Semantic Data Platform for Healthcare

	D4.2 – Terminology management platform	1
semcare	WP4: Terminological resource acquisition	
	Dissemination level: Public	
	Author: C. Riede (Averbis), J. Kors (EMC)	
	Version: 1.1 Final	

TABLE OF CONTENTS

	D4.2 – Terminology management platform	2
semcare	WP4: Terminological resource acquisition	
	Dissemination level: Public	
	Author: C. Riede (Averbis), J. Kors (EMC)	
	Version: 1.1 Final	

TABLE OF FIGURES

DOCUMENT INFORMATION

Grant Agreement Number	ICT-611388		Acronym	SEMCARE
Full title	Semantic Data Platform for Healthcare			
Project URL	www.semcare.eu			
EU Project officer	Saila Rinne (Saila.Rinne@ec.europa.eu)			

Deliverable	Number	4.2	Title	Terminology management platform
Work package	Number	4	Title	Terminological resource acquisition

Delivery date	Contractual	31/12/14	Actual	09/01/2015
Status	Version V1.1 Final		Draft ☐ Final ☑	
Nature	Report ☐ Prototype ☑ Other ☐			
Dissemination Level	Public ☑ Confidential ☐			

Authors (Partner)	C. Riede (Averbis), J. Kors (EMC)			
Responsible Author	Claudia Riede		Email	Claudia.riede@averbis.com
	Partner	AVERBIS	Phone	+49 761 203 97690

DOCUMENT HISTORY

NAME	DATE	VERSION	DESCRIPTION
C.Riede (AVERBIS)	08.11.2014	0.1	First draft
X. Wang (SGUL)	14.11.2014	0.2	Review and comments
J. Kors (EMC)	07.01.2015	1.0	Review
C. Riede (AVERBIS)	08.01.2015	1.1	Final version

D4.2 – Terminology management platform
WP4: Terminological resource acquisition
Dissemination level: Public
Author: C. Riede (Averbis), J. Kors (EMC)
Version: 1.1 Final

4

DEFINITIONS

- Partners (also named as beneficiaries) of the SEMCARE Consortium are referred to herein according to the following codes:

  AVERBIS - Averbis GmbH (Germany) Coordinator

  EMC - Erasmus Universitair Medisch Centrum Rotterdam (Netherlands) – Beneficiary

  MUG - Medical University of Graz (Austria) – Beneficiary

  SGUL - Saint George's University of London (UK) – Beneficiary

  SYNAPSE - Synapse Research Management Partners S.L. (Spain) – Beneficiary

- **Project:** The sum of all activities carried out in the framework of the Grant Agreement.

- **Consortium**: The SEMCARE Consortium, conformed by the above-mentioned legal entities.

- **AEP:** Averbis Extraction Platform; text analysis tool to extract information units such as facts and relations from unstructured text

- **EHR:** Electronic health record; clinical data record of a patient.

- **Terminology:** General term for information artefacts that provide controlled terms for a domain, identifiers of meaning and semantic relations. e.g. SNOMED, ICD-10, MeSH

- **Term Browser:** Tool provided by Averbis to load, view, modify and export terminologies. It can also be used to create new terminologies.

- **UMLS:** Unified Medical Language System, metathesaurus.

- **OBO:** OBO format is the text file format used by OBO-Edit, the open source, platform-independent application for viewing and editing ontologies.

- **SKOS:** Simple Knowledge Organization System. SKOS is part of the Semantic Web family of standards

- **LOINC:** Logical Observation Identifiers Names and Codes. LOINC was initiated in 1994 by the Regenstrief Institute and provide a universal code system for tests, measurements, and observations.

- **SNOMED CT:** systematically organized computer processable collection of medical terms

- **UIMA:** Unstructured Information Management Architecture; framework by Apache enabling the generation of analysis pipelines for arbitrary content such as text, image or video data

D4.2 – Terminology management platform
WP4: Terminological resource acquisition
Dissemination level: Public
Author: C. Riede (Averbis), J. Kors (EMC)
Version: 1.1 Final

5

EXECUTIVE SUMMARY

The aim of the Terminology Management Platform is to provide a user-friendly software which can support the management of terminologies needed within the SEMCARE project. This document presents a user manual explaining how to import terminologies into the platform, how to browse the terminology, how to add or modify terms and how to export the terminology for usage in the text analytics pipeline and the integrated search platform. Besides the SEMCARE terminology, which has been developed especially for the use case, further terminologies which are also used for the text analytics pipeline or support the refinement of the SEMCARE terminology are additionally available in the Terminology Management Platform. They are described shortly in this document.

KEY WORDS (Wordle style)[1]

D4.2 – Terminology management platform
WP4: Terminological resource acquisition
Dissemination level: Public
Author: C. Riede (Averbis), J. Kors (EMC)
Version: 1.1 Final

1. Introduction

The research project SEMCARE aims to build a semantic data platform to support case identification for clinical studies and to facilitate clinical decision making based on patient-level criteria (e.g., age, gender, diagnosis, indication, symptoms, lab results).

Most patient-level data can be found in unstructured text documents. Text mining tools will be used to extract and convert relevant data. These text-mining tools require multilingual terminologies in English, German and Dutch to standardize and address specific idiosyncrasies of medical language.

The SEMCARE Terminology Management Platform is using the Averbis Terminology Platform, called Term Browser. It provides software services to support the management of the terminologies used for the SEMCARE text analytics pipeline. It is possible to import terminologies into the platform, to browse easily the terminologies, to add and modify concepts and terms. The modified terminology can be exported for utilisation in the text-mining platform.

2. Terminology Management Platform

2.1. Prototype Data

The prototype is available via the internet with the URL http://www.termbrowser.de.

Username and password will be sent on request, please send a message via the SEMCARE homepage (http://semcare.eu/contact/).

2.2. Available Terminologies

The partners of the SEMCARE project have agreed on an initial use case **'Risk Stratification and Differential Diagnosis of Patients suffering from transient loss of consciousness'.** Based on the topic of interest as defined in deliverable D3.2 Requirement specification[2] a distinct SEMCARE terminology has been created. Further terminologies are used to enrich the SEMCARE terminology but are also needed to cover laboratory data, medication and diagnosis.

The following terminologies are currently available in the Terminology Management Platform:

---

[2] D3.2: Requirements Specification, chapter 2.1.3.

D4.2 – Terminology management platform
WP4: Terminological resource acquisition
Dissemination level: Public
Author: C. Riede (Averbis), J. Kors (EMC)
Version: 1.1 Final

8

- UMLS SEMCARE:

  multilingual terminology provided by EMC[3]

- SEMCARE:

  terminology created for the defined use case. The UMLS SEMCARE terminology has been used to enrich the vocabulary. Additional concepts and terms have been added based on the free-text phrases found in the test documents. The terminology is multilingual, for all concepts English, German and Dutch terms are included.

- semCare Modifier:

  list of descriptive and evaluative qualifiers, which are linked to the corresponding concepts in the SEMCARE terminology by cross references. The terminology is multilingual, for all concepts English, German and Dutch terms are included.

- Labor 2.0:

  terminology for laboratory parameters developed by Averbis. Its hierarchical structure is based on the hierarchy used in LOINC. The terminology is multilingual, for all concepts English, German and Dutch terms are included.

- SNOMED CT:

  English version from 2014

- ICD AHA 2014:

  German version of ICD-10-GM 2014, enriched by Averbis with further synonyms in order to use it for text mining purposes.

Further terminologies can be imported into the Terminology Management Platform, especially terminologies covering the medication in the different countries.

---

[3] D4.1: Initial multilingual biomedical terminology.

---

3. User Manual

The terminology management platform offers both reading and editing functions. Terminologies can be exported in different formats, e.g., OBO[4] or SKOS[5].
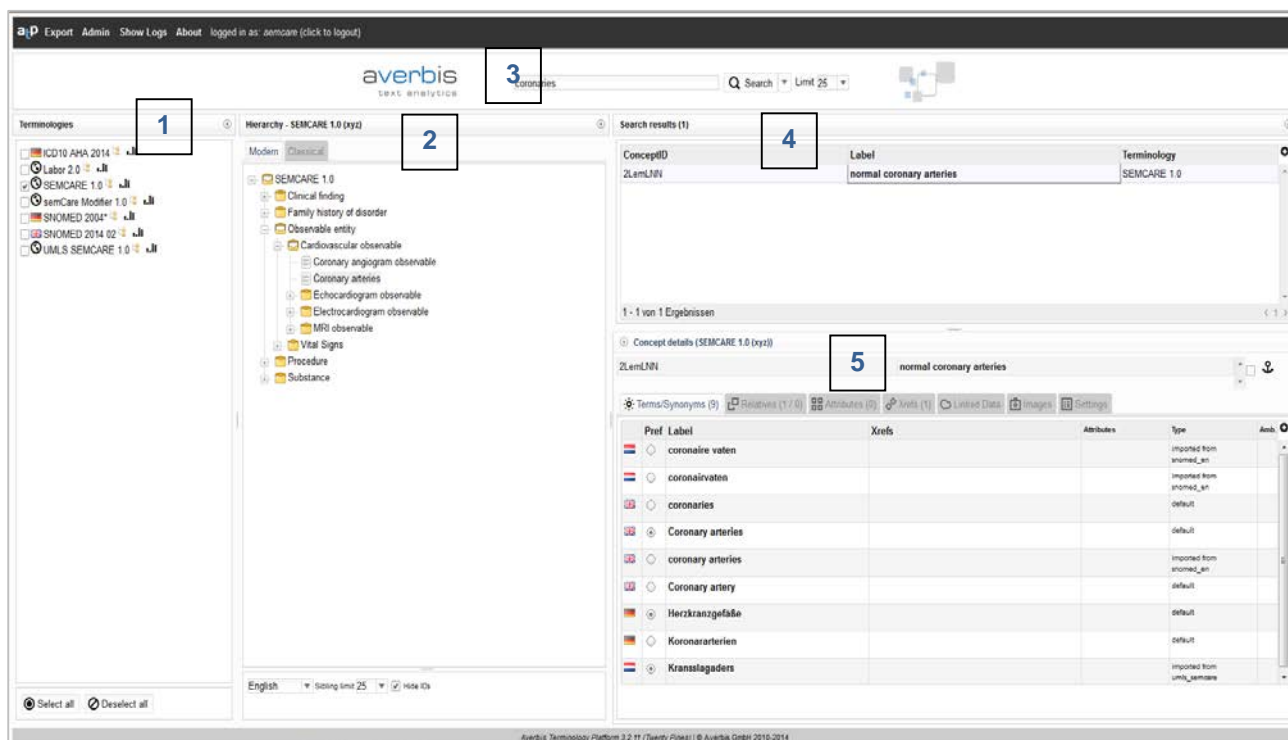
3.1. User Interface

Figure 1: User Interface

Description of the areas:

1. List of the available terminologies. The terminologies to be used in the search can be selected by ticking the box before the name of the terminology.
   By clicking on the name of the terminology this terminology can be browsed in the area rightwards.
2. The hierarchical structure of a selected terminology. The result of a free-text search will be shown here in its hierarchical position.
3. The search box which allows a free-text search.
4. List of the search results.

---

[4] OBO Flat File Format: http://oboformat.googlecode.com/svn/trunk/doc/GO.format.obo-1_4.html.

[5] Simple Knowledge Organization System: http://www.w3.org/2004/02/skos/.

D4.2 – Terminology management platform
WP4: Terminological resource acquisition
Dissemination level: Public
Author: C. Riede (Averbis), J. Kors (EMC)
Version: 1.1 Final

10

5. Details of the concept.

An area can be faded out in order to enlarge the disposable space for the remaining areas by clicking on ⏴ top right.

3.2. Search

A basic functionality of the Terminology Platform is the possibility to search terms via a free-text search in different terminologies. All terminologies which are marked in the left area *Terminologies* are included into the search. The search covers all synonyms of a concept.

Search box

The search box allows a free text search. During the entry of the word the AutoSuggest function helps to find the desired term.
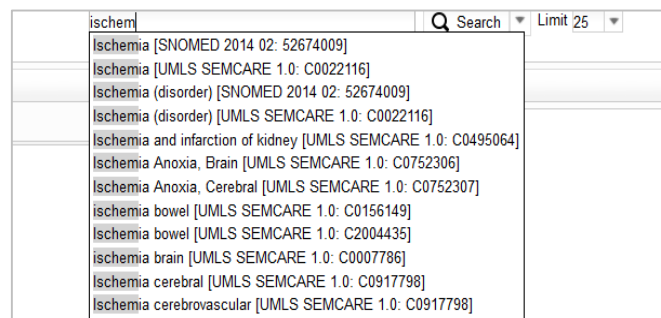
Figure 2: Example of the AutoSuggest function

Search results

The search results are listed on the right hand side**.**
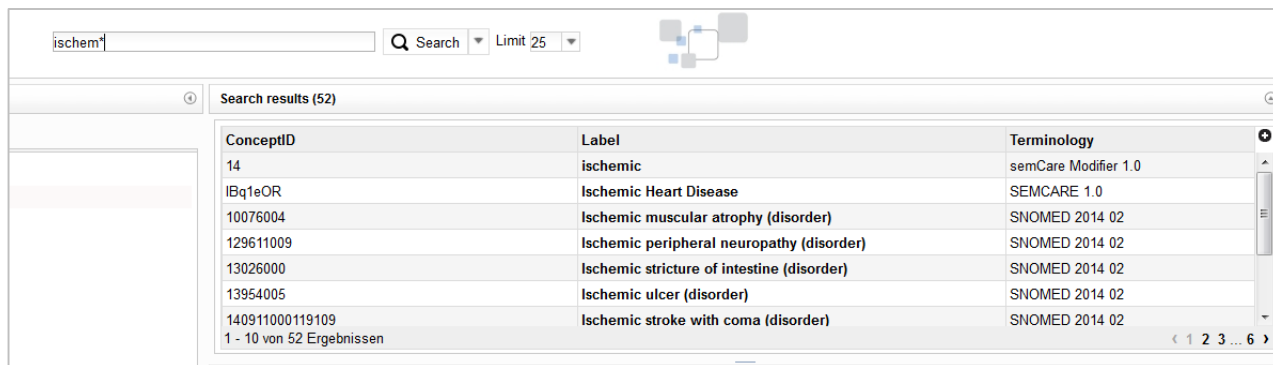
Figure 3: Search result for "ST segment"

Displayed are the ID and preferred term of the concepts and the name of the corresponding terminology. By clicking on one of the search results more details of the selected concept are shown in the area

below and additionally the hierarchical tree structure of the terminology is expanded to show the position of the concept.

The asterisk ("*") can be used as a wildcard character to expand the search.

Figure 4: Search result using the wildcard character in "ischem*"

3.3. Hierarchical Navigation

Besides searching for a specific term it is possible to navigate through the hierarchic structure of a terminology. By clicking on the name of a terminology in the left area *Terminologies* this terminology is displayed in the area *Hierarchy* opening the highest level of the hierarchical tree. More detailed sub-trees can be opened by clicking on the "+" symbol before the appropriate superordinate concept.
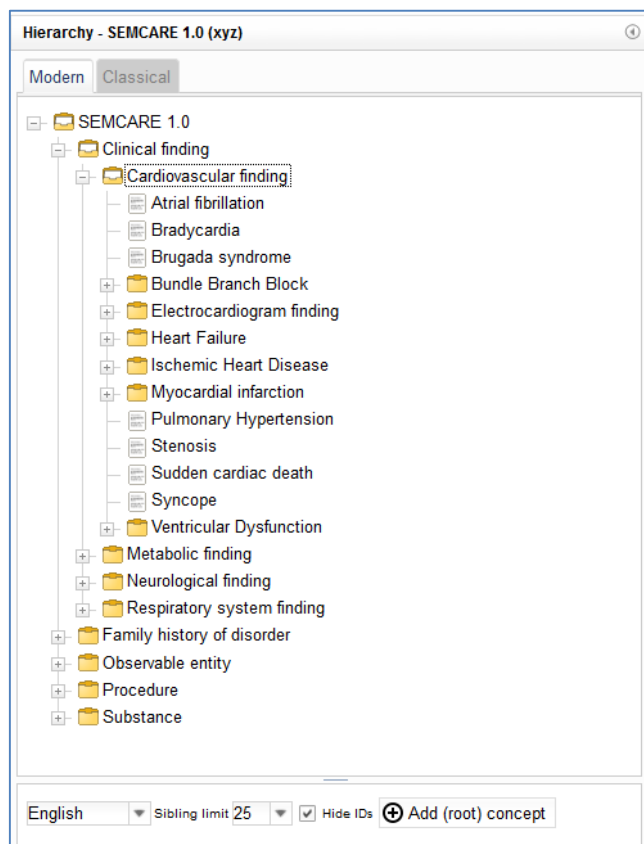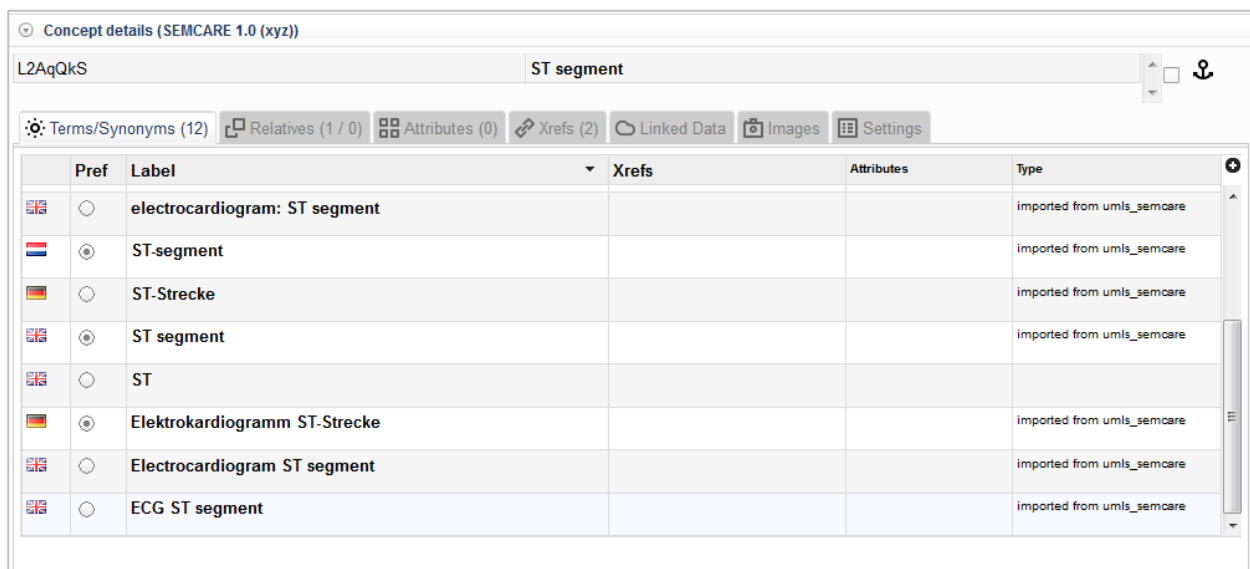
Figure 5: Part of the hierarchy of the SEMCARE terminology

If the selected terminology is multilingual, the language can be selected below. By default the concept IDs are not visible; this can be changed by removing the tick besides *Hide IDs*. The classical view is only available for monolingual terminologies.

3.4. Concept Details

For each concept there is additional information available, called *Concept details*, which are displayed in the lower right area.
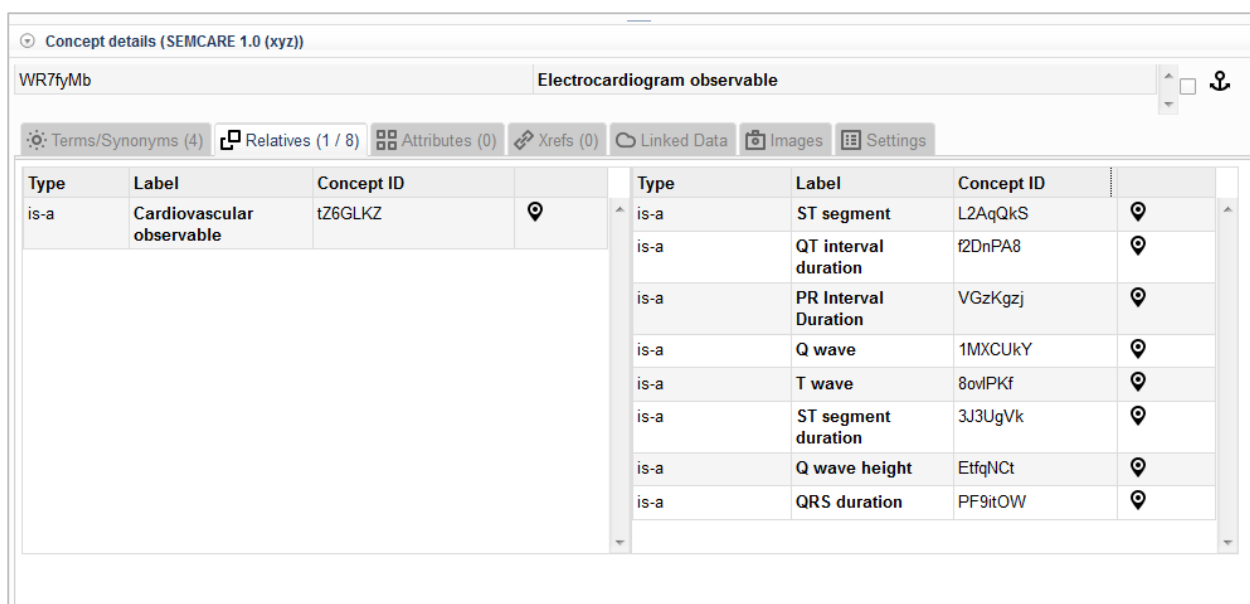
Figure 6: Concept details for "ST segment"

In the first tab, **Terms/Synonyms**, all terms or synonyms of the concept are listed. For each term the language, the type and possible attributes are specified. In each language a preferred term can be defined, this is indicated in the column *Pref*. Term attributes provide additional information in a typified format, where types can be defined by the user. E.g., common typing errors can be added to a concept as synonyms and can be marked by the attribute "typing error".

In the second tab, **Relatives**, the direct relatives of the concept within the terminology are listed.

Figure 7: Concept details: Relatives

The direct parents are listed in the left table and the direct children are listed in the right table. With a click on the pin symbol next to a concept ID it is possible to switch to the corresponding concept.

Concept Attributes can be used for specifying additional information on concept level, e.g. describing the scope of the concept ("Ejection fraction (MRI)": only to be used when the ejection fraction is measured by a MRI).

In the tab **Xrefs** the relations with concepts in other terminologies are listed. The relationship between the concept and the referenced terminology can be described by the type. If no type is specified the terminology is containing the same concept.

Figure 8: Concept details: Xrefs

If the referenced terminology is part of the Terminology Management Platform it is possible to go directly to the referenced concept by clicking on the pin symbol.

The tabs *Linked Data*, *Images* and *Settings* can be used in prospective versions to display information form external sources or images.

3.5. Editing existing terminologies

Import

Users with the appropriate rights can import additional content into an existing terminology under the menu item *Admin*.
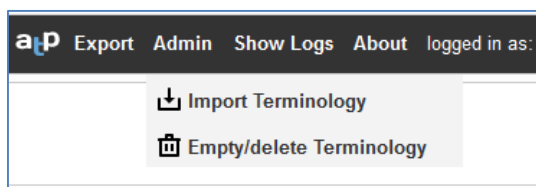

Figure 9: Import Terminology

D4.2 – Terminology management platform
WP4: Terminological resource acquisition
Dissemination level: Public
Author: C. Riede (Averbis), J. Kors (EMC)
Version: 1.1 Final

15

In the opening dialogue window the target terminology and the source file have to be chosen. Additionally the existing terminology can be emptied before import. Allowed import formats are csv and OBO.
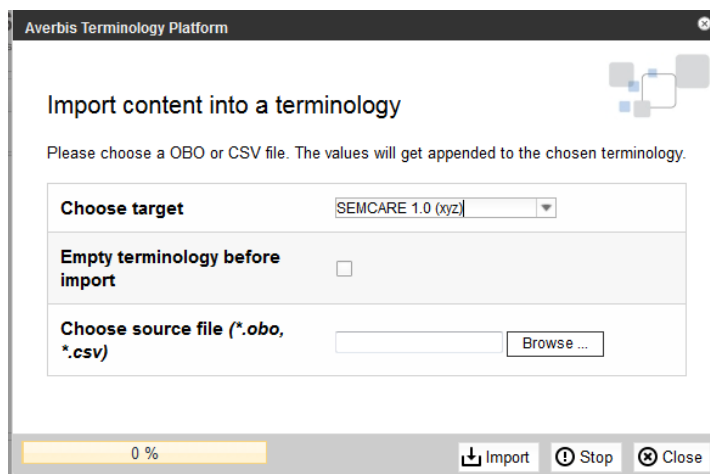

Figure 10: Dialogue window: Import

The fields of a source file in csv format must be separated by ";". The following fields are expected:

Table 1: Import fields

ConceptID	TermID	TermLabel	TermType	Language	isPreferred	parentConceptIDs	MappingForms[6]
ID of the concept	ID of the term (can be empty)	label of the term	type of the term (if empty then type = default)	identification code of the language, e.g. de, en	1 = preferred term

0 = no preferred term | concept ID of the parent concept

[optional] | mapping forms of the term

[optional} |

Manual editing

Terminologies can be edited manually provided that the terminology is editable and that the user has write permission.

Concepts:

With a right-mouse click on the concept in the hierarchical view the concept can be deleted, its hierarchy

---

[6] if a homonym term has an additional specification which will not be found in a text: (e.g. note <finan.> and note<mus.>) the text analytics pipeline can use the mapping form (e.g. note)

D4.2 – Terminology management platform
WP4: Terminological resource acquisition
Dissemination level: Public
Author: C. Riede (Averbis), J. Kors (EMC)
Version: 1.1 Final

16

can be edited or a new child can be added. Below the hierarchical tree there is also a ⊕ – Button to add new concepts.
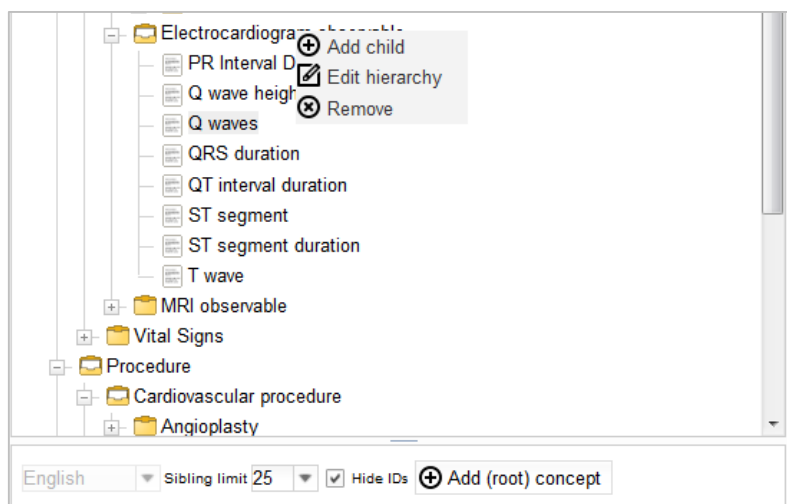

Figure 11: Concept hierarchy: Manual changes
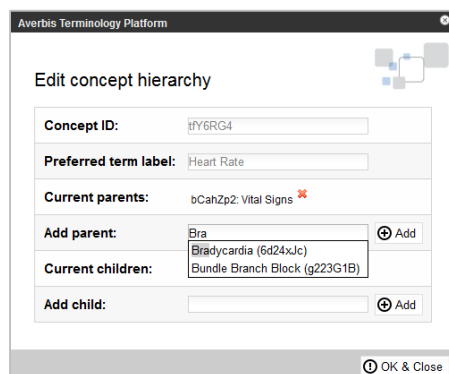
- **Editing concept hierarchy**

Figure 12: Editing concept hierarchy

Both parents and children can be added. In the fields after *Add parents* and *Add child* any term of the desired concept can be entered. The system will suggest applicable concepts. Existing parent or children relationships can be deleted by clicking on the red X.
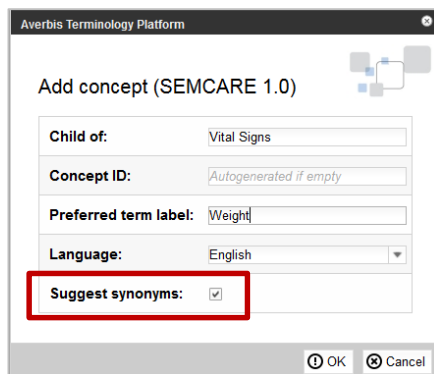
- **New concepts**:

Figure 13: Add concept

The concept ID can be entered manually, otherwise the ID is generated automatically.

By ticking the box *Suggest synonyms* all synonyms from terminologies that were marked as selected in the left area of the Terminology Platform, are proposed. The terms can be selected individually. The checkmark on concept level inserts an XREF to the referencing terminology.
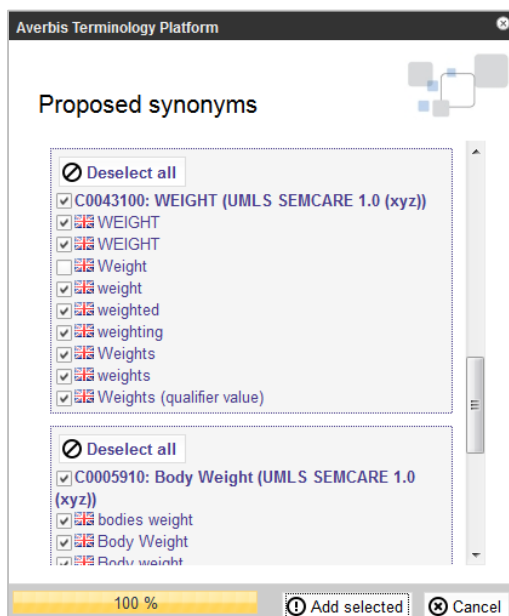

Figure 14: Proposing synonyms

	D4.2 – Terminology management platform	
	WP4: Terminological resource acquisition	18
semcare	Dissemination level: Public	
	Author: C. Riede (Averbis), J. Kors (EMC)	
	Version: 1.1 Final	

Terms

The terms of a concept can be edited in the *Concept details* area.
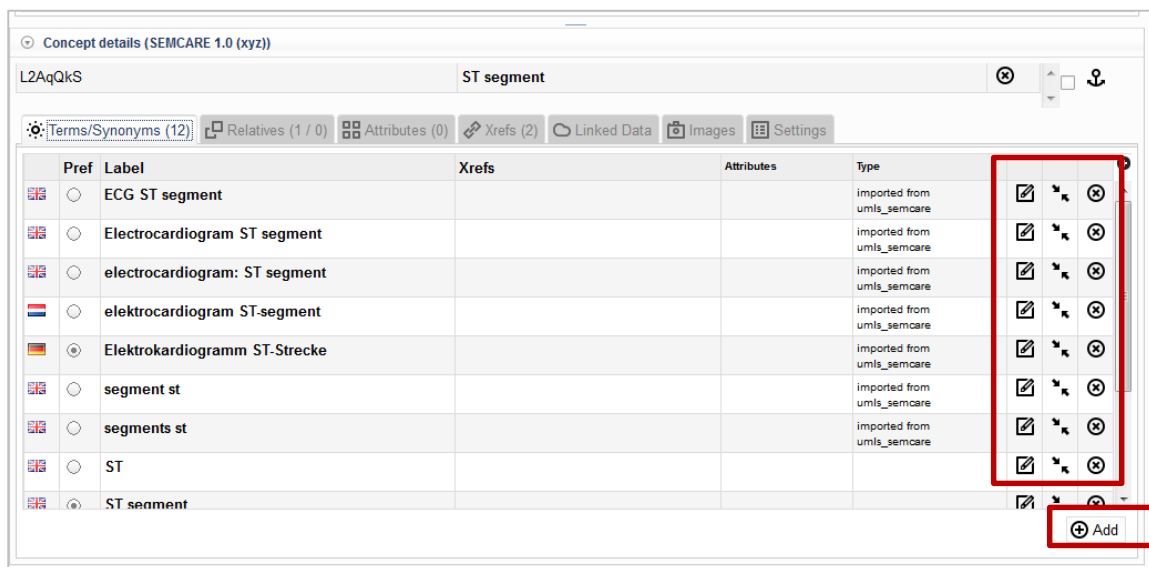

Figure 15: Editing terms

⊗   to delete the term

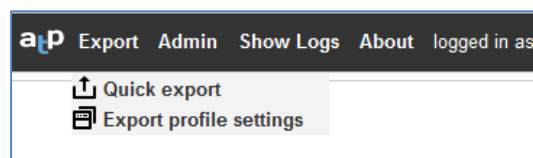↘↖   to search for synonyms in the terminologies available in the Terminology Platform

✎   to edit the term

Additional synonyms can be entered by clicking on the ⊕ – Button at the bottom right.

3.6. Export of terminologies

Users with the appropriate rights can export a terminology under the menu item *Export*.

Quick export

With the *Quick export* function terminologies can be exported in the data formats OBO, SKOS, and csv. The export in csv format includes only the fields that were accepted when the terminology was imported (see Table 1: Import fields).
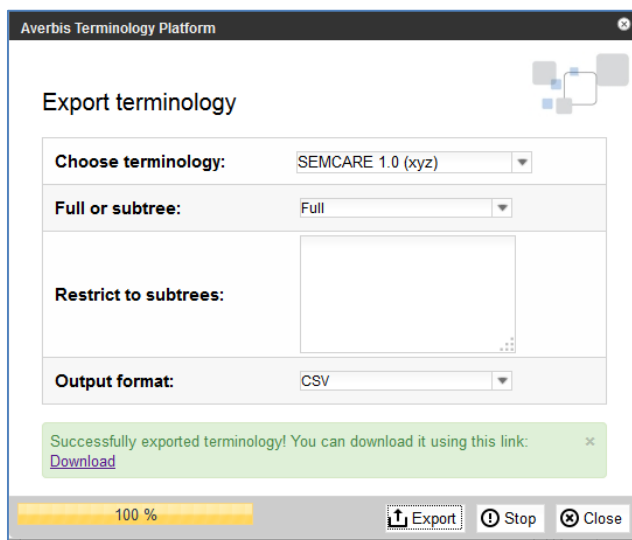
Figure 16: Quick export

The exported terminology can be downloaded directly in the dialogue window.

Export profile settings

For the export of terminologies into a ConceptMapper dictionary which is used in the text mining pipeline the *Export profile settings* modus is required.

First, an export profile has to be defined or an existing profile has to be selected. The name of the selected profile is displayed at the top of the window
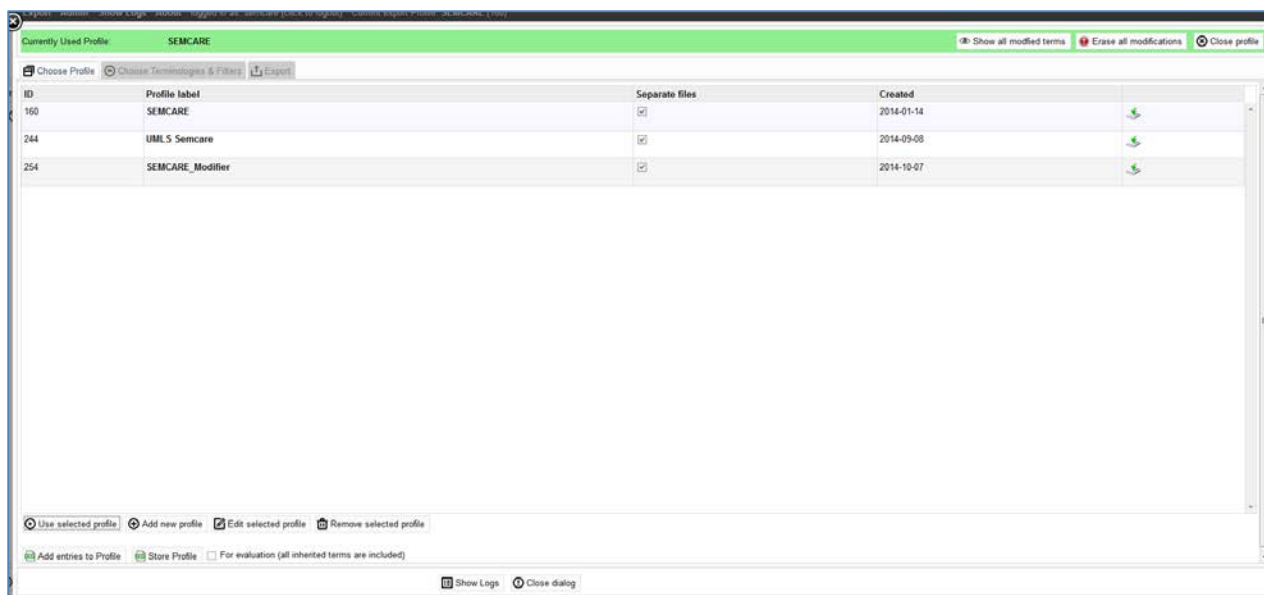

Figure 17: Choose Profile

D4.2 – Terminology management platform
WP4: Terminological resource acquisition
Dissemination level: Public
Author: C. Riede (Averbis), J. Kors (EMC)
Version: 1.1 Final

20

For each export profile at least one terminology has to be allocated. There are several predefined filters which can be applied, e.g. the *Default Filter English* identifies all terms containing acronyms and stop words and sets them to the EXACT modus for the export. Using the Averbis Concept Mapper format for the export the term or concept will then be written into a separate dictionary (see below). The results of a filter can be downloaded as csv file for evaluation instead of applying the filter directly.
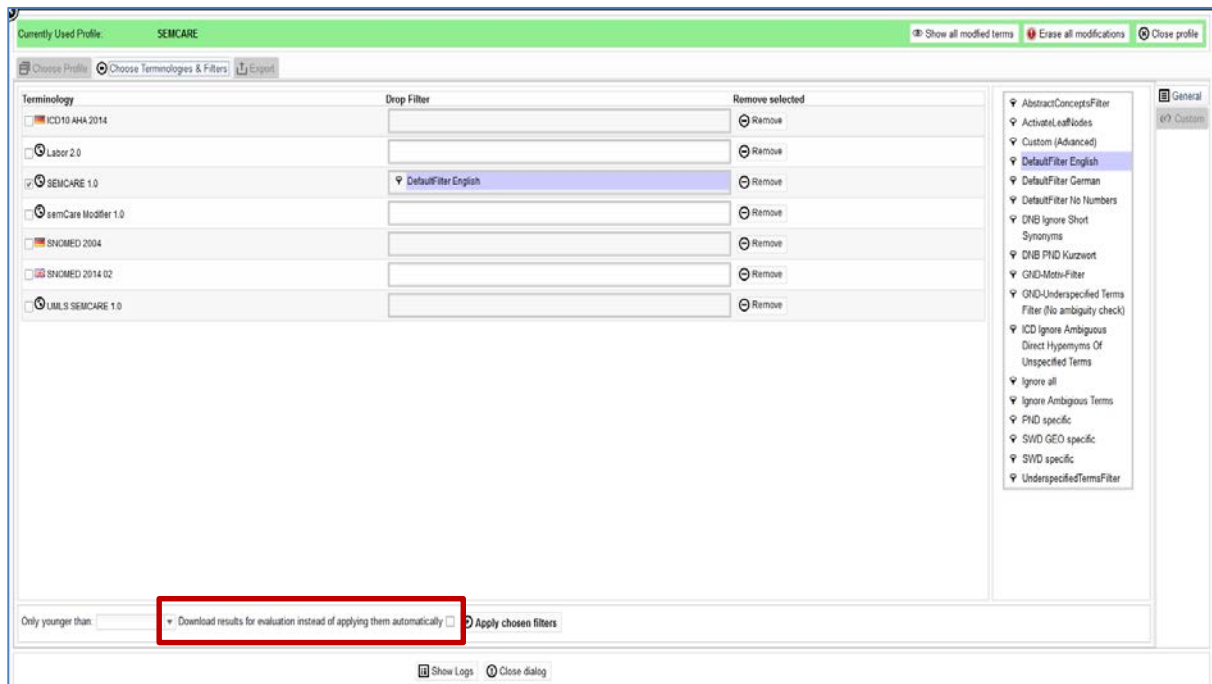

Figure 18: Export filters

After closing the export dialogue window (clicking on *Close dialog* at the bottom of the window), the export mode of a term or a concept can be set.
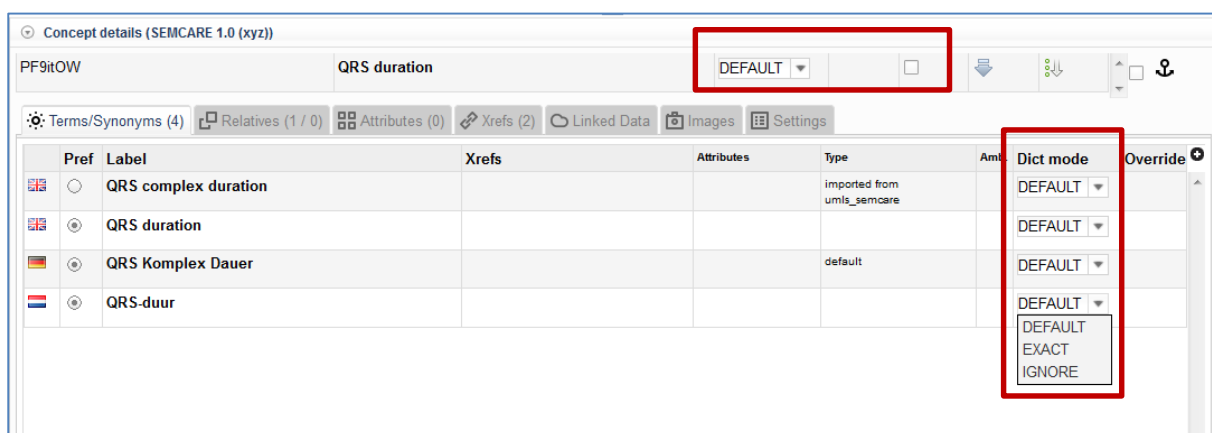

Figure 19: Manual setting of export modes

By setting the export mode to IGNORE the term or concept will not be included into an export. The IGNORE mode can be extended to subordinate concepts if the tick box on the right side of the drop-down menu on concept level is marked.

By setting the export mode to EXACT the term or concept is exported into a separate dictionary. The text mining pipeline will use the terms from this exact dictionary only in this exact form and allows no variations like differing lower or upper case or grammatical inflections.

For the export the following formats are available: Averbis Concept mapper Format, Averbis MSI Format, Averbis Lemmatizer Format, OBO, SKOS and csv. The text analytics pipeline expects the Averbis Concept mapper Format, which considers all three export modes.
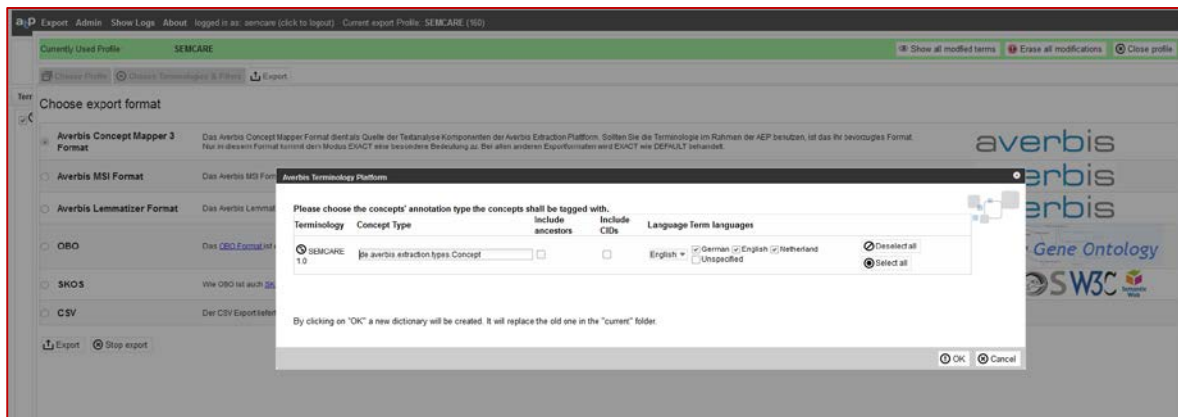

Figure 20: Averbis Concept Mapper export

In the dialogue window the concept type and the language(s) can be defined. The resulting dictionary is stored automatically in a defined data folder in the Averbis file system, but its current version can be downloaded in the register *Choose Profile*.