Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 4903–4910
Marseille, 11–16 May 2020

On the Formal Standardization of Terminology Resources: The Case Study of TriMED

F. Vezzani[1], **G. M. Di Nunzio**[2,3]
[1]Dept. of Linguistic and Literary Studies, [2]Dept. of Information Engineering, [3]Dept. of Mathematics
University of Padua, Italy
federica.vezzani@phd.unipd.it, giorgiomaria.dinunzio@unipd.it

Abstract

The process of standardization plays an important role in the management of terminological resources. In this context, we present the work of re-modeling an existing multilingual terminological database for the medical domain, named TriMED. This resource was conceived in order to tackle some problems related to the complexity of medical terminology and to respond to different users' needs. We provide a methodology that should be followed in order to make a termbase compliant to the three most recent ISO/TC 37 standards. In particular, we focus on the definition of i) the structural meta-model of the resource, ii) the data categories provided, and iii) the TBX format for its implementation. In addition to the formal standardization of the resource, we describe the realization of a new data category repository for the management of the TriMED terminological data and a Web application that can be used to access the multilingual terminological records.

Keywords: terminology standardization, medical language, methodologies for LRs construction.

1. Introduction

Terminology standardization has long been known as a fundamental requirement in dealing with special languages (Wüster, 1971; Galinski and Nedobity, 1988; Sager, 1998). This process reflects on two aspects related to terminology, as the study of the "micro-language" used by specialists in particular domains of the human activity (Balboni, 2000). Firstly, from a semantic perspective, the standardization of terminology consists in dealing with the meaning of terms. This process aims to 1) create a shared common consensus on the use of technical terms among a community of experts, and 2) disambiguate their meaning to the advantage of monoreferentiality (a term has a single referent in the real world) and monosemy (a term has a single and univocal meaning) (Magris et al., 2002). The importance of this aspect is reflected in such contexts where the transmission of information is "vital". Communication in the healthcare domain, for example, is often characterized by an opaque lexicon which needs to be disambiguated (Rouleau, 2003; Castro et al., 2007).

Secondly, from a formal perspective, the standardization of terminology, interpreted in this context as a collection of terminological data, reflects on the organization of structured data for linguistic resources. To this purpose, the International Organization for Standardization (ISO),[1] and in particular the Technical Committee ISO/TC 37 working on *Language and Terminology*, provides a large number of standards for "language resource management" (subcommittee 4) and "management of terminology resources" (sub-committee 3). One of the main goal is to provide structured and standardized terminological data which can be reused in different terminology management systems. This aspect is important not only for the correct implementation of a terminological resource, but also in a broader sense of 'Open Science', where the concept of "reusability" is one of the fundamental keystone of the FAIR (Findability, Accessibility, Interoperability and Reusability) principles (Wilkinson et al., 2016) of the European Open Science Cloud (ESOC).[2]

1.1. Our Contribution

In this paper, we focus on the terminology standardization by restricting our field of analysis to the medical domain. In this context, we describe the process of formal standardization of an existing multilingual database named TriMED (Vezzani et al., 2018), which was conceived in order to tackle some problems related to the complexity of medical terminology, such as the transmission and the comprehension of medical information. Our resource is designed in order to respond to the ISO standards for the implementation and management of terminology resources. In particular, we present:

1. The standardization of the structure of TriMED according to Terminological Markup Framework (TMF) meta-model (ISO-16642, 2017);
2. The description of its Data Categories Specifications through the implementation of the TriMED Data Category Repository (ISO-12620, 2019);
3. The implementation of the resource in the TermBase eXchange (TBX) format (ISO-30042, 2019);
4. The Web application to access the data stored in the Termbase.

The reminder of this paper is organized as follows: in Section 2., we present, in terms of both semantic and formal standardization, the existing linguistic resources available for the medical domain and the purposes behind their implementation. In Section 3., we describe the application of the above mentioned standards applied to TriMED resource by focusing on the 1) structural meta-model 2) data categories, and 3) TBX implementation. In Section 4. we present our Web Application and, finally, in Section 5., we give our conclusions and some hints on future works.

[1] https://www.iso.org/home.html

[2] https://www.eosc-portal.eu

2. Related Works

The standardization of medical terminology aims, first of all, at the optimization of the communication between experts working in their sub-domains of specialization (Wermuth and Verplaetse, 2019). To this purpose, nomenclatures, vocabularies, terminologies and codes have been developed to enable effective communication between medical experts and to facilitate the recording of patient data. The Unified Medical Language System[3] (UMLS) is a compendium gathering a large number of biomedical resources, such as: 1) SNOMED CT,[4] the reference terminology for clinical concepts which is used for clinical documentation and reporting; 2) MeSH Thesaurus,[5] a controlled vocabulary for the indexing and retrieval of the biomedical literature; 3) ICPC2,[6] a classification system which allows to classify patient data and clinical activity; 4) ICD11,[7] a classification system used to codify diagnoses for recording morbidity and mortality statistics. The formal standardization of these medical resources is regulated by the international standard ISO 17117-1:2018 developed by the Technical Committee 215 working on *Health informatics* (ISO-171171, 2018). According to the standard, these resources should have a conceptual orientation: they are configured as ontologies, that is collections of clinical concepts and the corresponding medical terms which designate their meaning.

The above mentioned resources are mainly used by experts for purely medical purposes in order to promote peer communication and patient care. However, many studies focus also on the problem of comprehension of medical information in the patient-physician dialogue (McCray, 2005; Jucks and Bromme, 2007; Tran et al., 2009; Elhadad and Sutaria, 2007). Indeed, physicians tend to use medical jargon, also referred to as *medicalese* (Hadden et al., 2018), which consists in words, phrases or concepts that can be misunderstood or misinterpreted by patients, or in general non-experts (Deuster et al., 2008; Castro et al., 2007).

In this context, there are numerous resources conceived to be used by non-experts in order to facilitate the communication and the transmission of information in the healthcare domain. The aim is to favour the understanding of the medical term rather than the standardization of its meaning. For example, the Consumer Health Vocabulary Initiative (Zeng and Tse, 2006) has led to the development of a useful resource,[8] for English language, allowing to translate technical terms into a consumer friendly language. In parallel, the initiative of (Cardillo et al., 2009) has led to the development of the Italian Consumer-oriented Medical Vocabulary,[9] that is an Italian vocabulary showing the different ways patients and healthcare consumers in general express and think about health topics. In this context, the Spanish Ministry of Economy and Competitiveness funded another project which has led also to the development of the terminological database VariMed[10]: this resource focuses on terminological variations in the medical field along with pragmatic information and is intended to facilitate communication between healthcare professionals and patients (Sánchez and Velasco, 2013).

From a multilingual viewpoint, several initiatives have been carried out at the European level. The European Commission (DG III) commissioned the Multilingual Glossary of Popular and Technical Medical Terms[11] project which was realized by the Department of applied Linguistics of the Heymans Institute of Pharmacology and Mercator School in 1995-2000. The resource groups nine glossaries of 1830 scientific and popular medical terms about medicinal product package inserts in nine official languages of the European Union. In 2018, the Terminology Coordination Unit (TermCoord) of the European Parliament in collaboration with the Paris-Diderot University, the University of Granada and the University of Padua, engaged in the YourTerm MED[12] project consisting in the realization of a structured linguistic database based on the cognitive structure of the Frame-Based Terminology (Faber, 2015). This multilingual tool is conceived in order to facilitate communication between healthcare professionals on mission and their patients: in particular, it is intended to meet the immediate terminological needs of Médecins sans Frontières/Doctors Without Borders (MSF) in their international missions, so that physicians can consult a multilingual resource on site which allows to take care of patients.

Finally, we want to mention also two important initiatives which have led to the development of structured and standardized medical resources. The TermSciences[13] initiative deals with the construction of a multi-purpose and multilingual terminological database from various source vocabularies produced by major French research institutions (Khayari et al., 2006). In the context of the formal standardization of terminological resource, this was the first public initiative to implement the ISO standard dealing with the Terminological Markup Framework meta-model for the interchange of computerized lexical data.[14] Secondly, the Medical Enduser and Reference Interface Terminology[15] (MERITERM) project, developed by (Roumier et al., 2011), deals with the implementation of a multilingual interface terminology linked to large existing linguistic resources and to a broad array of international medical classifications according to terminology standards and linked data principles (Warnier et al., 2012).

3. Formal Standardization of TriMED

In the context of the formal standardization of terminological resources, we describe the work of remodeling an existing multilingual termbase for the medical domain named

[3] http://bit.ly/2IWDfuQ
[4] http://www.snomed.org
[5] https://meshb.nlm.nih.gov/search
[6] http://bit.ly/2IHNK6e
[7] https://icd.who.int/en/
[8] http://bit.ly/2L3tSw4
[9] http://bit.ly/2L31Fpp
[10] http://varimed.ugr.es
[11] http://bit.ly/2UEMxxK
[12] http://bit.ly/2VqClO0
[13] http://www.termsciences.fr
[14] The authors refer to the previous version of the ISO 16642 on *Computer Applications in Terminology* published in 2003 and which has been replaced by the current version dated 2017 (ISO-16642, 2017).
[15] http://www.meriterm.org

TriMED (Vezzani et al., 2018). This termbase collects structured terminological records providing a wide range of linguistic information depending on the user identification in three languages: Italian, English and French. Indeed, the resource was conceived in order to: 1) help patients or, in general, non experts to understand difficult technical terms by providing their equivalent in a more familiar register; 2) support scientific translators both in the "decoding" (understanding) and "transcoding" (translating) of health information (Jammal, 1999) from a source language into a target language; 3) create an unique access point for healthcare professionals by the integration of standard medical terminologies, ontologies and classification systems.

The terminological records in TriMED are structured in *tibbles* ('modern' tables in the R programming language (Müller and Wickham, 2019)) where each row contains the information about a term and each column a linguistic feature of that term. Therefore, the objective of this work is to design a general methodology, described in the following sections, to re-model a plain table-based terminological dataset into a new data model compliant with the three ISO standards.

3.1. Structural Meta-Model

For the representation of the TriMED database structure, we have referred to the ISO standard 16642 about *Terminological Markup Framework* (TMF) meta-model (ISO-16642, 2017). TMF is an international standard that provides a framework for the representation of Terminological Data Collections (TDCs), such as multilingual terminological databases, in eXstensible Markup Language (XML). The TMF standard is based on two levels of abstraction. The first and most abstract level concerns the description of the meta-model that supports analysis, design and exchanges at a very general level, so that the meta-model is independent from any specific implementation or software. The second abstract level is the data model concerning the categories of data that can be associated within the meta-model levels and which are specific to each TDCs (see subsection 3.2.). This standard does not represent a particular format but rather a meta-model which is based on the traditional concept-oriented view of a terminological entry dating back to Wüster's early works (Picht and Schmitz, 2001): a concept is described in *n* languages and is designated by *m* terms for each language.

In particular, the structure of the meta-model is configured as follows:

- A Terminological Data Collection (TDC) contains any number of Terminological Entries (TEs).
- Each TE refers to a single concept which can therefore be represented in *n* languages in the Language Sections (LSs).
- For each language, there are *m* Term Sections (TSs) containing the terms that, in that language, describe the concept and all the related features.
- Each TS can contain any number of Term Component Section (TCS) providing information about parts of a term such as morphemes, phonemes, syllables, or single words from a multiword term.

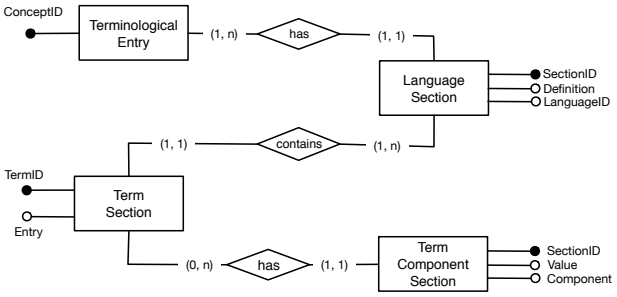


Figure 1: The Entity Relationship (ER) schema of the TMF meta-model

As shown in Figure 1, this standard proposes adopting the hub and spoke model (Van Campenhoudt, 2017) and distinguishes the hierarchical levels (Romary, 2001) to which different data categories can be associated: 1) the "conceptual" data common to all languages; 2) the language-specific data; 3) the term-specific data.

In TriMED, a single concept, which is exclusive for a terminological entry and defined by its identifier, can be expressed by *n* languages. Language sections have their own identifier and include the definition expressing the related concept, as well as the language code defined by the ISO standard 639.[16] Thereafter, for each language, all the terms (TSs) expressing such concept are associated with all their related terminological data (not shown in Figure 1 for space reasons). Finally, in case of a multi-words term, the properties of parts of the term are grouped in the TCS.

3.2. Terminological Data

The four main nodes of the meta-model can be accompanied by a large number of terminological data depending on the purpose of the resource. As described at the beginning of this section, we provide a wide range of information in the TriMED resource in oder to satisfy the information needs of i) patients, ii) translators, and iii) healthcare professionals. In Table 1, we show a list of all the terminological data provided in our model of record for the three working languages. Other administrative information specifically related to the record are provided, such as: author, updater, origination and modification date, user suggestion and notes.

In order to frame all the data categories associated to a term and the concept it designates, we have referred to the ISO standard 12620 about *Data Category Specifications* (ISO-12620, 2019). This international standard classifies all the categories of data that can be associated to the TMF meta-model levels: 1) the "conceptual" data common to all languages; 2) the language-specific data; 3) the term-specific data. For their description and implementation, the standard refers to a Data Category Repository (DCR) named DatCatInfo.[17] This resource replaces the previous ISO-

[16] https://www.iso.org/iso-639-language-codes.html

[17] http://www.datcatinfo.net/

Table 1: TriMED Terminological Data

Field	Terminological Data
Morphology	part of speech, grammatical gender, grammatical number, derivative forms
Phonetics	IPA transcription
Etimology	derivation, composition
Variation	orthographic variant, abbreviation, full form, acronym
Semantics	definition, semic analysis, (quasi-)synonym, hyponym, hyperonym
Phraseology	phraseological unit, collocation, colligation
Pragmatics	context of use
Register	common name, scientific name, international classification codes
Domain	subject field, subdomain
References	source (URLs)

cat Repository which was conceived precisely for the description of language resources metadata and for supporting interoperability and reusability of terminological data (Kemps-Snijders et al., 2008; Windhouwer, 2012; Broeder et al., 2014; Windhouwer and Schuurman, 2014). The new DatCatInfo repository collects a list of all the data categories and data category specifications that can be stored in a terminological data collection. A data category is a class of information that forms part of a data collection or annotation scheme for a given language resource (for example, /definition/,/part of speech/[18] etc.), whereas a data category specification provides the complete and formal representation of a data category (for example, its name, definition, examples, comments, etc.).

3.2.1. TriMED Data Category Repository

The novelty of this standard is the possibility for all implementors to create their own Data Category Repository (DCR) specific for a particular resource. In this way, the definition of a clear framework for specifying, managing, and using data categories will increase interoperability of terminology resources. According to this, we have designed the TriMED Data Category Repository which gathers all the data category specifications, that is the description of all the terminological data we provide in our model of terminological record. The Web application has been implemented with the Shiny R package (Chang, 2015) and it is available online.[19] To our knowledge, this is the first DCR implemented according to this standard and it collects 36 data category specifications.

The TriMED DCR allows us to disambiguate the meaning of some categories of data. The user who consults the resource, therefore, has all the information necessary to understand the data provided. In addition, the DCR provides guidelines for the correct compilation of the terminology records in terms of consistency of the data provided. This is because the set of values of each data category is defined by the DCR itself.

[18]For compliance with the standard the data categories are written between slashes.

[19]http://purl.org/trimed/dcr

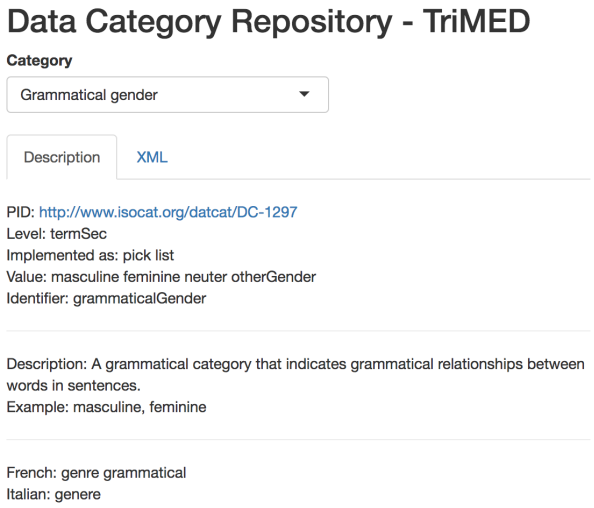


Figure 2: TriMED Data Category Repository

3.3. TBX Implementation

Once we have defined the structural meta-model of the terminological data collection and our Data Category Repository for the specifications of terminological data, we describe in this section the implementation format chosen for the TriMED resource. For its representation, we have referred to the ISO standard 30042 dealing with *TermBase eXchange* (TBX) format (ISO-30042, 2019). This document defines the TBX framework, expressed in the XML markup language, for the analysis, descriptive representation and dissemination of structured terminological data. Moreover, the TBXinfo.net [20] website provides many examples for those wishing to structure their data in TBX format. The main purpose of the TBX framework is to ensure that lexical data can be used in different software. For example, in the process of translation with CAT tools this format is the most commonly used for terminology management systems (Bowker and Fisher, 2010).

The structure of the TBX format consists of two interacting components: 1) a basic structure that reflects the TMF meta-model (see section 3.1.); and 2) a formalism aiming to define modules containing a list of data categories. The combination of these two components defines a particular "dialect", that is, an XML markup language conforming to TBX. Dialects may differ in terms of permitted data categories and meta-model levels where these categories may be entered. On the TBXinfo.net website, three public dialects are described and recommended for the exchange of terminological data: 1) TBX-Core, 2) TBX-Min, 3) TBX-Basic.[21] These dialects provide a restricted set of data categories, and for this reason the standard gives the opportunity to create "private dialects" for those who wish to represent terminology data categories that are not included in the *Core*, *Min*, *Basic* modules.

[20]https://www.tbxinfo.net

[21]https://www.tbxinfo.net/tbx-dialects/

3.3.1. TBX-TriMED dialect

The standard specifies that, for those data categories already documented, it is necessary to refer to public modules, while for the other categories it is necessary to create one's own module.

In order to frame all the TriMED data categories, we decided to create a new private dialect through the formal definition of the *Trimed* module which has be written in prose according to the TBX Module Description (TBXMD) formalism provided in the ISO standard. The TriMED dialect, therefore, includes the following modules: *Core*, *Min*, *Basic* as public modules, and the new private module *Trimed*[22] Here is an instance excerpt of the TriMED dialect:

```
<?xml version=''1.0'' encoding=''UTF−8''?>

<tbx xmlns:tbx3=''urn:iso:std:iso:30042:ed−2''
    xmlns:min=''http://www.tbxinfo.net/ns/min''
    xmlns:basic=''http://www.tbxinfo.net/ns/basic''
    xmlns:trimed=''http://www.trimed.org/ns/sample''
    type=''TBX−TriMED'' style=''dct'' xml:lang=''en''>
    <tbxHeader>
        <fileDesc>
            <sourceDesc>
                <p>A sample of the TriMED termbase
                consisting of one concept entry</p>
            </sourceDesc>
        </fileDesc>
    </tbxHeader>

    <text>
        <body>
            <conceptEntry id=''c1''>
                <basic:subjectField>medicine</basic:subjectField>

                <langSec xml:lang=''en''>
                    <descripGrp>
                        <basic:definition>The identification of diseases by the
                        examination of symptoms and signs and by other
                        investigations.</basic:definition>
                        <trimed:semicAnalysis>/identification//disease/
                        /examination//symptoms/</trimed:semicAnalysis>
                    </descripGrp>

                    <termSec>
                        <term>diagnosis</term>
                        <min:partOfSpeech>noun</min:partOfSpeech>
                        <basic:grammaticalGender>neuter
                        </basic:grammaticalGender>
                        <trimed:derivative>to diagnose, diagnostic,
                            prediagnosis, prediagnoses</trimed:derivative>
                    </termSec>
                </langSec>

                <langSec xml:lang=''fr''>
                    <descripGrp>
                    <basic:definition>Connaissance que l'on peut avoir d'une
                    maladie en observant les signes de celle−ci. </basic:definition>
                    <trimed:semicAnalysis>/diagnose//etude//maladie/
                    /symptome//medecine/</trimed:semicAnalysis>
                    </descripGrp>

                    <termSec>
                        <term>diagnose</term>
                        <min:partOfSpeech>nom</min:partOfSpeech>
                        <basic:grammaticalGender>feminin
                        </basic:grammaticalGender>
                        <trimed:derivative>diagnostic,
                        diagnostique, diagnostiquer</trimed:derivative>
                    </termSec>
                </langSec>

            </conceptEntry>
        </body>
    </text>
</tbx>
```

In the root element *tbx* the following information are specified: 1) the TBX-Core,[23] TBX-Min, TBX-Basic and TBX-TriMED namespaces; 2) the value of the type attribute, that is the name of the TBX dialect; 3) the style of the instance, that is the Data Category as Tag (DCT) format;[24] and 4) the working language (xml: lang) of the document. The structure follows the TMF hierarchical meta-model: in this order, i) *conceptEntry*, ii) *langSec*, and iii) *termSec* are defined. At the concept level, represented by its identifier (id=c1), the /subject field/ data category belonging to the *Basic* module is provided. Then, the same concept is expressed in two language sections containing the terms designating the concept in English and French. At the language level, we provide the information about the /definition/ and /semic analysis/ of the concept. These two data categories are extracted from two different modules, that is *Basic* and *Trimed*. Each *langSec* contains a *termSec* with the term designating the concept and other data categories, such as /part of speech/, /grammatical gender/ and /derivative forms/ belonging to the modules *min, basic, trimed*.

4. Web Application

The TriMED Web application has been implemented with the Shiny R package (Chang, 2015). It provides a different interface with a different subset of pieces of linguistic information depending on the user identification:

- Patients can look for the technical term and consult the corresponding popular term and its definition; [25]
- Physicians can consult the translation of each term and have access to other standard medical terminologies, ontologies and classification systems, such as MeSH terms and SNOMED concepts; [26]
- Scientific translators can examine a bilingual record with the medical terms in the source and target languages and all the relevant information in order to support the process of translation. [27]

In Figure 3, we show the main panels of the application: patient (Fig. 3a), physician (Fig. 3b), and translator (Fig. 3c)

4.1. Patients

A patient who is looking for the information about a popular term starts the search by selecting the language (English, French, or Italian) and then by typing the term in the search box (Fig. 3a). The system automatically filter words, character by character, and shows the possible alternatives in the box. After a careful analysis, we chose to display during the search both the popular and the technical term in order to give an immediate feedback to the user. When the user selects the term, the definition and the information about terminological variation are displayed in the page.

[22] The TriMED dialect is currently under revision. A partial dialect definition package is available at: `https://github.com/trimed-dialect/TriMED.git`

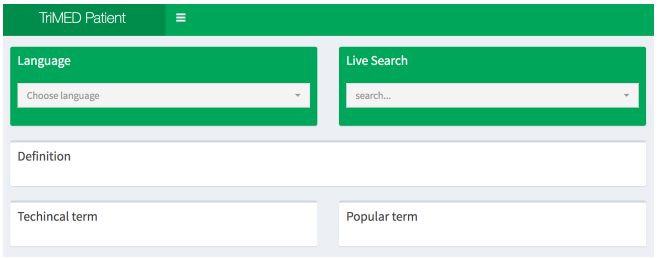
[23] That is *urn:iso:std:iso:30042:ed-2* which has to be used as the default namespace for TBX document instances of all dialects.

[24] There are two XML isomorphic styles that can be used to represent terminological data: Data Category as Attribute (DCA) and Data Category as Tag (DCT); for more information, see section 6 of the ISO standard or `https://www.tbxinfo.net/dca-v-dct/`

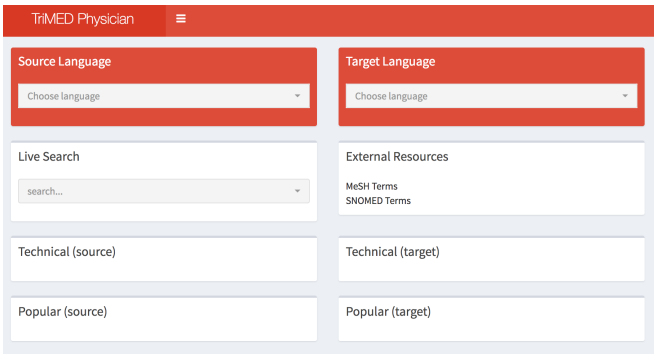
[25] `http://purl.org/trimed/patient`

[26] `http://purl.org/trimed/physician`

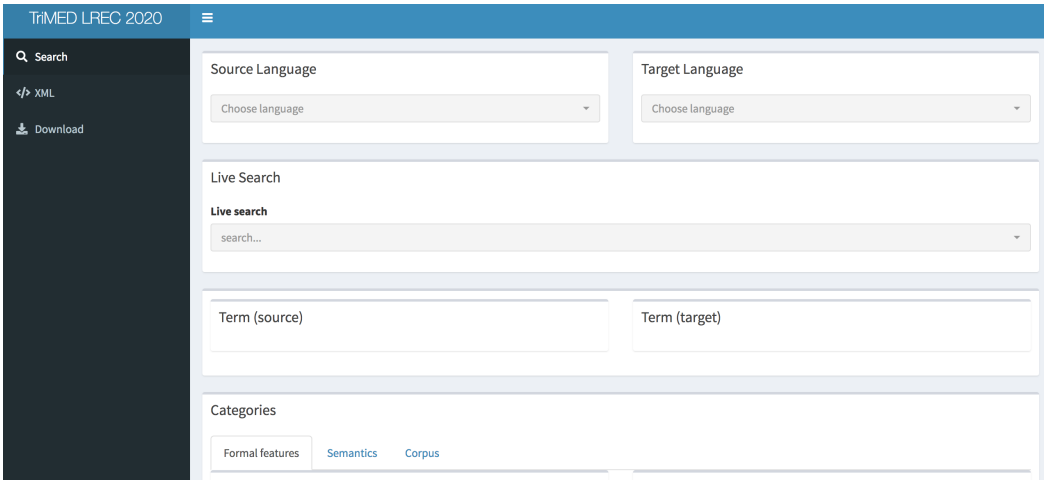
[27] `http://purl.org/trimed/translator`



(a) Patient interface



(b) Physician interface



(c) Translator interface

Figure 3: The interfaces of the TriMED Web application.

4.2. Physicians

The interface for the physician allows the doctor to access a bilingual terminological record (Fig. 3b). First, the user chooses the source (and optionally the target language), then it is possible to type the term for which a translation is needed, in both technical and popular register. Once the term is selected, the interface shows the definition of the term (and the translation of the term if the target language has been selected) and the link to the medical ontologies MESH and SNOMED.

4.3. Translators

A scientific translator can access all the linguistic features of the multilingual version of every record in TriMED and download a standard TBX record that can be used in any Computer Assisted Translation software that supports this standard. This interface has a left sidebar (Fig. 3c) that allows a translator to choose among three visualization options: Search, XML, and Download. Initially, the translator must choose at least a source language in the Search visualization to activate the search box. The search process for the source language is identical to the physician interface. Once the translator selects a term, the interface shows the linguistic features grouped in three tabs: Formal Features, Semantics, and Corpus. If the user selects a target language, the XML and the Download visualization areas will display 1) the TBX version of the multilingual record and 2) a Download button to download the TBX record.

5. Conclusions

In this paper, we presented a general methodology for the formal standardization of terminological resources according to the three most recent standards provided by the ISO TC/37, sub-committee 3, dealing with the management of terminology resources. We applied this methodology on the TriMED multilingual termbase and, in particular, we focused on 1) the definition of the resource structure according to Terminological Markup Framework (TMF) meta-model; 2) the description of its data categories specifications through the implementation of the TriMED Data Category Repository, and to our knowledge this is the first implementation of a DCR compliant with the new ISO standard; 3) the implementation of the resource in the TermBase eXchange (TBX) format; and 4) the realization of the Web application to access the data stored in the termbase. This work of re-modeling the TriMED resource was guided by the need for FAIR (Findability, Accessibility, Interoperability and Reusability) data declared by the European Open Science Cloud (ESOC).

The TriMED terminological database currently contains a total of 1188 multilingual terminological records. The work on the compilation of the records has been running since 2018 and, at present time, we can give the following number of completed standardized records per language: 263

for English, 132 for French, and 347 for Italian. Since this is an ongoing project, the number of completed records per language will increase in the next weeks. As future works, we are working on the full specification of the TriMED dialect definition package. Moreover, we are planning to include additional medical data from international classification systems, such as ICP2 and ICD11 codes. Finally, we are thinking about the design of a Web application where external users can contribute to the compilation of the missing terminological records and add new ones.

6. Acknowledgements

This work was partially supported by the ExaMode Project, as a part of the European Union Horizon 2020 Program under Grant 825292.

7. Bibliographical References

Balboni, P. E. (2000). Le microlingue scientifico-professionali. *Torino: Utet*.

Bowker, L. and Fisher, D. (2010). Computer-aided translation. *Handbook of translation studies*, 1:60–65.

Broeder, D., Schuurman, I., and Windhouwer, M. (2014). Experiences with the ISOcat data category registry. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 4565–4568, Reykjavik, Iceland, May. European Language Resources Association (ELRA).

Cardillo, E., Serafini, L., and Tamilin, A. (2009). A lexical-ontological resource for consumer healthcare. In *International Conference on Electronic Healthcare*, pages 131–138. Springer.

Castro, C. M., Wilson, C., Wang, F., and Schillinger, D. (2007). Babel babble: physicians' use of unclarified medical jargon with patients. *American journal of health behavior*, 31(1):S85–S95.

Chang, W., (2015). *Shiny: Web Application Framework for R*. R package version 0.11.

Deuster, L., Christopher, S., Donovan, J., and Farrell, M. (2008). A method to quantify residents' jargon use during counseling of standardized patients about cancer screening. *Journal of general internal medicine*, 23(12):1947.

Elhadad, N. and Sutaria, K. (2007). Mining a lexicon of technical terms and lay equivalents. In Proc BioNLP Workshop, pages 49-56. ACL.

Faber, P. (2015). Frames as a framework for terminology. *Handbook of terminology*, 1(14):14–33.

Galinski, C. and Nedobity, W. (1988). Special languages, terminology planning and standardization. In *Standardization of Technical Terminology: Principles and Practices (Second Volume)*. ASTM International.

Hadden, K., Coleman, C., and Scott, A. (2018). The bilingual physician: Seamless switching from medicalese to plain language. *Journal of Graduate Medical Education*, 10(2):130–133.

ISO-12620. (2019). Management of terminology resources – Data category specifications. Standard, International Organization for Standardization, Geneva, CH, March.

ISO-16642. (2017). Computer applications in terminology – Terminological markup framework. Standard, International Organization for Standardization, Geneva, CH, November.

ISO-171171. (2018). Health informatics – Terminological resources – Part 1: Characteristics. Standard, International Organization for Standardization, Geneva, CH, April.

ISO-30042. (2019). Management of terminology resources – TermBase eXchange (TBX). Standard, International Organization for Standardization, Geneva, CH, April.

Jammal, A. (1999). Une méthodologie de la traduction médicale. *Meta: journal des traducteurs/Meta: Translators' Journal*, 44(2):217–237.

Jucks, R. and Bromme, R. (2007). Choice of words in doctor-patient communication: an analysis of health-related internet sites. Health Commun, 21(3):267-77.

Kemps-Snijders, M., Windhouwer, M., Wittenburg, P., and Wright, S. E. (2008). ISOcat: Corralling data categories in the wild. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, May. European Language Resources Association (ELRA).

Khayari, M., Schneider, S., Kramer, I., Romary, L., et al. (2006). Unification of multi-lingual scientific terminological resources using the iso 16642 standard. the termsciences initiative. *arXiv preprint cs/0604027*.

Magris, M., Musacchio, M. T., Rega, L., and Scarpa, F. (2002). *Manuale di terminologia: aspetti teorici, metodologici e applicativi*. Hoepli.

McCray, A. (2005). Promoting health literacy. J of Am Med Infor Ass, 12:152-163.

Müller, K. and Wickham, H., (2019). *tibble: Simple Data Frames*. R package version 2.1.3.

Offenga, F., Broeder, D., Wittenburg, P., Ducret, J., and Romary, L. (2006). Metadata profile in the ISO data category registry. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy, May. European Language Resources Association (ELRA).

Picht, H. and Schmitz, K.-D. (2001). *Terminoligie und Wissensordnung: Ausgewählte Schriften aus dem Gesamtwerk von Eugen Wüster*. TermNet publisher.

Romary, L. (2001). An abstract model for the representation of multilingual terminological data: Tmf-terminological markup framework. In *TAMA 2001*.

Rouleau, M. (2003). La terminologie médicale et ses problèmes. Tribuna, Vol. IV, n. 12.

Roumier, J., Vander Stichele, R., Romary, L., and Cardillo, E. (2011). Approach to the creation of a multilingual, medical interface terminology. In *Workshop Proceedings of the 9th International Conference on Terminology and Artificial Intelligence*, pages 13–15. INALCO.

Sager, J. C. (1998). Terminology: standardization. *& "Terminology: theory". In Mona Baker (Ed.) Routledge Encyclopedia of Translation Studies. GB: Routledge*.

Sánchez, M. T. and Velasco, J. A. P. (2013). Las barreras en la comunicación médico-paciente: el proyecto

varimed. In *Translating culture*, pages 593–606. Editorial Comares.

Tran, T., Chekroud, H., Thiery, P., and Julienne, A. (2009). Internet et soins : un tiers invisible dans la relation médecine/patient. Health Commun, 21(3):267-77.

Van Campenhoudt, M. (2017). Standardised modelling and interchange of lexical data in specialised language. *Revue française de linguistique appliquée*, 22(1):41–60.

Vezzani, F., Nunzio, G. M. D., and Henrot, G. (2018). Trimed: A multilingual terminological database. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation, LREC 2018, Miyazaki, Japan, May 7-12, 2018.*

Warnier, M., Roumier, J., Jamoulle, M., Cardillo, E., Vander Stichele, R., and Romary, L. (2012). Publishing a multilingual medical terminology according to terminology standards and linked data principles. In *SABRE Conference 2012*.

Wermuth, C. and Verplaetse, H. (2019). Medical terminology in the western world. current situation. In *Handbook of Terminology, Volume 2 [HoT2]: Terminology in the Arab world*, volume 2, pages 84–108. John Benjamins; Amsterdam.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., et al. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific data*, 3.

Windhouwer, M. and Schuurman, I. (2014). Linguistic resources and cats: how to use ISOcat, RELcat and SCHEMAcat. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3806–3810, Reykjavik, Iceland, May. European Language Resources Association (ELRA).

Windhouwer, M. (2012). RELcat: a relation registry for ISOcat data categories. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 3661–3664, Istanbul, Turkey, May. European Language Resources Association (ELRA).

Wüster, E. (1971). Principles of special language standardization. *Muttersprache*, 81(5):289–295.

Zeng, Q. T. and Tse, T. (2006). Exploring and developing consumer health vocabularies. *Journal of the American Medical Informatics Association*, 13(1):24–29.