facts” does not allow automatic testing of data quality for controlled vocabularies. PoolParty validator

and SKOSify are to tools that mainly focus on conformance with SKOS ontologies and only define few application-specific quality constraints. Our contribution was therefore to focus on narrowing this gap and provide a framework for semi-automatic quality assessment of controlled vocabularies [16].

In the first step we used a catalog of quality issues for controlled vocabularies that divide them into four main categories and 29 single quality issues (Mader 2015):

- Labeling and Documentation Issues
    - Omitted or Invalid Language Tags
    - Incomplete Language Coverage
    - No Common Language
    - Undocumented Concepts
    - Overlapping Labels
    - Missing Labels
    - Unprintable Characters in Labels
    - Empty Labels
    - Ambiguous Notation References
- Structural Issues
    - Orphan Concepts
    - Disconnected Concept Clusters
    - Cyclic Hierarchical Relations
    - Valueless Associative Relations
    - Solely Transitively Related Concepts
    - Unidirectionally Related Concepts
    - Omitted Top Concepts
    - Top Concepts Having Broader Concepts
    - Hierarchical Redundancy
    - Reflexive Relations
- Linked Data Specific Issues
    - Missing In-Links
    - Missing Out-Links
    - Broken Links
    - Undefined SKOS Resources
    - HTTP URI Scheme Violation
- SKOS Semi-Formal Consistency Issues
    - Relation Clashes
    - Mapping Clashes
    - Inconsistent Preferred Labels
    - Disjoint Labels Violation

- Mapping Relations Misuse

The experience with controlled vocabularies gained in other projects (e.g. MEKETRE, From Object To Icon...) allowed us the implementation of qSKOS, a tool for finding quality issues in SKOS vocabularies. qSKOS allows semi-automatic quality testing of the controlled vocabularies. It can be run locally, remotely or as part of the Poolparty Thesaurus Manager. We did several runs of qSKOS to continually review the quality of the vocabulary. A summary overview of current quality issues of the vocabulary can be found in list 1. It is possible to repeat the qSKOS check via the online tool hosted at Poolparty. A detailed description of all detected quality issues can be found at the Wiki page of qSKOS.

List 1 - Summary of Quality Issue Occurrences of the controlled vocabulary

- Empty Labels: OK (no potential problems found)
- Omitted or Invalid Language Tags: OK (no potential problems found)
- Incomplete Language Coverage: FAIL (292)
- Undocumented Concepts: FAIL (22)
- No Common Languages: OK (no potential problems found)
- Missing Labels: FAIL (2)
- Overlapping Labels: OK (no potential problems found)
- Orphan Concepts: FAIL (8)
- Disconnected Concept Clusters: FAIL (2)
- Cyclic Hierarchical Relations: OK (no potential problems found)
- Valueless Associative Relations: FAIL (8)
- Solely Transitively Related Concepts: OK (no potential problems found)
- Omitted Top Concepts: OK (no potential problems found)
- Top Concepts Having Broader Concepts: OK (no potential problems found)
- Hierarchical Redundancy: OK (no potential problems found)
- Mapping Relations Misuse: OK (no potential problems found)
- Reflexively Related Concepts: FAIL (1)
- Ambiguous Notation References: OK (no potential problems found)
- Unprintable Characters in Labels: OK (no potential problems found)
- Missing Out-Links: FAIL (177)
- Undefined SKOS Resources: OK (no potential problems found)
- Unidirectionally Related Concepts: FAIL (4)
- HTTP URI Scheme Violation: OK (no potential problems found)
- Relation Clashes: FAIL (3)
- Mapping Clashes: OK (no potential problems found)
- Inconsistent Preferred Labels: OK (no potential problems found)

- Disjoint Labels Violation: OK (no potential problems found)

	All concepts	omitted or invalid language tag	incomplete language coverage	no common language	undocumented concepts	overlapping labels	missing labels	empty labels	unprintable characters in labels	ambituous notation references
ODT	233	3	16	-	2	0	0	0	0	0
Geonames	688	0	43	-	60	162	9	0	1	0
NYTL	1920	0	0	-	1862	0	1	0	0	0
PXV	2112	1578	0	x	1492	7	2	0	0	0
Reegle	2952	3	1450	-	3	22	0	3	0	2
PRECIOUS	292	0	22	-	0	2	2	0	0	0

Table 1 - Comparison of PRECIOUS vocabulary with other small sized Web vocabularies in detail of Labeling and Documentation (Mader 2015)

We compared the PRECIOUS controlled vocabulary in all four main categories with other smaller public web vocabularies. Table 1 gives an example of the results in the qSKOS category Labeling and Documentation issues. To get a picture of all, we summed up the issues of all four categories over all vocabularies and compared them against one another. In comparison with the other web vocabularies, although most vocabularies are many times bigger, the PRECIOUS vocabulary with a total issue score of 815 scores comparatively well (see Table 2).

	Labeling and Documentation Issues	Structural Issues	Linked Data Specific Issues	Consistency Issues	SUM	Total Issues per concept
ODT	254	68	80	1	403	1,73
Geonames	963	689	715	1	2368	3,44
NYTL	3783	1921	3268	0	8972	4,67
PXV	5191	2310	2839	6	10346	4,90
Reegle	4435	12756	2287	322	19800	6,71
PRECIOUS	318	23	471	3	815	2,79

Table 2 - Summary of PRECIOUS vocabulary with other small sized Web vocabularies

3.4. Comparison of Collaborative Vocabulary Management Systems

In this chapter we will investigate which systems are suitable for the development for PRECIOUS controlled vocabulary.

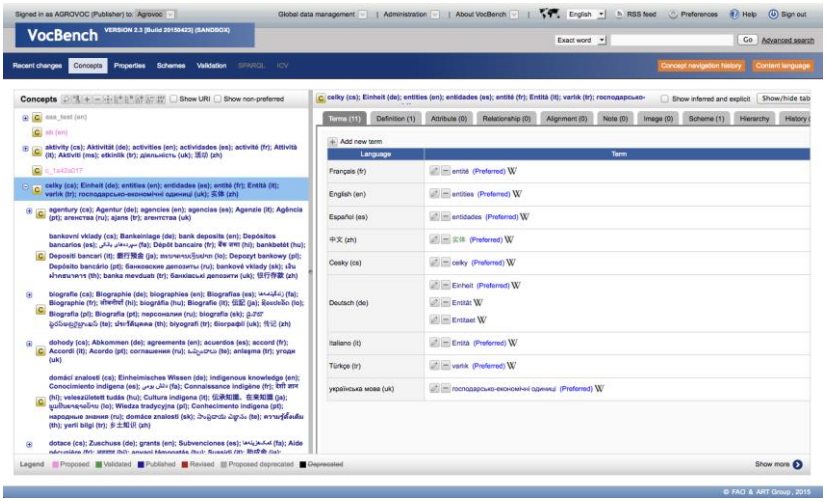

Figure 6 - Vocbench Vocabulary Management System

Our requirements for the system are:

- Collaborative contribution: Allow multiple partners to work on the vocabulary simultaneously. As indicated above (section 3.1), the work areas in the PRECIOUS project does have overlaps. In order to work simultaneously with several partners and allow moderation at the same time a collaborative system is mandatory.
- Workflow based: A system that allows the integration into CVDC (section 3.2), with optimal support for the creation of the vocabulary.
- User Roles: Offers various user roles for admins, editors, publishers...
- Usability: Easy to use interface, also for non-computer experts. Providing a high usability of the system for people from different disciplines of PRECIOUS.
- Open access: Public interface that allows other external partners to benefit from the vocabulary
- SPARQL Interface for further development of the vocabulary
- Export of various representation types (HTML, RDF, Turtle,...)

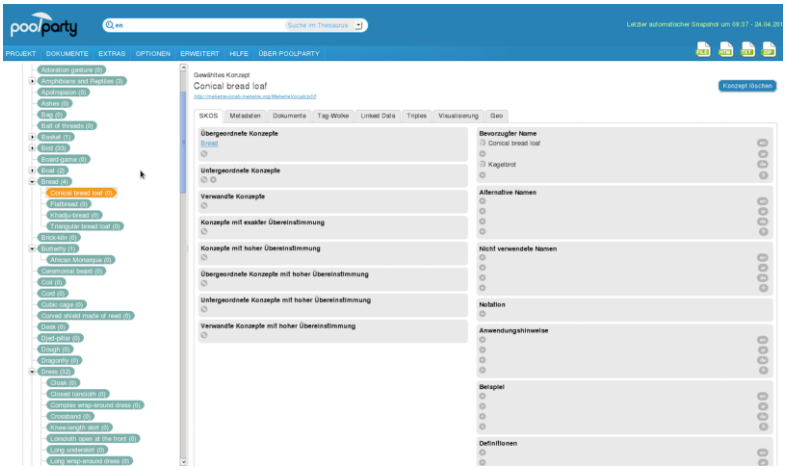

Figure 8 - Poolparty Thesaurus Manager

Vocabulary management systems that meet the defined requirements and came into our selection were Vocbench developed by FAO of the UN and University of Rome, Poolparty Thesaurus Management developed by Semantic Web Company GmbH and iqvoc developed by innoQ.

The following table will give an overview about all three systems

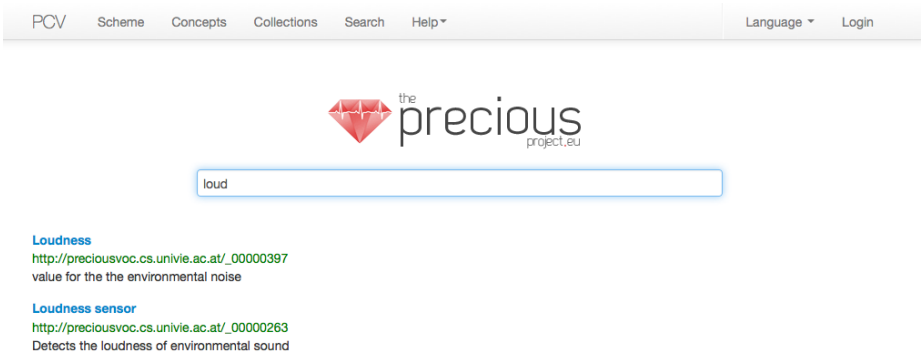

Figure 8 - iqvoc Vocabulary Management System

	Vocbench	**Poolparty Thesaurus Manager**	**iqvoc**
Producer	FAO of the UN and Università di Roma	Semantic Web Company GmbH	innoQ Deutschland GmbH
License	GPL license	Commercial software - academic license can be negotiated	Apache License, Version 2.0
Functionality	managing editorial workflow of creating/editing/publishing thesauri	comfortable taxonomy editor, smart workflow management, vocabulary mapping & linking, multiple projects, quality management	creating, editing, publishing thesauri, simple user interface, collaborative workflows
Website	http://aims.fao.org/tools/vocbench-2	https://www.poolparty.biz/taxonomy-thesaurus-management/	http://iqvoc.net/
Demo	http://202.73.13.50:55481/vocbench/ user: AGROVOC, password: AGROVOC	a free demo account can be requested	http://try.iqvoc.net/en.html
Resources	http://eprints.rclis.org/17735/ https://www.youtube.com/results?q=vocbench	http://bid.ub.edu/27nagy3.htm	http://ceur-ws.org/Vol-699/Paper2.pdf

Table 3 - Comparison of Vocabulary Management Tools

We examined closely the advantages and disadvantages of all systems and came to the conclusion that Poolparty and iqvoc both are vocabulary systems that offer a big variety of

functions and add-ons for professional vocabulary management. However the PRECIOUS environment need a lot of people without vocabulary development knowledge to work with the vocabulary system and needs strict, clear and clean forms and interfaces. The additional benefit of Poolparty and Vocbench would not justify individual multiple trainings with all partners, to teach them how to work with this more difficult systems. Therefore we finally concluded to use the proven iqvoc vocabulary management system.

3.5. Dissemination

The controlled vocabulary is publicly available hosted on a server of the University of Vienna. PRECIOUS controlled vocabulary offers a search page to quickly find concepts by the label, definition or other explanation fields. Furthermore concepts can be browsed in a hierarchically tree. The single view of each concepts, does provide all details about it and shows internal relations using SKOS-related terms like skos:broader, skos:narrower, skos:relate, and external relations like skos:closeMatch, skos:exactMatch, skos:broadMatch, skos:narrowMatch, skos:relatedMatch defined in the SKOS reference. Each resource of the vocabulary can be downloaded in three representations (HTML, RDF/XML and RDF/Turtle). The iqvoc vocabulary management system supports server-driven content negotiation for all available representations.

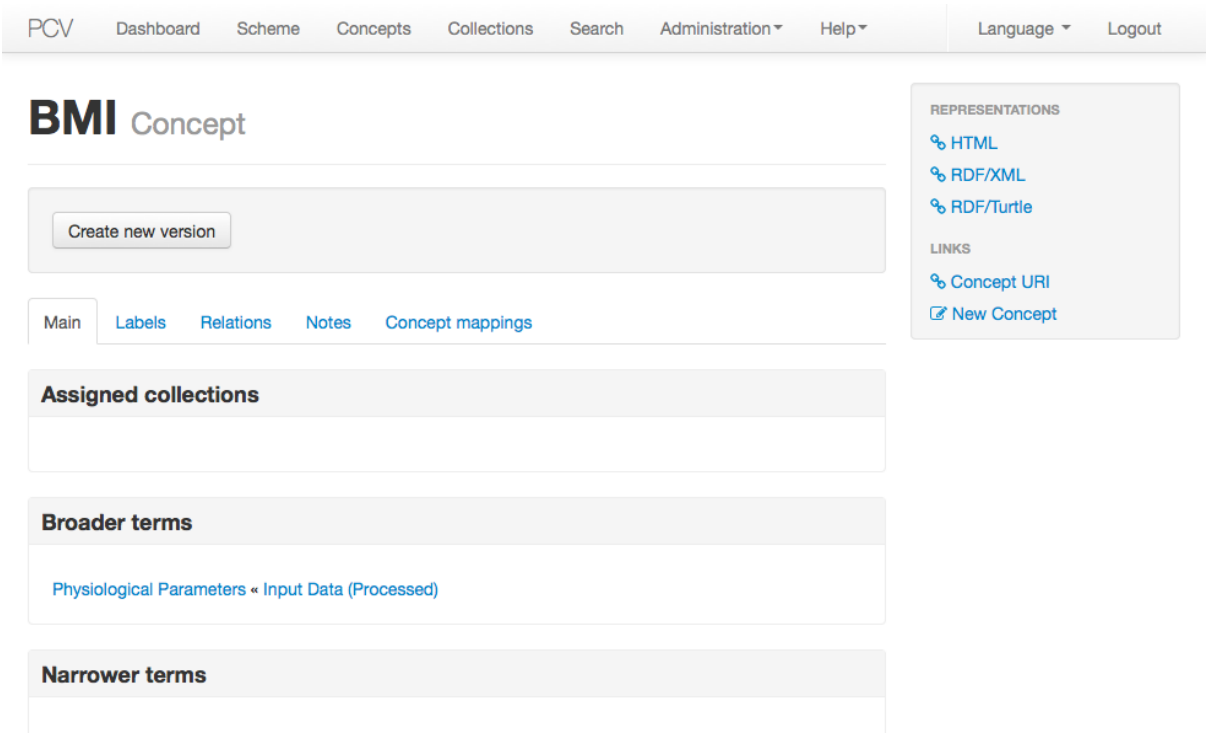

Figure 9 - HTML representation of resource BMI

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF                                xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"  xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:skos="http://www.w3.org/2004/02/skos/core#"        xmlns:dct="http://purl.org/dc/terms/"
xmlns:foaf="http://xmlns.com/foaf/spec/"                 xmlns:void="http://rdfs.org/ns/void#"
xmlns:iqvoc="http://try.iqvoc.net/schema#"         xmlns="http://preciousvoc.cs.univie.ac.at/"
xmlns:coll="http://preciousvoc.cs.univie.ac.at/collections/"
xmlns:schema="http://preciousvoc.cs.univie.ac.at/schema#">
<rdf:Description rdf:about="http://preciousvoc.cs.univie.ac.at/_00000029">
 <skos:prefLabel xml:lang="en">Weight</skos:prefLabel>
</rdf:Description>
<rdf:Description rdf:about="http://preciousvoc.cs.univie.ac.at/_00000002">
            <skos:prefLabel xml:lang="en">Physiological Parameters</skos:prefLabel>
</rdf:Description>
<rdf:Description rdf:about="http://preciousvoc.cs.univie.ac.at/_00000045">
 <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
 <skos:inScheme rdf:resource="http://preciousvoc.cs.univie.ac.at/scheme"/>
 <skos:prefLabel xml:lang="en">BMI</skos:prefLabel>
 <skos:related rdf:resource="http://preciousvoc.cs.univie.ac.at/_00000029"/>
 <skos:broader rdf:resource="http://preciousvoc.cs.univie.ac.at/_00000002"/>
 <skos:changeNote>
```

Code 1 - Excerpt of RDF/XML representation of resource BMI

3.6. Semantic Relations

The concepts of the controlled vocabulary can not only be accessed from the PRECIOUS infrastructure. As all concepts are uniquely identified by an URI it is possible to make semantic relations from and to different internal and external semantic web vocabulary schemes. Interlinked concepts help applications such as information retrieval packages to make use of severals knowledge organization systems.

Currently (ongoing process) there are 115 concepts of the PRECIOUS vocabulary manually interlinked with established taxonomies in medical and health domains like HL7, MESH and for more generally terms HL7 and UO. Most concepts from the psychology experts of PRECIOUS already existed in an non-direct-linkable publication [18]. To allow us a correct interlinking with this concepts we had to create a taxonomy by ourselves. We downloaded the taxonomy pdf file and created a hierarchical taxonomy with Protege. For all 16 main subjects we created disjoint top-level classes and included their in total 93 assigned subclasses below. For the dissemination we published the created taxonomy at Bioportal, a public taxonomy hoster. This allowed us to interlink all concepts of the psychology domain with the created online BCT (Behaviour Change Technique) taxonomy. The following list contains all schemes used for interlinking of PRECIOUS concepts.

- HL7: Health Level Seven International Founded in 1987, Health Level Seven International (HL7) is a not-for-profit, ANSI-accredited standards developing organization dedicated to providing a comprehensive framework and related standards for the exchange, integration, sharing, and retrieval of electronic health information that supports clinical practice and the management, delivery and evaluation of health services.
- LOINC: Logical Observation Identifiers Names and Codes LOINC is a common language (set of identifiers, names, and codes) for clinical and laboratory observations.
- MESH (Medical Subject Headings) MeSH (Medical Subject Headings) is the NLM controlled vocabulary thesaurus used for indexing articles for PubMed.
- UO: Units of Measurement Ontology The Ontology of Units of Measurement is developed as part of the OBO Foundry initiative.
- SNOMEDCT Systematized Nomenclature of Medicine - Clinical Terms: SNOMED CT (Systematized Nomenclature of Medicine--Clinical Terms) is a comprehensive clinical terminology, originally created by the College of American Pathologists (CAP) and, as of April 2007, owned, maintained, and distributed by the International Health Terminology Standards Development Organisation (IHTSDO), a not-for-profit association in Denmark. The CAP continues to support SNOMED CT operations under contract to the IHTSDO and provides SNOMED-related products and services as a licensee of the terminology.
- NDFRT National Drug File - Reference Terminology NDF-RT is an extension of the VHA National Drug File (NDF). It organizes the drug list into a formal representation. NDF-RT is used for modeling drug characteristics including ingredients, chemical structure, dose form, physiologic effect, mechanism of action, pharmacokinetics, and related diseases.
- BCTT Behaviour Change Technique Taxonomy A taxonomy of 93 behaviour change techniques within 16 conceptual groupings. The published journal supplementary file detailing the taxonomy has been submitted but it should be read in conjunction with the journal article

4. Semantic analysis of textual social media data

Beside physical parameters also psychological factors like mood have a significant influence on well-being and health of individuals. As written in [2,4] the quality of emotions contributes significantly to the eating habits. Nowadays many people use social media platforms for social interactions and express their moods and activities via textual communication and social interactions (see also Figure 10).

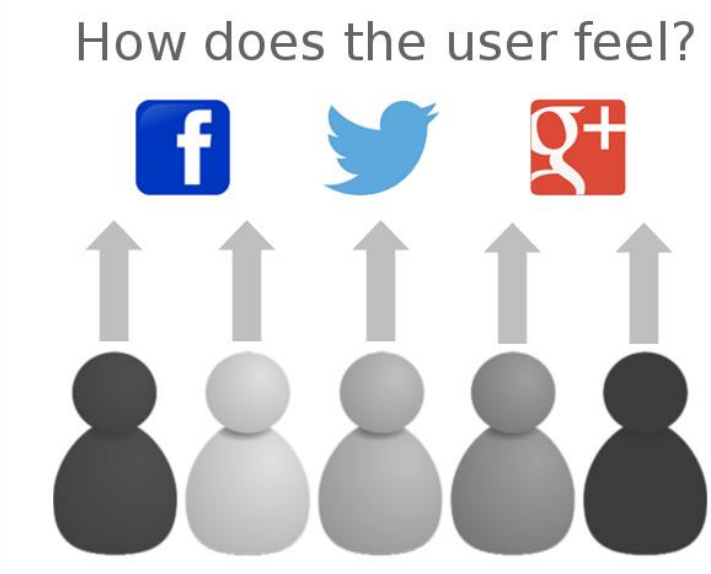

Figure 10: Users and information flow to social media networks

One way to get to know more about the mood of a person is to analyze the person's textual communication in social media. For this purpose we conducted an experiment, which results were published in [1] and presented at the Semantics[2] conference in Vienna. The first experiment consisted of three major steps (cf. Figure 11):

1) Data Acquisition
2) Feature Extraction
3) Model Building and Statistical Evaluation

The data acquisition is the collection of mood classified Facebook posts. The feature extraction is about extracting linguistic and text-based features. After these two steps ground truth data files were generated to build statistical mood classifier models.
Later a second experiment was done using these ground truth data files to get a broader view on the possible statistical results using different machine learning algorithms...
In the following it is described in detail how these three parts of this first experiment were conducted. The results are discussed subsequently. Afterwards the results of the second experiment are named and compared to the results of the first experiment...

[2] http://www.semantics.cc/

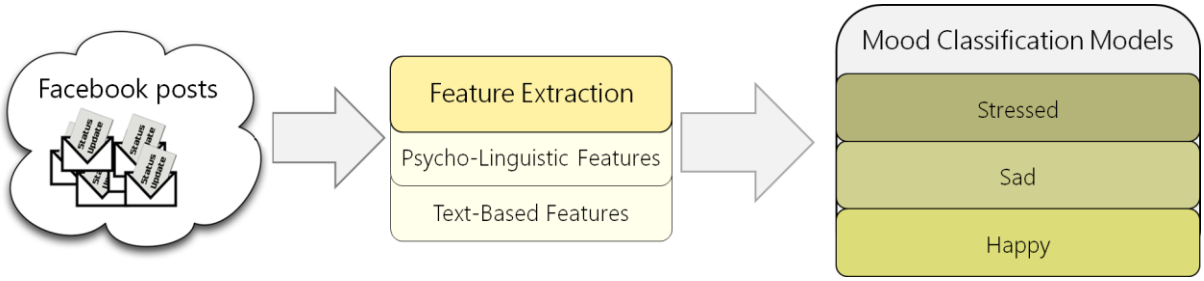

Figure 11: Experimental Setup for building a mood classifier for Facebook posts

4.1. Data Acquisition [1]

To create a ground-truth dataset from real world social media text communications, we performed a mechanical turk study[3]. We asked each turker to provide us five of their Facebook posts related to 'Sad', 'Happy', and 'Stress' moods. In order to ensure the quality of the work by coders, we requested them to provide the Facebook profile address of popular persons (singer, sportsperson, etc.) and provide us five posts of the selected person with 'Happy' tone. The goal was to detect possible inconsistencies and ensure that answers were specific and not given randomly. 100 turkers participated in this study and only those who had already received a total of Human Intelligence Tasks (HITs) higher than 5000 and HIT Approval Rate higher than 98% were accepted. In total, we collected 1500 posts related to three moods (500 happy, 500 sad, and 500 stress). It is important to note that to be allowed to participate turkers had to be active users of social media platforms, such as Facebook, and provide us their user ID.

4.2. First Semantic Layer - Feature Extraction [1]

For developing mood classifiers, we first set up two sets of features for classification:

1. Text-based Features (TB): We use a standard classification feature setup that is common in text and sentiment classification. Posts are represented as vectors of unigram and bigram features. Before feature extraction, posts are lowercased, URLs are removed, and numbers are normalized (canonical form). Next, feature reduction takes place. First, features that occur fewer than five times are removed. Second, features are subsequently reduced to the top 120 features in terms of log likelihood ratio.
2. Psycholinguistic Features (PLB): We utilized an established source of text analysis dictionary, Linguistic Inquiry and Word Count (LIWC[4]), to develop a set of features. LIWC was demonstrated by previous work [3] as a useful resource for identifying emotions of user-generated content. For LIWC, we used affective-indicative categories like positive/negative emotions, anxiety, sadness, and anger.

Subsequently, we chose three classifier algorithms to evaluate their performance for mood identification: Logistic Regression (LR), Support Vector Machine (SVM), and Bayesian network (BN) classifier. Also, for developing the mood classifier, we used two modeling

[3] http://aws.amazon.com/mturk/
[4] http://www.liwc.net/

approaches: (1) Balanced binary class for each mood, meaning three separated classifiers were developed for each mood. (2) Multi-class for all moods, meaning a classifier for predicting all moods with multi classes was developed. For analyzing the influence of the different sets of features on their performance, each classifier was set with all combinations of the feature sets and they were evaluated against each other. Finally, to evaluate the performance of the classifiers, we used four measures: precision (P), recall (R), F1-measure (the harmonic mean between precision and recall) and AUC (Area Under Curve) the Receiver Operator Curve (ROC).

Stress	Sad	Happy
hours	missing	halloween
late	sad	happy
nervous	miss	awesome
times	rip	favorite
anxiety	sick	fun
test	friend	excited
phone	cold	today
tomorrow	pretty	finally
won't	lost	birthday
minutes	damn	weekend
stressed	ugh	enjoying
break	sucks	dinner
work	put't	cool
start	sad	Love
hour	put	mom
missing	watching	birthday
makes	hot	great
coffee	snow	amp
omg	rain	party
suck	give	blast

Table 4: The most common top-20 terms extracted for each mood category extracted from Facebook posts collected via Mechanical Turk[5].

4.3. Second Semantic Layer - Building classifier models from Features

Classification results for different modeling approaches and various combinations of features are shown in Table 5 - Table 7. The results demonstrate the effectiveness of using text-related features for inferring individual moods. Nevertheless, for all moods, training a classification model using both sets of features shows improved performance compared to the same
models trained using one set of features. We observe that the Bayesian network classifier performs best for almost all moods with different combinations of features.

4.4. Results of First Experiments [1]

The best performances are observed for the mood 'Stress', while the worst are for 'Sad'. More precisely, in the case of the 'Stress' mood, we are able to achieve an F1 score of 0.88, coupled with high precision and recall, when using the Bayesian network classifier in combination with all the features. Similarly, for the same setting, we achieve an F1 score of 0.86, coupled with also high precision and recall for 'Happy' mood. However, we find a lower

[5] http://aws.amazon.com/mturk/

level of F1 score (0.79) when using the same classifier for 'Sad' mood, but it is still the best performance setting for this mood. With regard to binary or multiclass modeling, we observe that the binary class classifier using both sets of features outperforms other models and, in particular, outperforms models with binary classes using only one set of features. As the text-related features play an important role for mood classification, we computed a ranked list of terms from a set of 1500 posts for each mood (500 posts for each mood) as an illustrative example. For ranking the terms, we used the Mutual Information (MI) measure from the information theory which can be interpreted as a measure of how much the joint distribution of features Xi (terms in our case) deviate from a hypothetical distribution in which features and categories are independent of each other. Table 4 shows the top 20 terms extracted for each category. Obviously, many of the 'Happy' posts contain terms expressing sympathy or commendation. 'Sad' posts, on the other hand, often contain negative adjectives.

Features	Classifier	Binary-Class				Multi-Class			
		P	**R**	**F1**	**ROC**	**P**	**R**	**F1**	**ROC**
TB	LR	0.81	0.82	0.80	0.84	0.50	0.71	0.59	0.84
	SVM	0.84	0.83	0.81	0.68	0.48	0.82	0.61	0.78
	BN	0.86	0.86	0.85	0.91	0.60	0.98	0.70	0.91
PLB	LR	0.81	0.81	0.81	0.85	0.58	0.72	0.64	0.86
	SVM	0.81	0.82	0.79	0.67	0.58	0.70	0.63	0.82
	BN	0.80	0.79	0.79	0.82	0.53	0.75	0.63	0.84
Both	LR	0.85	0.85	**0.85**	0.91	0.65	0.73	0.71	0.92
	SVM	0.85	0.85	0.84	0.75	0.62	0.77	0.69	0.86
	BN	**0.88**	**0.86**	**0.86**	**0.94**	0.70	0.91	0.80	0.96

Table 5: Results of binary-class and multi-class classifiers using different combination of features for ‘happy’ mood. The features used are TB (text-based), PLB (psycho-linguistic-based), and a combination of both. These features combinations were used to build mood classification models using three machine learning algorithms: Logistic Regression (LR), Support Vector Machine (SVM), Bayesian Network (BN).

Features	Classifier	Binary-Class				Multi-Class			
		P	**R**	**F1**	**ROC**	**P**	**R**	**F1**	**ROC**
TB	LR	0.79	0.80	0.79	0.76	0.53	0.49	0.50	0.76
	SVM	0.79	0.80	0.79	0.76	0.63	0.51	0.56	0.73
	BN	0.87	0.87	0.88	0.87	0.87	0.73	0.80	0.96
PLB	LR	0.67	0.73	0.68	0.67	0.41	0.42	0.43	0.68
	SVM	0.62	0.74	0.64	0.50	0.42	0.48	0.45	0.66
	BN	0.74	0.71	0.71	0.77	0.42	0.44	0.42	0.69
Both	LR	0.77	0.78	0.77	0.78	0.56	0.54	0.55	0.81
	SVM	0.80	0.81	0.79	0.67	0.58	0.56	0.57	0.75
	BN	**0.89**	**0.89**	**0.88**	**0.92**	0.87	0.76	0.81	0.96

Table 6: Results of classifier for ‘stress’ mood (binary-class and multi-class,

binary-class and multi-class classifiers using different combination of features for 'stress' mood. The features used are TB (text-based), PLB (psycho-linguistic-based), and a combination of both. These feature combinations were used to build mood classification models using three machine learning algorithms: Logistic Regression (LR), Support Vector Machine (SVM), Bayesian Network (BN).

Features	Classifier	Binary-Class				Multi-Class			
		P	**R**	**F1**	**ROC**	**P**	**R**	**F1**	**ROC**
TB	LR	0.77	0.78	0.76	0.75	0.50	0.40	0.44	0.75
	SVM	0.78	0.79	0.75	0.61	0.53	0.43	0.48	0.72
	BN	0.83	0.83	0.83	0.87	0.86	0.61	0.71	0.93
PLB	LR	0.71	0.75	0.70	0.68	0.45	0.34	0.39	0.68
	SVM	0.65	0.74	0.64	0.50	0.43	0.30	0.35	0.63
	BN	0.73	0.76	0.71	0.66	0.47	0.27	0.34	0.68
Both	LR	0.76	0.77	0.76	0.79	0.56	0.56	0.56	0.83
	SVM	0.78	0.79	0.77	0.65	0.56	0.57	0.57	0.77
	BN	**0.80**	**0.74**	**0.79**	**0.91**	0.75	0.76	0.75	0.95

Table 7: Results of 'sad' mood classifier (binary class - and multi-class classifier, text-based and/or psycho-linguistic-based features, machine learning algorithms: Logistic Regression (LR), Support Vector Machine (SVM), Bayesian Network (BN), cf. Table 5 & 6)

4.5. Further Experimental Results

In this scope a Java program was developed to evaluate additional Machine Learning algorithms on the previously created features sets using text-based and psycho-linguistic-based features and binary classifiers with the Weka API[6]. The results show that none of the additionally tested binary-class classifiers could exceed the results of the Bayesian Network binary-class classifier. The best results were again achieved by the classifier for the happy mood.

In case of binary class classifiers the best received F1 score was 0.88 and was achieved with C4.5 decision tree. Also in case of 'sad' and 'stress' classifier the C4.5 decision tree achieved the best F1 scores: stress F1=0.82, sad F1=0.79 (cf.Table 8).

In case of multi-class classifiers overall the results of these tests exceeded the one from the binary classifiers. The highest F1 score (0.84) was achieved for the happy mood with the Naive Bayes classifier (cf. Table 9).

[6] http://www.cs.waikato.ac.nz/ml/weka/documentation.html

Mood	Classifier	Binary-class			
		P	R	F1	ROC
happy	Naive Bayes	0.87	0.84	0.84	0.91
	C4.5 Decision Tree	0.87	0.88	0.88	0.89
	Linear Logistic Regression	0.85	0.85	0.85	0.90
	Stochastic Gradient Descent	0.84	0.84	0.84	0.77
	Voted Perceptron	0.80	0.81	0.79	0.69
sad	Naive Bayes	0.80	0.74	0.75	0.81
	C4.5 Decision Tree	0.79	0.79	0.79	0.80
	Linear Logistic Regression	0.78	0.79	0.77	0.81
	Stochastic Gradient Descent	0.77	0.79	0.78	0.67
	Voted Perceptron	0.72	0.76	0.69	0.57
stress	Naive Bayes	0.79	0.70	0.72	0.82
	C4.5 decision tree	0.82	0.82	0.82	0.82
	Linear Logistic Regression	0.79	0.80	0.76	0.70
	Stochastic Gradient Descent	0.80	0.81	0.80	0.69
	Voted Perceptron	0.66	0.74	0.66	0.54

Table 8: Results (Precision (P), Recall(R), F-Measure (F1) and Receiver Operator Curve(ROC)) from the evaluation of binary classifiers (Naive Bayes, C4.5 Decision Tree, Linear Logistic Regression, Stochastic Gradient Descent, Voted Perceptron) for 'Happy''. 'Sad'. and 'Stress'' mood

Moods	Classifier	Multi-class			
		P	R	F1	ROC
happy	Naive Bayes	0.737	0.850	0.790	0.933
	C4.5 Decision Tree	0.752	0.762	0.757	0.874
	Linear Logistic Regression	0.759	0.822	0.789	0.932
sad	Naive Bayes	0.760	0.641	0.696	0.883
	C4.5 Decision Tree	0.696	0.756	0.725	0.854
	Linear Logistic Regression	0.679	0.653	0.666	0.858
stress	Naive Bayes	0.729	0.732	0.730	0.878
	C4.5 decision tree	0.756	0.679	0.715	0.845
	Linear Logistic Regression	0.698	0.667	0.682	0.851
average	Naive Bayes	0.742	0.741	0.739	0.898
	C4.5 decision tree	0.734	0.733	0.732	0.857
	Linear Logistic Regression	0.712	0.714	0.712	0.880

Table 9: Results from the evaluation of multi-class classification with different machine learning algorithms - classes are 'Happy'. 'Sad'. and 'Stress'. The first three lines contain Precision (P), Recall(R), F-Measure (F1) and Receiver Operator Curve(ROC) for each single mood (happy, sad, stress), the last line contains the average P,R,F1 and ROC of all three moods

5. Food sensor data analysis

Food intake monitoring of daily food intake and dietary behaviour is as an important process in understanding the development related medical conditions. These conditions include obesity which occurs when energy intake from food exceeds energy expended, and is a major determinant on human health [26] with strong links to non-communicable diseases, such as, T2D and CVD.

First semantic layer: food recognition and raw nutritional data

The food intake monitoring could be considered to be in two parts or phases, that is: the food intake detection (detecting the action that the user is eating) and food recognition (recognising the type of food, nutritional content, amount of the food, etc.). The purpose of food intake monitoring is to develop an understanding of a user's nutritional behaviour (food choices, eating patterns, etc.) and assess their risk to diet-related diseases. This knowledge will in turn inform appropriate personalised dietary advice that influences nutritional behaviour with the target of maintaining optimal health and disease prevention. The effectiveness of personalised dietary advice interventions is being further enhanced through advances in nutritional genomics (or nutrigenomics).[27] Nutritional genomics is creating new knowledge for further personalisation of dietary advice by taking into individual differences metabolic factors, particularly, genetic background and biomarker (phenotypic) status, and their individualised influence on diet and obesity. The current primary scope of the PRECIOUS food intake monitoring and feedback framework considers only nutritional behaviour, but provides flexibility for expansion of the scope to include metabolic factors and related aspects of nutritional genomics.[28]

The PRECIOUS leverage digital health tools for food intake monitoring and personalised dietary advice (feedback) based diet/obesity risks related to T2D and CVD. To that end, the project considers use of smart user devices (smartphone, smart watches, etc.) to recognize the food type. Specifically, this can be through use of PRECIOUS app interface for manual entry of food name or use the application's barcode scanning feature (with device camera) to scan the barcode on packaging of the food product to obtain the product name. Alternatively, the food recognize could be implemented by user taking a photo which is then sent to cloud servers for digital image treatment to recognize the food using a combination of digital image processing and machine learning algorithms. A more detailed description of food intake monitoring methods and related challenges is provided in deliverables D3.4 and D4.1.

When the food is identified by the PRECIOUS app it is then matched to an entry in a food database. From the database the related nutritional information is retrieved and returned to the user as feedback. The overall process of recognition of food type and return of the raw nutritional information constitutes the first semantic layer for food intake monitoring. The feedback of this nutritional information to the user is based on the EU Regulation No 1169/2011 [29] on provision of food information to consumers entered into application on 13 December 2014 . Nutritional labelling is associated with the labelling on packaged food products. However, the requirements on how food nutritional information is displayed on digital devices like smartphones should in principle adhere to the same display requirements standard food labelling as specified by The European Food Information Council (EUFIC) [30]. The EUFIC guidelines specify the way in which labels should be set out and what

mandatory information should be included. For the back of nutritional table, the mandatory nutrients are (and they must be listed in this order); energy (kJ and kcal), fat, saturates, carbohydrate, sugars, protein, salt (sodium can no longer be listed). Companies can voluntarily also provide mono-unsaturates, polyunsaturates, polyols, starch and fibre. It is mandatory that the amount per 100g in the food is detailed; however, it is optional to provide the amount per portion. If declared it is critical to clearly define what constitutes as portion and reference the intake values. Companies can only detail vitamin and mineral contents of products if they are present in significant amounts (defined in point 2 of Part A of Annex XIII in EU FIC). If a claim is made about vitamins or minerals it is mandatory that their content is declared in the nutritional table. It is not a mandatory requirement for companies to provide front of pack nutritional labelling; however, if they wish to do so, the content of that declaration is limited to;

a) The energy value; or
b) The energy value together with the amounts of fat, of which is saturates, sugars and salt
Additional forms of expression, including colour coding, can also be used.

Second semantic layer: Representation of nutritional information to the user
The EUFIC guidelines on recommended daily intake of different nutrients in based on reference intakes. The reference intakes are not intended as targets, due to differences in energy and nutrient requirements for different people. But instead they provide a useful indication of how much energy the average person needs and how a particular nutrient fits into their daily diet. The daily reference intakes for adults are:

- Energy: 8,400 kJ/2,000kcal
- Total fat: 70g
- Saturates: 20g
- Carbohydrate: 260g
- Total sugars: 90g
- Protein: 50g
- Salt: 6g

Similarly the EUFIC daily reference intakes for voluntary listed nutrients (vitamins and minerals) are shown in the Appendix section.

The feedback of nutritional information back to the user also presents challenges in ensuring that the information could be easily understood and meaningfully interpreted by non-expert users. This is because level of detail in the nutritional information (based on the EUFIC guidelines) may be overwhelming to some non-expert user and lacks additional advice in the context of the recommended intake for the user. This requires additional modification of the nutritional information. To that end, this modification or simplification of the EUFIC labelling for enhanced understanding has been left to individual member states.
The European commission is due to complete a review of the different approaches used by member states and this is due to be submitted to the European parliament and the Council by 13th December 2017. For instance, in the UK, the Department of Health has provided guidance to food manufactures, on the use of traffic light colour coding to help consumers

understand where a food product sits in terms of fat, saturates, sugar and salt (Green=LowHigh, Amber=Medium, Red=HighLow) [31] (see Figure 12 below).

Table 2: Criteria for 100g of food (whether or not it is sold by volume)

Text	LOW	MEDIUM	HIGH	
Colour code	Green	Amber	Red	
Fat	≤ 3.0g/100g	> 3.0g to ≤ 17.5g/100g	> 17.5g/100g	> 21g/portion
Saturates	≤ 1.5g/100g	> 1.5g to ≤ 5.0g/100g	> 5.0g/100g	> 6.0g/portion
(Total) Sugars	≤ 5.0g/100g	> 5.0g and ≤ 22.5g /100g	> 22.5g/100g	> 27g/portion
Salt	≤ 0.3g/100g	> 0.3g to ≤ 1.5g/100g	>1.5g/100g	>1.8g/portion

Note: portion size criteria apply to portions/serving sizes greater than 100g

Table 3: Criteria for drinks (per 100ml)

Text	LOW	MEDIUM	HIGH	
Colour code	Green	Amber	Red	
Fat	≤ 1.5g/100ml	> 1.5g to ≤ 8.75g/100ml	> 8.75g/100ml	>10.5g/portion
Saturates	≤ 0.75g/100ml	> 0.75g to ≤ 2.5g/100ml	> 2.5g/100ml	> 3g/portion
(Total) Sugars	≤ 2.5g/100ml	> 2.5g to ≤ 11.25g/100ml	> 11.25g/100ml	> 13.5g/portion
Salt	≤ 0.3g/100ml	>0.3g to ≤0.75g/100ml	> 0.75g/100ml	> 0.9g/portion

Note: Portion size criteria apply to portions/serving sizes greater than 150ml

Figure 12: UK Department of Health colour coding criteria for different macro nutrients in 100g of food (Note: portion size criteria apply to portions/serving sizes greater than 100g)

As part of the Department of Health's Guidance for front of pack labelling energy information (kJ and kcal) is excluded from the colour coding. Instead it should be displayed on a neutral/colourless background without association to the colour coding. The amount each nutrient within the food contributes towards the individual's reference intake can also be expressed as a percentage.

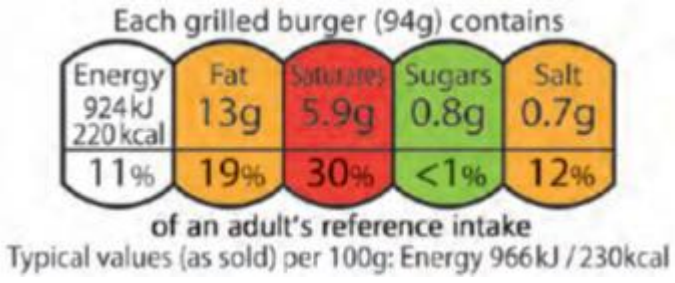

Figure 13 Example nutritional information display for a packet of 4 beef burgers sold raw:

The UK colour coding proposal has been well received by some member states and is an approach that is also considered a suitable approach for the PRECIOUS nutritional feedback in the second semantic layer until a community-wide simplification format comes into force. The level (colour code) of a particular nutrient is evaluated by the VIM engine based on the food nutritional information and amount food amount or portion size entered by user, measured from digital image or from barcode (see Figure 14).

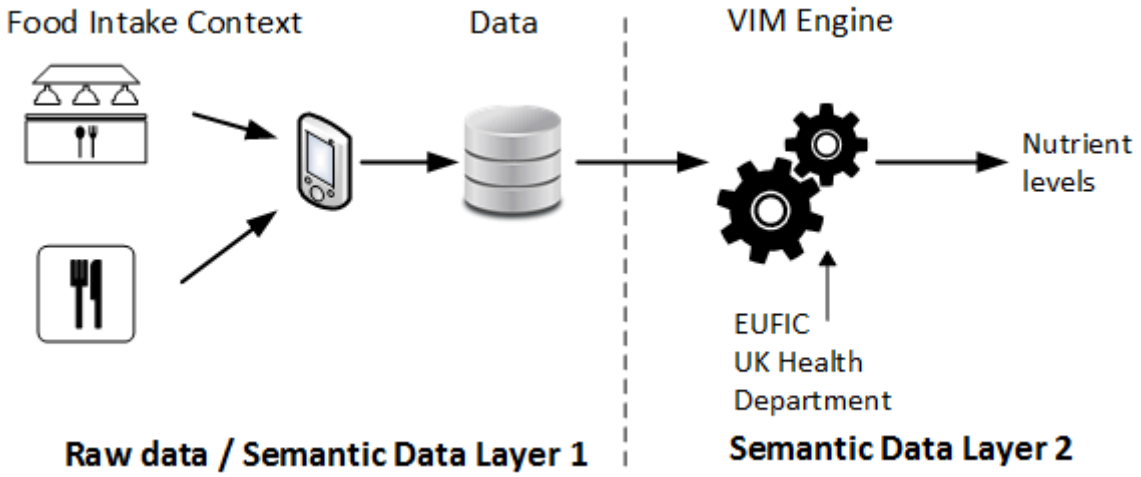

Figure 14: Overview of VIM and rules for food intake

The adaptation of the of the UK three-level colour coding scheme semantics of Figure 14 to the higher level context is shown in Table 10.

UK colour code	Semantic data layer 2	Human meaning
Green	Low	The content of particular nutrient is low for a particular food amount or portion
Amber	Medium	The content of particular nutrient is medium for a particular food amount or portion
Red	High	The content of particular nutrient is high for a particular food amount or portion

Table 10: Adaptation of UK department three-level colour coding scheme for description of levels of different nutrients

The discussion on semantic aspects has focused on semantic representation from the perspective of dietary tracking tools (food intake monitoring and recording). However, the PRECIOUS project considers diet to be a more broader all-encompassing concept that also tasks into account inter-related behaviours. This means that apart from dietary tracking tools, the dietary arm of the PRECIOUS project also includes dietary challenges and information modules (described in detail in deliverable D3.4). Both the dietary guidelines and information modules are underpinned by EUFIC guidelines described above.

6. Heart rate sensor data analysis

Stress, recovery and physical activity analysis for Precious VIM is based on sophisticated utilization of collected sensor data about the users' beat-to-beat heart rate (heart rate variability, HRV) and optionally acceleration data. Figure 15 describes the analysis process and the semantic data layers related to HRV and acceleration data analysis for the PRECIOUS VIM. The raw sensor data is complemented with the users' background parameter information such as age, height, weight and general physical activity level (activity class). The data is delivered to Firstbeat analytics engine via API.

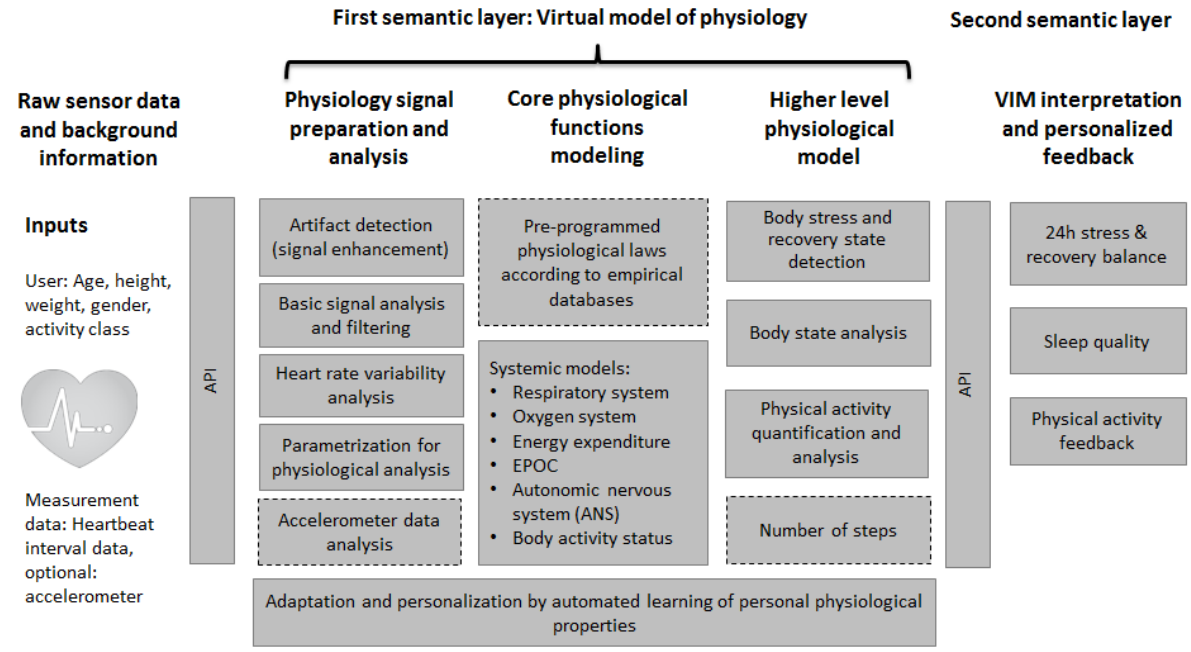

Figure 15 Different layers for heart rate sensor data analysis

In the first semantic layer the raw data along with the background parameters is analyzed by going through several phases. The first phase contains possible artefact detection and correction of the raw R-R interval and acceleration data, parametrization of the data, analysis of acceleration data, and basic HRV analyses. Thereafter the user's core physiological functions are modelled wherein pre-programmed physiological laws are utilized together with systemic models of body's physiology that represent basic physiological phenomena. These are for example respiration activity analysis, oxygen consumption and energy expenditure analysis, excess post-exercise oxygen consumption (EPOC) analysis, autonomic nervous system status analysis and the body's activity status analysis. Each of these phases contain several subphases wherein the specific phenomena are modelled by mathematical algorithms (for further information please see references/Firstbeat white papers [17,37-40]). The physiological system level parameters are used for higher level physiological model, which forms the interpretation of the virtual model of the user's physiology and contains information regarding periods of stress, recovery, and physical activity in terms of duration and intensity. Variables such as number of steps taken can be produced.

In the second semantic data layer the aforementioned physiological parameters are interpreted for making conclusions about the user's status compared to general, international, and database-based guidelines. The results are interpreted to belong to categories such as worse than recommended (poor) or sufficient/better than recommended (good) or being at average level (moderate). The bigger the amount of data or the more longitudinal the data the stronger or more accurate the conclusions about the user status can be.

From controlled vocabulary perspective the most important for the actors in the PRECIOUS system (e.g. users, developers, health staff) is to understand the concept regarding collection of raw data and the second semantic layer, i.e. interpreted results.

The results of heart rate sensor data analysis can be visualized to the user in a form of graphs, bars, points and so on. One possibility is to show a lifeline about physiological reactions at different time points. Figure 16 describes one example of possible ways to visualize the data.

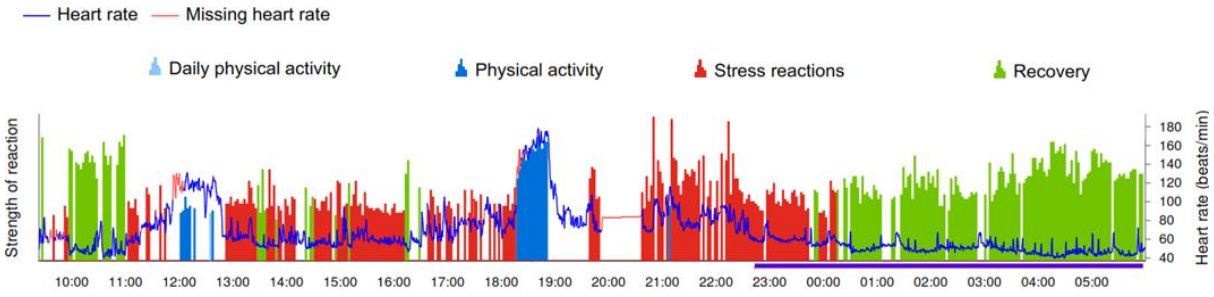

Figure 16. Example of heart rate sensor data visualization.

7. Mobile phone sensor data analysis

Contemporary mobile phones (smartphones) now come equipped with an increasing number of inbuilt sensors, such as, accelerometers, gyroscopes, barometer, proximity sensors, and so on (see Figure 17). The sensors have many uses in supporting the operation of the phone (e.g. using gyroscope to control the screen rotation) and leveraging the raw data from the different sensors to create rich mobile apps. To that end, the fact that 79% of users aged 18-44 have their smartphones for up 22 hours a day [32] makes smartphones a powerful tool for implementation of health and wellness apps. The use of wearable devices connected to smartphones (e.g. using Bluetooth) enables even more accurate sensor measurement data for health applications. In the PRECIOUS project the smartphone is considered as a baseline device for monitoring physical activity, sleep duration and food intake via the PRECIOUS app.

Figure 17: Example evolution in the variety/number of embedded sensors for different Samsung Galaxy smartphone models (source: Qualcomm)

Raw Mobile Sensor Data to Semantic Data Layer 1

In most cases user's Physical Activity (PA) can be derived from the raw measurement data of their smartphone's inbuilt accelerometer. The accelerometer sensors present in nearly all smartphones and wearable devices are a reliable and energy-efficient (battery-preserving) way to track PA. The accelerometer uses non-gravitational acceleration measurements to detect vibrations of a mobile device due to movement state changes (standstill, velocity etc.). The smartphone accelerometer typically measures acceleration (in m/s^2 or g-force) for three axes, namely: forward (x-axis or roll), side (y-axis or pitch) and vertical (z-axis or yaw). The accelerometer measurements can then be used to approximate the movement of a user holding or wearing while performing a PA (e.g. running) as shown in Figure 18. The detection of peaks and variations of accelerations are essential in all three axes in order to detect a unit cycle (stride or step) of walking or running. For example Figure 18 shows measurements

corresponding to the three axes for a running user with one axis showing relatively large periodic acceleration changes no matter how the device is held or worn.

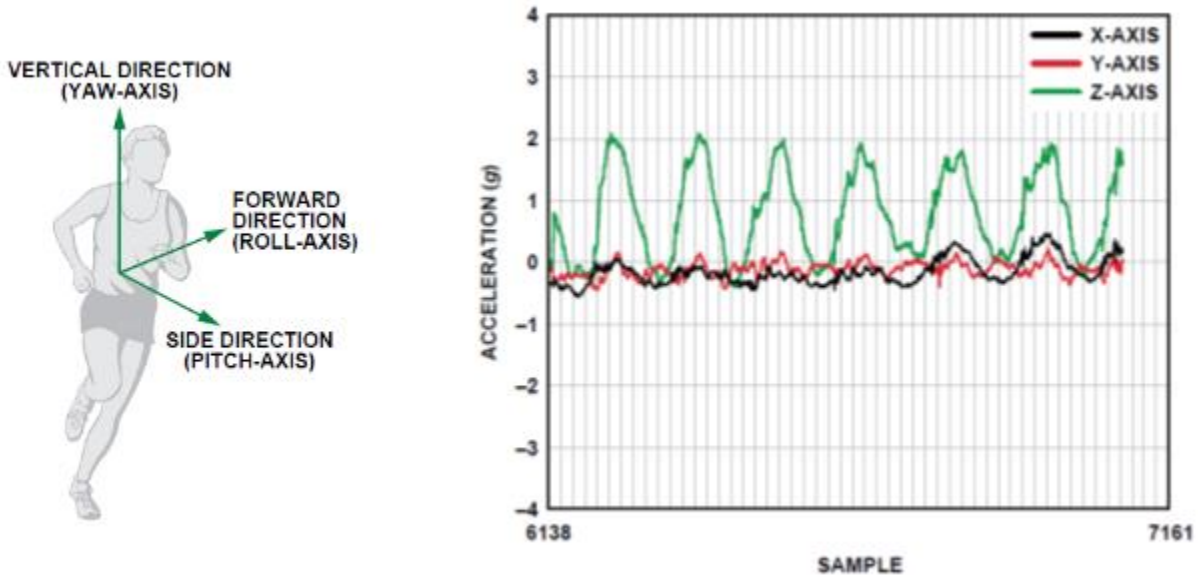

Figure 18: Accelerometer axes definition (left) and example acceleration G-force measurements (right) on 3-axes for a user performing a running activity (source: [33])

The results of analysis of the accelerometer can be in form of step counts or recognition of type of activity (e.g. user is riding a bicycle, running, standing still etc.) using machine learning classifier algorithms that are part of the Google Services API [33]. The Google API sends broadcast with the details of the recognized PA type for the preceding 20 seconds and the PRECIOUS app receives them and processes the data using simple filtering and state machines.

Semantic Data Layer 2

The PA data must be projected to higher semantic layers to the user in a more meaningful and easy to understand way. Therefore, in addition to step counts and activity detection the PA data the PRECIOUS VIM engine and rules for PA are also used to translated into additional parameters (Figure 19), such as, distance covered (for walking or running) and energy expended (in terms of calories burned). The distance covered may be evaluated from GPS sensor in smartphone or wearable. In case the GPS measurements are unavailable (e.g. GPS switched off), knowing the gender and height of the user and the number of steps, it is possible to estimate distance travelled by walking by multiplying the number of steps with the step or stride length given by (Crosby, 2015) [34]:

- For Men: $step\_length[m] = 0.415 \cdot height[m]$
- For Women: $step\_length[m] = 0.413 \cdot height[m]$

Furthermore, from the walking distance and duration, one can calculate the average speed, as well as, use the following expression to estimate the energy expended (burned calories) based on the user's body weight [35]:

$$E = (0.46 \cdot V^2 - 1.24 \cdot V + 1.69) \cdot W$$

where E is the estimated burned calories, V is the average walking speed in miles per hour and W is the weight in kg. For physical activities, such as, biking, the energy expended can be obtained from the same expressions assuming biking being equivalent to walking speed of 3 mph and multiplying the duration of the biking activity in minutes by 270 steps/min.

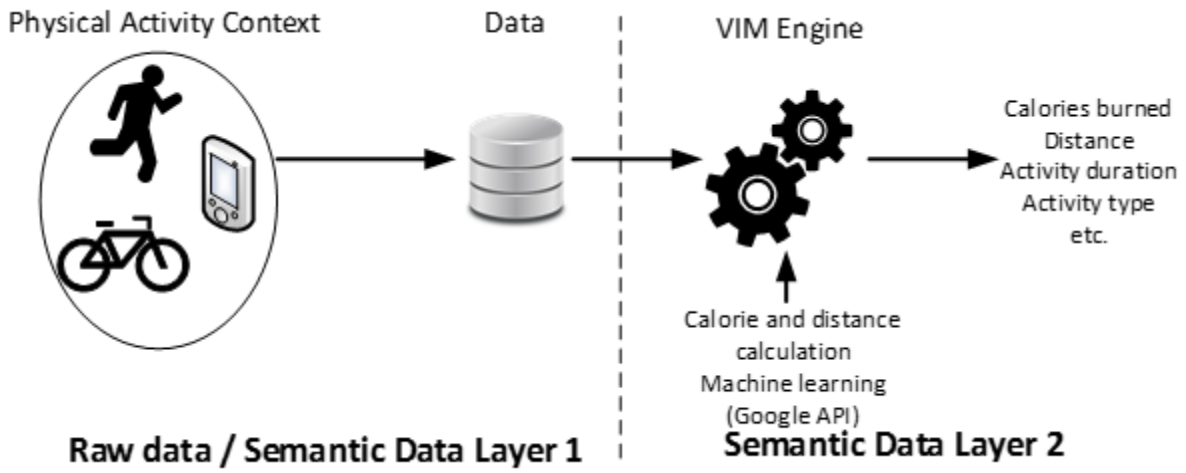

Figure 19: PRECIOUS Physical Activity system

8. Ambient sensor data analysis

8.1. Indoor Environmental Quality Variables Overview

Generally, to assess well-being of occupant in dwelling, the indoor environmental quality (IEQ) is important to monitor and is part of the risk factors identified in the PRECIOUS project. In details, IEQ is related to the user context awareness (UCA) at home and the data analysis of related variables which is the topic of the present chapter. The PRECIOUS project defined the following IEQ variables within deliverable 3.1 and 3.2: temperature, humidity, noise, light

In the deliverable D4.1, the infrastructure to collect data from ambient sensors has been described. A new protocol, xAAL, has been proposed to fight interoperability issues in the home automation domain and it is used to gather sensor data characterizing the PRECIOUS home user space. The xAAL infrastructure deployed in the home environment allows also the access to the PRECIOUS services. Indeed a gateway forwards home user data in the PRECIOUS database in order to provide inputs to the VIM. The VIM is the heart of the PRECIOUS system. It processes data, analyzes context and finally allows to deliver specific user feedback. The process offering feedback on environmental quality to PRECIOUS users is based on rules (see D3.2 section 4.6).
Each indoor environmental quality variable is stored in the backend of the PRECIOUS platform. These will be gathered automatically with xAAL according to the location of the sensors in various rooms (e.g. bedroom, living room) and include the following:

- Temperature (in °C), measured by the thermometer sensor
- Humidity (in %), measured by the hygrometer sensor
- Noise (in dB(A)), measured by the sound meter
- Light (in lux), measured by the light sensor

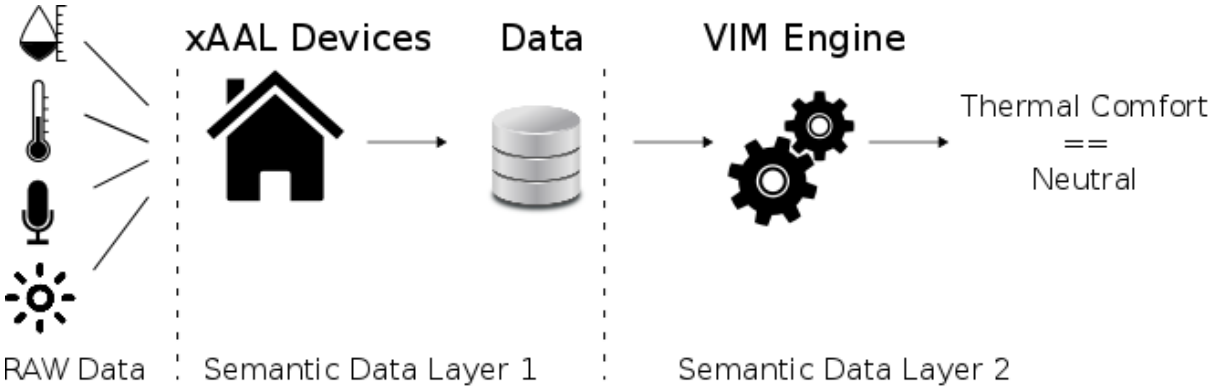

Figure 20: Overview of PRECIOUS system: IEQ variables - VIM - Semantic Data Layer

The VIM Engine described is part of the “Domain knowledge VIM processing module” in the high-level diagram of PRECIOUS system implementation depicted Figure 1. The semantic data layer 2 is used to produce high level context such as “Thermal Comfort” produced with IEQ variables temperature & humidity. The semantic meaning extract from the semantic data

layer 1 is mainly driven by European (or international) standards, norms or recommendations from World Health Organization (WHO) and/or European guidelines if standards/norms do not exist.
In the following, we will describe the IEQ variables gathered (raw data, Figure 20) and the related semantic data extraction (semantic layer 1 & 2) according to the domain-knowledge as described in D3.2.

8.2. Raw Data to Semantic Data Layer 1

The transparent sensing layer at home is realized by the xAAL infrastructure. The xAAL specification defines the notion of schema which is a description of a xAAL device. In the case of the IEQ variables identified in the PRECIOUS project, we have the following schemas:

- “thermometer.basic”
- “hygrometer.basic”
- “luxmeter.basic”
- “loudness.basic”

By design the xAAL infrastructure provides a first semantic meaning to data related to xAAL devices as for example the temperature attributes shown Table 11 extracted from the Json Schema (see Appendix 10.1)

```
"temperature": {
        "description": "Temperature",
        "unit": "°C",
        "direction": "out",
        "type": {
          "$schema": "http://json-schema.org/schema#",
          "type": "number"
        }
```

Table 11: Temperature attributes of the xAAL Json Schema representing ‘thermometer.basic’

This first description of the temperature data is associated to the semantic data layer 1 as shown Figure 1. It is a low level context information or primary context information from sensors.

8.3. Semantic Data Layer 1 to Semantic Data Layer 2

This section deals with techniques to produce a high level of abstraction using semantic data from the layer 1. All the methods described here will be implemented in the PRECIOUS cloud server as shown Figure 1. The implementation will be detailed in the deliverable D4.2 “System integration report”.

8.3.1. Thermal Comfort

For thermal comfort, the European standard EN15251 (European committee for Standardisation) will be the reference. The semantic meaning related to the the thermal comfort is extracted from the PMV-PPD (ISO) model. The Predicted Mean Vote (PMV) index is defined by :

$$PMV = [0.303.e^{(-0,036.M)} + 0,028].\{(M-W) - 3,05*10^{(-3)}.[5733 - 6,99.(M-W) - p_a] - 0,42.[(M-W) - 58,15] - 1,7.10^{(-5)}.M.(5867 - p_a) - 0,0014.M.(34 - t_a) - 3,96,10^{(-8)}.f_{cl}.[(t_{cl}+273)^4 - (t_r+273)^4] - f_{cl}.h_c.(t_{cl} - t_a)\}$$

$$t_{cl} = 35,7 - 0,028.(M-W) - I_{cl}.\{3,96.10^{-8}.f_{cl}[(t_{cl}+273)^4 - (t_r+273)^4] + f_{cl}.h_c.(t_{cl} - t_a)\}$$

$$h_c = \begin{cases} 2,38.|t_{cl} - t_a|^{0.25} & for \;\; 2,38.|t_{cl} - t_a|^{0.25} > 12,1\sqrt{v_{ar}} \\ 12,1\sqrt{v_{ar}} & for \;\; 2,38.|t_{cl} - t_a|^{0.25} < 12,1\sqrt{v_{ar}} \end{cases}$$

$$f_{cl} = \begin{cases} 1,00 + 1,290.I_{cl} & for \;\; I_{cl} \le 0,078\, m^2.K/W \\ 1,05 + 0,645.I_{cl} & for \;\; I_{cl} > 0,078\, m^2.K/W \end{cases}$$

with :

- M , the metabolic rate in watts per square meter (W/m^2)
- W , the effective mechanical power (W/m²)
- f_{cl} , the clothing area factor
- t_{cl} , the clothing surface temperature (°C)
- h_c , the heat convective transfer coefficient (W/m²/°C)
- p_a , the partial water vapor pressure in the air (Pa)
- t_a , the air temperature (°C)
- I_{cl} , the thermal resistance of clothing
- t_r , the mean radiant temperature (°C)

The PMV values are described in the Table 1, where each value is link to thermal comfort meaning.

PMV value (PMV processing)	Semantic data layer 2
+3	Hot
+2	Warm
+1	Slight Warm
0	Neutral
-1	Slight Cold
-2	Cool
-3	Cold

Table 12: PMV values and thermal comfort meaning

The Predicted Percentage of Dissatisfied (PPD) allows to estimate the percentage of a large group of user related to the thermal sensation scale (PMV) :

$$PPD = 100 - 95 . e^{(-0{,}03353 . PMV^4 + 0{,}2179 . PMV^2)}$$

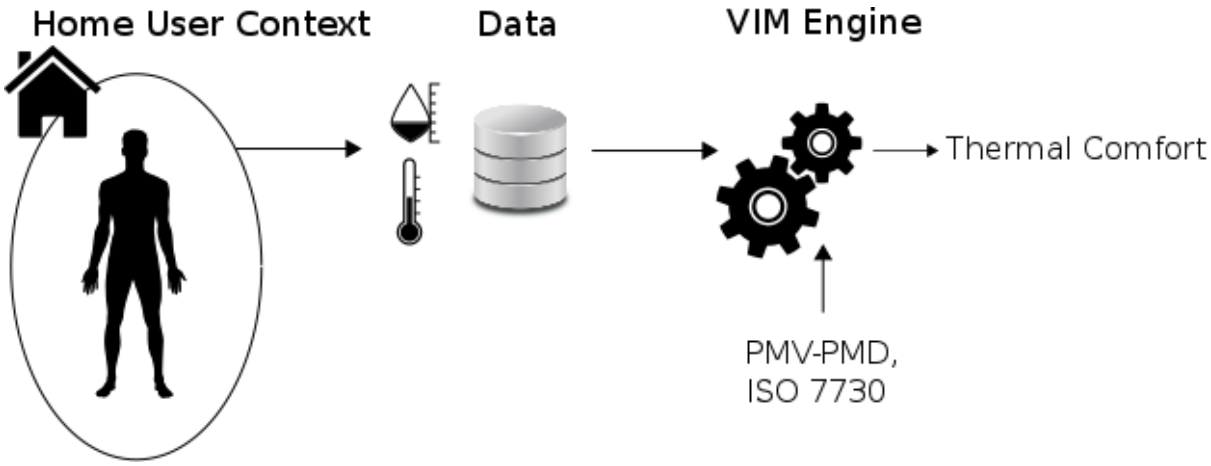

Figure 21: Overview of VIM & rules for thermal comfort

According to humidity and temperature data, the thermal comfort of the user context at home has been estimated.

8.3.2. Noise comfort

The ambient sound level will be monitor and specially in bedrooms during night. Indeed, the sleep disturbance is one risk factor that PRECIOUS try to reduce. The EN15251:2007 defines recommendation for indoor noise level (Figure 22) if specific rules in countries are not defined. Noise affects human activities : sleep, rest, work, etc. Noise level

recommendations are not the same according the user activity. For instance, there is a special noise level when sleeping.

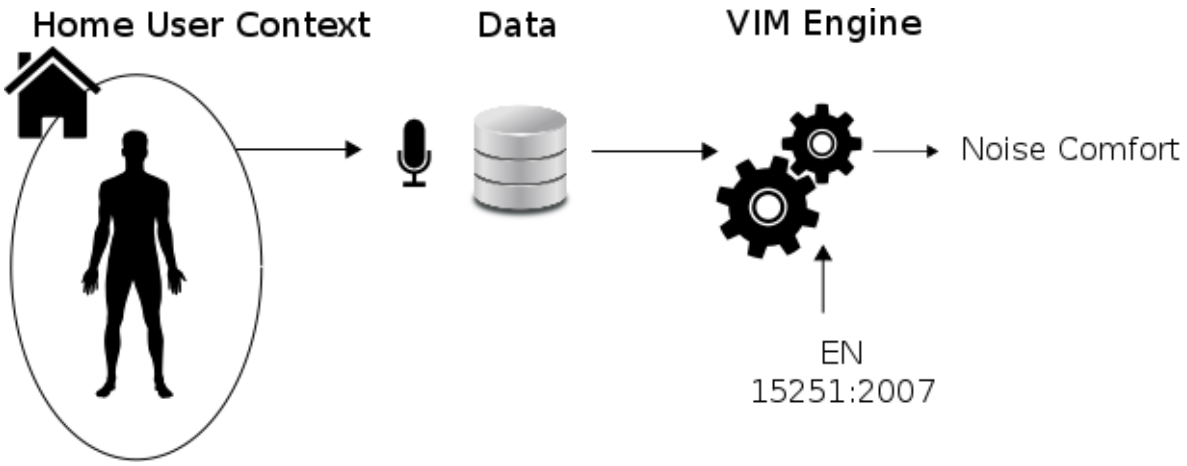

Figure 22: Overview of VIM & rules for noise comfort

The WHO guidelines for night noise recommends less than 40 dB(A) of annual average (Lnight) outside of bedrooms to prevent adverse health effects from night noise (Hurtley, 2009). The rules will be based on the “Equivalent continuous sound pressure level, (LAeq)” and Leq:

$$Leq = 10\log\left(\frac{1}{T}\int_0^T \frac{p^2(t)}{p_0^2}dt\right)$$

with

T , the measurement duration

$p(t)$, the sound pressure

p_0 , the reference sound pressure of 20µPa.

Thus, the LAeq will be compute as the average energy of the A-weighted sounder during a period T (e.g., T = 8 hours for the night period).

- Daytime & evening: LAeq = 35 dB, 16 hours
- Night time: LAeq = 30 dB, 8 hours

As explained in D3.2, we will only monitor the night time. With the threshold defined by the norm we will produce the following high level of context:

Processing	Semantic data layer 2	Human meaning
LAeq < 30 dB	Good	The exposure level to noise during sleeping activity is under the no-effect level.

LAeq > 30 dB	Bad	The exposure level to noise during sleeping activity is too high and could disturb the sleep quality.

Table 13: Threshold values and noise comfort meaning

In the PRECIOUS system, the human activity has not been considered. In the present study, we will make the assumption that the nighttime is comprised between 12 am and 6 am. It means that the period T used to process the LAeq will be adjusted to T=6 hours.

8.3.3. Light Comfort

There is not a specific reglementation for the lighting quality in dwelling. Consequently, the rules for light quality will be driven by the EN 15251:2007, EN 12464-1 and EN 12193 as depicted Figure 23.

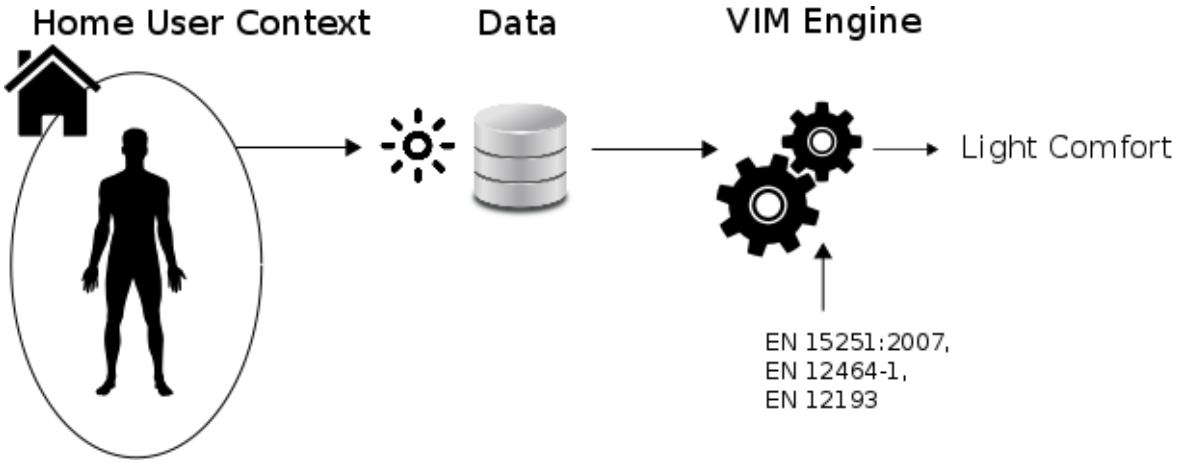

Figure 23: Overview of VIM & rules for light comfort

Standards for light quality are more dedicated to work tasks and visual comfort. For example, typical recommendation according the user tasks:

- 500lx on workplane
- 750lx for drawing (Europe)

In the bedroom, the light exposure should be also measured. Indeed, exposure to light during the night can affect the quality of the user sleeping activity [Kim et al]. A field study conducted by Burgess *et al.* [25] showed that home lighting before usual bedtime impacts also the circadian timing. However, there is no standards or european recommendations concerning the indoor light comfort. Different human activities are related to different light level. Kwon *et al.* [6] proposed a “LED Context Lighting System in Residential Areas” according to the human activities such as reading or waking up. In the PRECIOUS system, the human activity has not been considered. In the present study, we will make the assumption that the bedtime is comprised between 12 am and 6 am.

The PRECIOUS system proposes feedback to reduce risk factors. Consequently, the system only monitor the user context. In this case, we will address uniquely nighttime light level. The light level will be measured during the bedtime of a user. If the mean of illuminance is higher

than a *darkness threshold* during this period, the system will consider that the light level could have an impact on the sleeping activity.

Processing	Semantic data layer 2	Description
illuminance > darkness threshold	Bad	The exposure level to light during bedtime is too high which could impact your sleeping quality
illuminance < darkness threshold	Good	The exposure level to light during bedtime should not have impact your sleeping quality

Table 13: Threshold values and illuminance comfort meaning

9. Conclusions

Deliverables D3.1 and D3.2 provided the foundations upon which the architecture of the virtual individual model (VIM) within PRECIOUS is built. The present deliverable provides the description of data analysis techniques to obtain the VIM data set. For all VIM variables, the data acquisition, the first semantic and the second semantic layers have been explained. These high level data will be used by gamification services to engage users towards a healthy lifestyle (see D3.4).

First of all, a controlled vocabulary has been built with actors of the PRECIOUS project. It provides a common understanding of the data to all partners (users, developers, experts, health staff), establishes relationships to existing projects and data sources focused on e-health, allows the monitoring and maintaining of quality issues of the vocabulary, harmonizes data from different sensors and input providers, and the usage of a standardized data model for the entire project.

As shown in [1] Semantic analysis of textual social media can be used to reduce the total sparsity of information and uncertainty of the mood identification process. Furthermore results have shown that the results of mood classifiers built on multimodal and multivariate feature vectors for mood identification exceed the classifiers built with features of one kind.

The processings on heart rate data allows to generate report which can be visualized by users in a form of graphs, bars, points and so on. The heart rate data analysis of Firstbeat allows to produce periods of stress, recovery, and physical activity in terms of duration and intensity.

The food intake analysis has been detailed in two parts: the food intake detection and food recognition. The high level food intake context (i.e. nutrient levels for a particular food amount or portion) has been built with food recognition, raw nutritional data and EUFIC & UK health department guidelines.

The processing of raw mobile phone sensor data for physical activity monitoring and characterisation was described with a focus on the use of 3-axis accelerometer sensors integrated in smartphones. The smartphones provide a simple way to obtain useful insights on PA parameters, such as, step count, distance and energy expended, even if the accuracy is relatively limited (e.g. compared to the use of complementary wearable devices connected to the smartphone). The PRECIOUS project considers both use of smartphones only (baseline case) and a combination or smartphone and wearables in the field studies of WP5.

Indoor environmental quality has been characterized by the thermal comfort, the noise level and the light quality. The processings have been based on european norms or guidelines to build data of the semantic layer 2. Those data will be used by the domain specific rules engine to generate user feedback. In future works, the indoor human activity, i.e “is sleeping”, will be considered to improve the precision of the high level context produced at the semantic data layer 2.

10. References

[1] Roochi, E. M., Kalchgruber, P., Ramsauer, D. & Rawassizadeh, R., Leveraging Social Affect for Identifying Individual Mood, Sep 2015, Workshop on Data Science: Methods, Technology and Applications 2015 at Semantics. 4 S.

[2] M. Macht. How emotions a ect eating: A five-way model. Appetite, 50(1):1 { 11, 2008.

[3] E. Momeni, C. Cardie, and M. Ott. Properties, prediction, and prevalence of useful user-generated comments for descriptive annotation of social media objects. In

[4] Cardi, V.; Leppanen, J. & Treasure, J. 'The effects of negative and positive mood induction on eating behaviour: a meta-analysis of laboratory studies in the healthy population and eating and weight disorders.' Neuroscience and biobehavioral reviews, 2015, 57, 299-309

[5] Proceedings of the 7th International AAAI Conference on Weblogs and Social Media (ICWSM2013), Boston, USA, June 2013. AAAI.

[6] Sook-Youn Kwon, Kyoung-Mi Im, and Jae-Hyun Lim, “LED Context Lighting System in Residential Areas,” The Scientific World Journal, vol. 2014, Article ID 851930, 16 pages, 2014. doi:10.1155/2014/851930

[7] Aitchison, J., Gilchrist, A., & Bawden, D. (2000). Thesaurus construction and use: a practical manual. Psychology Press.

[8] Batini, C., & Scannapieca, M. (2006). Data Quality Dimensions. Data Quality: Concepts, Methodologies and Techniques, 19-49.

[9] Berardini, T. Z., Mundodi, S., Reiser, L., Huala, E., Garcia-Hernandez, M., Zhang, P., et al. (2004). Functional annotation of the Arabidopsis genome using controlled vocabularies. Plant physiology, 135(2), 745-755.

[10] Bizer, C., Heath, T., Berners-Lee, T. (2009): Linked data - the story so far. Int. J. Semantic Web Inf. Syst 5(3), 1–22, http://www.igi-global.com/articles/details.asp?ID=35386

[11] Davenport, T. H., & Prusak, L. (1998). Working knowledge: How organizations manage what they know. Harvard Business Press.

[12] Fisher, C., Lauría, E., & Chengalur-Smith, S. (2012). Introduction to information quality. AuthorHouse.

[13] Hogan, A., Harth, A., Passant, A., Decker, S., & Polleres, A. (2010). Weaving the pedantic web Proc. WWW2010 Workshop on Linked Data on the Web (LDOW)

[14] Jain, L. C., & Nguyen, N. T. (2008, December 19). Knowledge Processing and Decision Making in Agent-based Systems (Vol. 170). Springer.

[15] Jentzsch, A., Hassanzadeh, O., Bizer, C., Andersson, B., & Stephens, S. (2009). Enabling tailored therapeutics with linked data. Proceedings of the 2nd Workshop on Linked Data on the Web (LDOW2009).

[16] Mader, C., Haslhofer, B., & Isaac, A. (2012). Finding quality issues in SKOS vocabularies. Theory and Practice of Digital Libraries, 222-233.

[17] Firstbeat Technologies (2014). White paper on “Stress and Recovery Analysis Method based on 24-hour Heart Rate Variability”. Available online at: https://www.firstbeat.com/app/uploads/2015/10/Stress-and-recovery_white-paper_20145.pdf

[18] Michie, Susan et al. "The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions." Annals of behavioral medicine 46.1 (2013): 81-95.

[19] Miller, P., Styles, R., & Heath, T. (2008). Open Data Commons, a License for Open Data. LDOW, 369.

[20] National Information Standards Organization (US) (2005). Guidelines for the construction, format, and management of monolingual controlled vocabularies. NISO Press.

[21] Soergel, D. (2005). Thesauri and ontologies in digital libraries. Digital Libraries, 2005. JCDL'05. Proceedings of the 5th ACM/IEEE-CS Joint Conference on. IEEE.

[22] Suarez-Figueroa, M. C., Gomez-Perez, A., & Fernandez-Lopez, M. (2012). The NeOn methodology for ontology engineering. Ontology engineering in a networked world, 9-34.

[23] Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. Journal of management information systems, J. of Management Information Systems 12(4), 5–33

[24] Yu, L. (2011, January 3). A developer’s guide to the semantic Web. Springer Science & Business Media.

[25] Burgess, H. J., & Molina, T. A. (2014). Home Lighting Before Usual Bedtime Impacts Circadian Timing: A Field Study. Photochemistry and Photobiology, 90(3), 723–726.

[26] WHO. Diet, nutrition and the prevention of chronic diseases. Report of a joint WHO/FAO expert consultation. Geneva: WHO Press; 2003. http://www.who.int/dietphysicalactivity/publications/trs916/en/

[27] For summary of recent breakthrough see, for instance, http://nutrigenomics.ucdavis.edu/

[28] This is contingent on the level of integration with the dietary advise platform provided by FP7 QUALIFY http://www.qualify-fp7.eu/

[29] Regulation (EU) No 1169/2011 of the European Parliament on the provision of food information to consumers http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:32011R1169

[30] http://www.eufic.org/

[31] UK Government Guide to creating a front of pack (FoP) nutrition label for pre-packed products sold through retail outlets

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/300886/2902158_FoP_Nutrition_2014.pdf

[32] IDC, Always Connected How Smartphones And Social Keep Us Engaged, An IDC Research Report, Sponsored By Facebook, 2013.

[33] Zhao, N.: Full-Featured Pedometer. Design Realized with 3-Axis. Digital Accelerometer Analog Dialogue 44(6), 1–5 (2010)

[34] https://developers.google.com/android/reference/com/google/android/gms/location/package-summary

[35] Crosby, K (2015). Measuring your stride length. Retrieved from: https://www.walkingwithattitude.com/articles/features/how-to-measure-stride-or-step-length-for-your-pedometer

[36] Equation obtained by nonlinear regression of the data avaiable in http://calorielab.com/burned/

[37] Firstbeat Technologies (2012a). White paper on "VO_2 Estimation Method Based on Heart rate Measurement". Available online at: https://www.firstbeat.com/app/uploads/2015/10/white_paper_vo2_estimation.pdf

[38] Firstbeat Technologies (2012b). White paper on "An Energy Expenditure Estimation Method Based on Heart Rate Measurement". Available online at: https://www.firstbeat.com/app/uploads/2015/10/white_paper_energy_expenditure_estimation.pdf

[39] Firstbeat Technologies (2012c). White paper on "Indirect EPOC Prediction Method Based on Heart Rate Measurement". Available online at: https://www.firstbeat.com/app/uploads/2015/10/white_paper_epoc.pdf

[40] Firstbeat Technologies (2014). White paper on "Stress and Recovery Analysis Method based on 24-hour Heart Rate Variability". Available online at: https://www.firstbeat.com/app/uploads/2015/10/Stress-and-recovery_white-paper_20145.pdf

11. Appendix

11.1. xAAL Json Schema of ‘thermometer.basic’

```
{
  "title": "thermometer.basic",
  "description": "Simple thermometer",
  "lang": "en",
  "documentation": "http://recherche.telecom-bretagne.eu/xaal/documentation/",
  "ref": "http://recherche.telecom-bretagne.eu/xaal/documentation/thermometer.basic",
  "license": "Copyright Christophe Lohr Telecom Bretagne 2014 - Copying and distribution
of this file, with or without modification, are permitted in any medium without royalty
provided the copyright notice and this notice are preserved. This file is offered as-is,
without any warranty.",
  "extends": "any.any",
  "attributes": {
    "temperature": {
      "description": "Temperature",
      "unit": "°C",
      "type": {
          "$schema": "http://json-schema.org/schema#",
          "type": "number"
      }
    }
  },
  "methods": {
    "getAttributes": {
      "description": "Get measured temperature",
      "parameters": {
          "attributes": {
            "description": "List of wanted attributes",
            "unit": "",
            "direction": "in",
            "type": {
              "$schema": "http://json-schema.org/schema#",
              "type": "array",
              "items": {
                "enum": [ "temperature" ]
              },
              "additionalItems": false
            }
          },
          "temperature": {
            "description": "Temperature",
            "unit": "Â°C",
            "direction": "out",
            "type": {
              "$schema": "http://json-schema.org/schema#",
              "type": "number"
            }
          }
```

```
      },
      "relatedAttributes": [ ]
    }
  },
  "notifications": {
    "attributesChange": {
      "description": "Report temperature that have changed",
      "parameters": {
            "temperature": {
              "description": "Temperature",
              "unit": "Â°C",
              "type": {
                "$schema": "http://json-schema.org/schema#",
                "type": "number"
              }
            }
      }
    }
  }
}
```

11.2. EUFIC Daily Reference Intakes for Adults

The EUFIC recommended daily amounts for intake of vitamins and minerals for adults is shown in table below.

Vitamin A (µg)	800	Chloride (mg)	800
Vitamin D (µg)	5	Calcium (mg)	800
Vitamin E (mg)	12	Phosphorus (mg)	700
Vitamin K (µg)	75	Magnesium (mg)	375
Vitamin C (mg)	80	Iron (mg)	14
Thiamin (mg)	1.1	Zinc (mg)	10
Riboflavin (mg)	1.4	Copper (mg)	1
Niacin (mg)	16	Manganese (mg)	2
Vitamin B6 (mg)	1.4	Fluoride (mg)	3.5
Folic acid (µg)	200	Selenium (µg)	55
Vitamin B12 (µg)	2.5	Chromium (µg)	40
Biotin (µg)	50	Molybdenum (µg)	50
Pantothenic acid (mg)	6	Iodine (µg)	150
Potassium (mg)	2000		