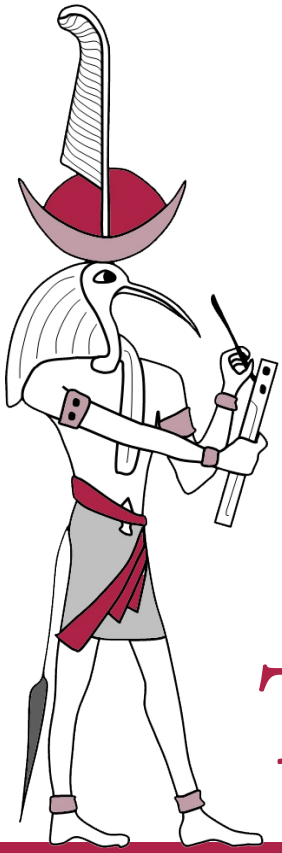
Terminologica



TOTh 2019

Terminologie & Ontologie : Théories et Applications

Presses Universitaires Savoie Mont Blanc

Terminologie & Ontologie : Théories et Applications

Actes de la conférence

TOTh 2019

Le Bourget du Lac – 6 & 7 juin 2019

Les ouvrages TOTh précédents sont disponibles sur le site du Comptoir des Presses d'Universités (www.lcdpu.fr) ou auprès de : contact@toth.condillac.org

Éditeur : Presses Universitaires Savoie Mont Blanc
27 rue Marcoz
BP 1104
73011 CHAMBÉRY CEDEX
www.univ-smb.fr

Réalisation : C. Brun, C. Roche
Collection « Terminologica »
ISBN : 978-2-919732-80-7
ISSN : 2607-5008
Dépôt légal : juillet 2020

Terminologie & Ontologie : Théories et Applications

Actes de la conférence

TOTh 2019

Le Bourget du Lac – 6 & 7 juin 2019

http://toth.condillac.org

avec le soutien de :
Université Savoie Mont Blanc
École d'ingénieurs Polytech Annecy Chambéry

Presses Universitaires Savoie Mont Blanc
Collection « Terminologica »

Comité scientifique

Président du Comité scientifique : Christophe Roche

Comité de pilotage

Rute Costa	Universidade Nova de Lisboa
Humbley John	Université Paris 7
Kockaert Hendrik	University of Leuven
Christophe Roche	Université Savoie Mont Blanc

Comité de programme 2019

Le comité de programme est constitué chaque année à partir du comité scientifique de TOTh en fonction des soumissions reçues. La composition du comité scientifique est accessible à l'adresse suivante : http://toth.condillac.org/committees

Guadelupe Aguado	Universidad Politécnica de Madrid – Spain
Amparo Alcina	Universitat Jaume I – Spain
Bruno Bachimont	Université Technologie de Compiègne – France
Jean-Paul Barthès	Université Technologie de Compiègne – France
Christopher Brewster	TNO – The Netherlands
Danielle Candel	CNRS, Université Paris Diderot – France
Sylviane Cardey	Université de Franche-Comté – France
Stéphane Chaudiron	Université de Lille 3 – France
Manuel Célio Conceição	Universidade do Algarve – Portugal
Rute Costa	Universidade NOVA de Lisboa – Portugal
Bruno Courbon	Université Laval – Canada
Lyne Da Sylva	Université de Montréal – Canada
Luc Damas	Université Savoie Mont-Blanc – France
Éric De La Clergery	INRIA – France
Dardo De Vecchi	Kedge Business School – France
Valérie Delavigne	Université Paris 3 – France
Sylvie Després	Université Paris 13 – France
Juan Carlos Diaz Vasquez	EAFIT University – Colombia
Hanne Erdman Thomsen	Copenhagen Business School – Denmark
Pamela Faber	Universidad de Granada – Spain
Christiane Fellbaum	Princeton University – USA
Iolanda Galanes	Universidade de Vigo – Spain
Christian Galinski	INFOTERM – Austria
François Gaudin	Université de Rouen – France
Teodora Ghiviriga	Alexandru Ioan Cuza University – Romania
Jean-Yves Gresser	ancien Directeur à la Banque de France – France
Ollivier Haemmerlé	Université de Toulouse – France
Gernot Hebenstreit	University of Graz – Austria
Amanda Hicks	University of Florida – USA
John Humbley	Université Paris 7 – France

Avant-propos

La Terminologie est une discipline scientifique à part entière qui puise à de nombreux domaines dont la linguistique, la théorie de la connaissance et la logique. Pour que cette diversité soit une richesse, il faut lui offrir un cadre approprié au sein duquel elle puisse s'exprimer et s'épanouir : c'est une des raisons d'être des Conférences TOTh créées en 2007. A ces conférences « mères » qui se tiennent chaque année à l'Université Savoie Mont-Blanc sont associées depuis 2011 les Journées d'étude TOTh dédiées à un thème plus spécifique organisées par une institution partenaire.

Dans ce contexte, la formation et la transmission des connaissances jouent un rôle essentiel. La *Formation TOTh* précédant la Conférence se déroule sur deux années consécutives dédiées pour l'une à la dimension linguistique et pour l'autre à la dimension conceptuelle de la terminologie, deux dimensions étroitement liées.

A la présentation de travaux sélectionnés par un Comité de programme international, la *Conférence TOTh* inclut une *Conférence invitée* et, selon les années, une *Disputatio*. La première, donnée par une personnalité reconnue dans son domaine vise l'ouverture à d'autres approches de la langue et de la connaissance. La seconde, à travers une lecture commentée effectuée par un membre du comité scientifique, renoue avec une forme d'enseignement et de recherche héritée de la scolastique.

Christian Galinski de Infoterm, a ouvert la conférence sur le sujet de « *The emergence of terminology science and terminological activities* ».

Cette année, comme en 2018, nous n'avons pas inclus de *Disputatio* par manque de temps. En effet, pour la première fois, TOTh a accueilli une session satellite, en parallèle avec la conférence, sur le thème de « Terminology and Text Mining » en lien direct avec les thèmes de TOTh. Nous avons également dédié une session de la conférence au projet Européen ELEXIS.

Les 29 communications et les 3 posters ont permis d'aborder de nombreux sujets tant théoriques que pratiques, autant d'exemples de la diversité et de la richesse de notre discipline. Je vous invite à découvrir à travers ces actes les 24 interventions qui ont donné lieu à publication.

Avant de vous souhaiter bonne lecture, j'aimerais terminer en remerciant tous les participants pour la richesse des débats et des moments partagés.

Christophe Roche
Président du comité scientifique

Sommaire

Using an Infrastructure for Lexicography in the Field of Terminology

Tanja Wissik*, Thierry Declerck*/**

*Austrian Academy of Sciences
Austrian Centre for Digital Humanities and Cultural Heritage
tanja.wissik@oeaw.ac.at
https://www.oeaw.ac.at/acdh/team/current-team/tanja-wissik
**DFKI GmbH
Multilinguality and Language Technology Lab
Stuhlsatzenhausweg, 3
D-66123 Saarbrücken
Germany
declerck@dfki.de
https://www.dfki.de/~declerck/

Abstract. In this contribution, we discuss the (re-)use of the ELEXIS research infrastructure for lexicography in order to deal with terminological data. We present central aspects of the ELEXIS infrastructure and the standards it both applies and further develops. We also present TBX, which is the main standard used for representing terminological data. We describe in some detail the OntoLex-Lemon specifications, which which result from a W3C Community Group and which play a central role in our work consisting in describing terminological data within an infrastructure for lexicography, as it supports linking knowledge organisation systems to a full lexical description. To exemplify this capability, we use multilingual terminology data, originally encoded in TBX, from the field of risk management.

1. Introduction

In many disciplines e-Research and the use of digital methods has become an omni-present research practice (Lusicky and Wissik (2017)). In general, research infrastructures enable e-Research by providing facilities, resources or services of a unique nature, to conduct and to support top-level research

activities in different domains. They include for example major scientific equipment or sets of instruments, knowledge-based resources like collections, archives and scientific data as well as e-Infrastructures, such as data and computing systems and communication networks, and any other tools that are essential to achieve excellence in research and innovation.[1]

2. Research Infrastructures

Research infrastructures (RIs) offer technical and social infrastructures in a more stable and sustainable way than research projects that run only for two to four years. As technical infrastructure they provide resources, tools and services to the scientific community in order to support top-level research activities. As social infrastructure, RIs provide platforms for collaborative research and knowledge transfer and promote the use of common methods and standards. They also play an important role in educating new generations of researchers.

The European Strategy Forum on Research Infrastructures (ESFRI) recognises over fifty different research infrastructures (ESFRI (2018)). While some infastructures are domain specific and others generic, ideally these domain specific and generic research infrastructures complement each other (Illmayer (2017)).

2.1. Generic Research Infrastructures

As Generic Research Infrastructures, for this paper, we understand Research Infrastructures that can be used by researchers from a variety of research fields. In the Humanities, e.g. CLARIN (Common Language Resources and Technology Infrastructure) and DARIAH (Digital Research Infrastructure for the Arts and Humanities) count as generic research infrastructures (Doel and Maes (2012)).

2.2. Domain Specific Research Infrastructures

Besides generic research infrastructures, there are domain specific research infrastructures. In the Humanities, e.g. EHRI (European Holocaust Research Infrastructure) or ARIADNE (Advanced Research Infrastructure

1 ESFRI Roadmap 2018, Part 1, p 11

for Archaeological Dataset Networking in Europe) can be seen as domain specific research infrastructure.

3. European Lexicographic Infrastructure (ELEXIS)

As an example of a domain specific research infrastructure we introduce the European Lexicographic Infrastructure (ELEXIS). ELEXIS is a research infrastructure project under H2020 aiming to develop a Research Infrastructure for Lexicography that provides online access to data, tools and services for lexicography research (cf. Declerck *et al*. 2018). In the following we describe the whole infrastructure, also the parts that are still in development. It consists of three sub-infrastructures LEX1, LEX2, LEX3 (Krek *et al*. (2018)) that are explained below in more detail.

The first part of the infrastructure (LEX1) includes conversion and alignment tools in order to harmonise and convert lexicographic resources into a uniform data format that allows their integration in the Linked Open Data (LOD) cloud[2], and more specifically in the Linguistic Linked Open Data (LLOD) cloud[3]. The LLOD was originally an initiative by members of the Open Knowledge Foundation that has gained a lot of attraction, and which was further developed in various projects, as described in (McCrae *et al*. (2016)). The second part of the infrastructure (LEX2) includes word sense disambiguation and entity linking tools dedicated to semantic processing of corpus data. These tools facilitate disambiguation and corpus analysis and open up the possibility to create lexicographic resources automatically from corpora. The third part of the infrastructure (LEX3) includes tools to support the retro-digitising process of dictionaries (Krek *et al*. (2018)).

As common data formats, the ELEXIS infrastructure makes use of the OntoLex-Lemon (McCrae *et al*. (2017)) and TEI Lex-0 models (Banski *et al*. (2017)) as outlined in (Ahmadi *et al*. (2019)) or in (McCrae *et al*. (2019)).

There are different ways, how the ELEXIS Infrastructure can be used for already existing dictionary data. For dictionaries that are not already in a digital format the retro-digitization tool in LEX3 is used. This applies OCR to the text and then processes it by adding XML markup in the format of TEI Lex-0. For dictionaries that are already available in a digital form, but not one that is supported directly by the project, the conversion tool of LEX1 is

2 See https://www.lod-cloud.net/ for more details.
3 See https://linguistic-lod.org for more details.

used to convert these resources to TEI Lex-0. If the dictionaries are already in TEI Lex-0 or have been converted to TEI Lex-0 by one of the two methods described above, they can be consumed directly by the interoperable REST interface. Furthermore, if the dictionaries are already available in Ontolex-Lemon, they can also be consumed directly be the REST interface. However, there are no conversion tools foreseen in the infrastructure to convert other formats to Ontolex-Lemon, for the time being it is handled outside of the infrastructure. Most of this mapping work onto OntoLex-Lemon is realised by a partner Research and Innovation project, Prêt-à-LLOD[4]. One of its tasks is to transform various types of linguistic data into an LLOD compliant format.

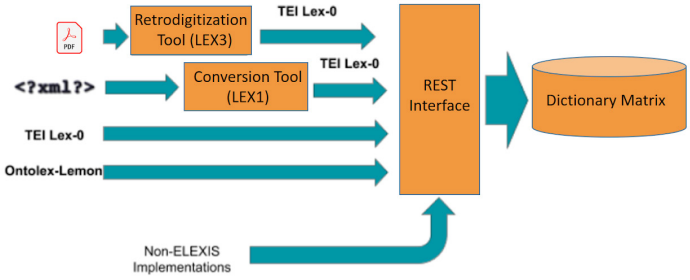


Fig. 1 – *Used formats and standards in order to access the ELEXIS Infrastructure via the REST Interface (Ahmadi et al., 2019, modified).*

All data, copyrights permitting, will be made available as part of the Linguistic Linked Open Data cloud.

In the next section we discuss in detail the different standards used for terminological and lexicographical resources, especially those relevant for the ELEXIS infrastructure and our use case.

4. Standards

One aspect of a sustainable infrastructure and the (re)-usability of resources, services and tools are common standards and data formats. In the

4 See https://www.pret-a-llod.eu/ for more details.

following section we describe standards that are relevant in the fields of lexicography and terminology.

4.1. Standards for Lexicography

In terms of used standards and formats, the field of lexicography is very heterogeneous. In the ELEXIS project, a survey of user needs was carried out (Kallas *et al*. (2019)). This survey had two parts, one for lexicographers and one for institutions. In the survey, there were also questions regarding used standards and formats. The survey has shown that many lexicographic projects use XML or databases and some RDF based formats, but there are still projects working with unstructured data and text format. Among those using XML, custom XML, TEI (P2 or P5) and TEI Lex-0 were mentioned (Kallas *et al*. (2019)). According to the survey data, two tendencies were observed: "a) a transition from non-structured data or text format to structured data format; b) still insufficient use of (standardised) structured formats enabling reliable re-use and linking of dictionary data" (Kallas (2019, 55)). In this respect, ELEXIS, as a promoter of common standards, plays a crucial role. In the following we describe some of the standards mentioned in the ELEXIS survey of user needs.

4.1.1. TEI Dictionary Chapter

In the TEI P5 Guidelines, the Dictionary Chapter defines how to encode lexical resources of all kinds, in particular human-oriented monolingual and multilingual dictionaries, glossaries, and similar documents (TEI P5 2019).

4.1.2. TEI Lex-0

The TEI Dictionary Chapter is very complex and allows many ways to encode a dictionary entry. TEI Lex-0 is simpler and aims at establishing a baseline encoding and a target format to facilitate the interoperability of heterogeneously encoded lexical resources (DARIAH WG Lexical Resources (2019)). TEI Lex-0 should not be thought of as a replacement of the Dictionary Chapter in the TEI Guidelines or as the format that must be used for editing or managing individual resources, especially in those projects and/or institutions that already have established workflows based on their own flavors of TEI. TEI Lex-0 should be primarily seen as a format that existing TEI dictionaries can be univocally transformed to in order to be queried, visualised, or mined in a uniform way (Romary (2015), DARIAH WG Lexical Resources (2019)).

4.1.3. OntoLex-Lemon

The OntoLex-Lemon model was originally developed with the aim to provide a rich linguistic grounding for ontologies, meaning that the natural language expressions used in the labels, definitions or comments of ontology elements are equipped with an extensive linguistic description.[5] This rich linguistic grounding includes the representation of morphological and syntactic properties of lexical entries as well as the syntax-semantics interface, i.e. the meaning of these lexical entries with respect to an ontology or to specialised vocabularies.

The main organising unit for those linguistic descriptions is the lexical entry, which enables the representation of morphological patterns for each entry (a multi word expression, a word or an affix). The connection of a lexical entry to an ontological entity is marked mainly by the *ontolex:denotes* property or is mediated by the *LexicalSense* or the *LexicalConcept* classes, as this is represented in Figure 2, which displays the core module of the model. OntoLex-Lemon builds on and extends the lemon model (Cimiano *et al*. (2016)). A major difference is that OntoLex-Lemon includes an explicit way to encode conceptual hierarchies, using the SKOS[6] standard. As can be seen in Figure 2, lexical entries can be linked, via the *ontolex:evokes* property, to such SKOS concepts, which can represent WordNet synsets. This structure is paralleling the relation between lexical entries and ontological resources, which is implemented either directly by the *ontolex:reference* property or mediated by the instances of the *ontolex:LexicalSense* class.

5 See (McCrae *et al*., 2012), (Cimiano *et al*., 2016) and also https://www.w3.org/2016/05/ontolex/

6 SKOS stands for "Simple Knowledge Organization System". SKOS provides "a model for expressing the basic structure and content of concept schemes such as thesauri, classification schemes, subject heading lists, taxonomies, folksonomies, and other similar types of controlled vocabulary" (https://www.w3.org/TR/skos-primer/).

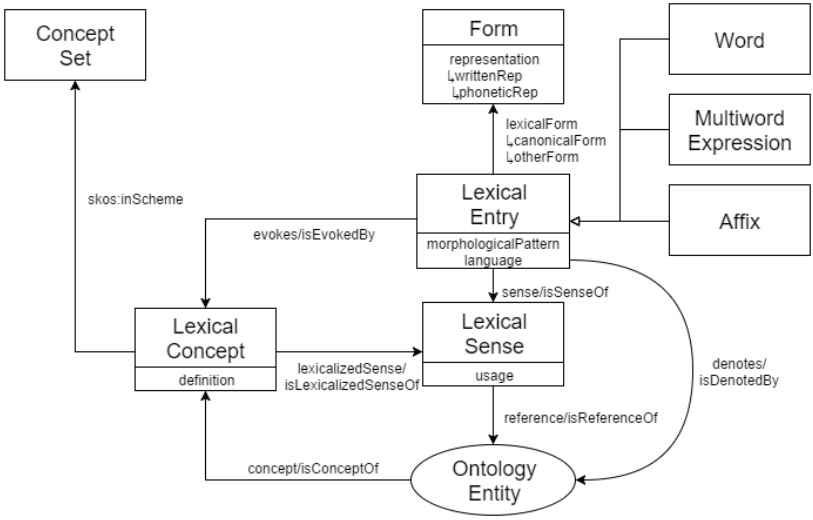


Fig. 2 – *The core module of OntoLex-Lemon : Ontology Lexicon Interface. Graphic taken from https ://www.w3.org/2016/05/ontolex/.*

More recent developments of the model have been described in (McCrae *et al*. (2017)). Currently three extension modules are being discussed : a lexicographic, a morphology and a corpus frequency module[7].

4.2. Standards for Terminology

Most standards in the field of terminology are language technology related standards. One of these standards is TBX, or TermBase eXchange, the other is TBX basic, a reduced version of TBX.

4.2.1. TBX

TBX is the international standard for representing and exchanging information on terminological data. It defines a family of formats that share a common structure and a limited range of information types. Each member of the family is called a dialect of TBX. The main purpose of TBX is to ensure "the independence of valuable terminological data from any particular software

7 See respectively https ://www.w3.org/2019/09/lexicog/, https ://www.w3.org/community/ontolex/wiki/Morphology and https ://github.com/acoli-repo/ontolex-frac

application used to access, display, update or otherwise process it" (Melby 2015, 393).

TBX is modular in order to support the varying types of terminological data, or data-categories, that are included in different terminological databases (termbases). TBX includes two modules : a core structure, and a formalism for identifying a set of data-categories and their constraints, both expressed in XML. The term TBX, when used alone, refers to the framework consisting of these two interacting modules.

The TBX entry model is organised in strict compliance to ISO 16642 (TMF) in that each of three elements : <termEntry>, <langSet> and <tig> in TBX respectively correspond to the three levels in TMF (entry level, language level and term level) (cf. Romary 2014). The data categories associated to these three levels are made of a) specific elements such as <term>, <note>, <ref> and <xref> and b) so called meta data-elements that may express a wide range of possible data categories, namely <admin>, <descrip> and <termNote>. For instance, <descrip type= "definition"> is how a definition is represented in TBX. (Romary (2014)).

4.2.2. TBX basic

TBX basic, as already mentioned, is a simpler and reduced version of TBX. It is also XML based as TBX and adheres to the same entry structure as described above (entry level, language level and term level) but only allows a limited set of data categories. Its purpose is to formalise the markup that is used in relatively simple terminology resources, in order to ensure interoperability (Terminorgs (2014)).

5. Use Case

5.1. Resource Description

The data we use for this use case is a data export of a multilingual terminology database on risk management provided by the University of Vienna available in the ELRC-SHARE repository[8]. The resource is in TBX-Basic format containing 1024 terms in the field of risk management in 5 languages : French, English, German, Romanian and Spanish. The termbase was created

8 https ://www.elrc-share.eu/repository

in order to improve domain communication and to facilitate mutual understanding across linguistic boundaries. The intended target users of this terminology resource were risk managers, civil engineers as well as teachers, students and translators (cf. Budin (2011, 23)).

The term entries contain terms, definitions (sometimes even more than one definition is provided) and context information, but for example, no grammatical information like part of speech is available as a separate data category. The resource contains single word expressions such as "risk" or "riesgo" as well as multiword expressions such as "risk reduction" or "reducción del riesgo" or "disaster risk reduction" or "reducción del riesgo de desastre".

5.2. TBX to Ontolex Lemon Transformation

So far there have been no mapping efforts to map common formats in terminology such as TBX to TEI Lex-0. However, there have been discussions on how to provide a representation of onomasiological data, such as terminological data in TEI in addition to the already existing "dictionaries" chapter (Romary (2014)). Furthermore, there have been initiatives to describe terminological data in RDF based representations (Cimiano *et al*. (2015), Rodriguez-Doncel *et al*. (2018)). From there it is possible to convert already existing terminological data to OntoLex-Lemon, which is the core of the LEX1 component of ELEXIS. As mentioned earlier, OntoLex-Lemon provides for a declarative interface between knowledge systems represented in SKOS and lexical data represented in OntoLex. In this, one can easily combine terminologies and lexicographical data and make them interoperable.

For exemplifying our approach, we display below some code from the OntoLex-Lemon encoding, in a simplified form, of one term taken from the Risk Management Terminology. The term is in the original TBX "risk". While the original terminology is repeating the term by each covered language ("risk", "risque", "Risiko", "Risc", "riesgo", for EN, FR, DE, RO and ES), in our conversion we have only one SKOS concept. We also note that in the original terminology the terms in the various languages are considering some typographical rules (for example capital letter for German nouns). We advocate for a language independent and neutral encoding of the terms of a termbase. This way, we implement a modular approach in which the terms are organised independently of the language data they use, but to which they are inter-linked by declarative property relations.

For reason of space and of simplicity we give just an example of the term "risk" below, but the OntoLex-Lemon model can also deal with terms involving multy-word expressions.

```
risk :ConceptSet_1
a ontolex :ConceptSet ;
rdfs :label "Risk Management Terminology from the University of
Vienna"@en.
```

The first code example displays the introduction of an instance of a skos :ConceptScheme (see Fig. 2 as orientation for all displayed code examples)

```
risk :LexicalConcept_1
a ontolex :LexicalConcept ;
rdfs :label „risk"@en ;
rdfs :label „riesgo"@es ;
rdfs :label „risque"@fr ;
rdfs :label „Risl"@de ;
skos :definition "Probabilidad de que un evento ocurra. Cálculo
matemático de pérdidas (de vidas, personas heridas, propiedad dañada y
actividad económica detenida) durante un periodo de referencia en una
región dada para un peligro en particular. Riesgo es el producto de la
amenaza y la vulnerabilidad."@es ;
skos :topConceptOf risk :ConceptSet_1 ;
ontolex :isConceptOf <https ://www.wikidata.org/wiki/Q104493> ;
ontolex :isEvokedBy risk :Word_1 .
```

Now we introduce the term that will cover all the language variations associated with it (so that we do not duplicate anymore the number of terms for one concept). In the code example before, we added just one definition, the Spanish one. The instance of *LexicalConcept* is related to a *LexicalEntry* by the property *isEvokedBy* and to an ontology entry in Wkidata by the property *isConceptOf*. Lexical data is thus no longer encoded with the conceptual space but linked to it.

The following code example is displaying the basic lexical information we are associated with the term "LexicalConcept_1" (to which it relates with the property "evokes". We specify there not only the language of the entry but also its part-of-speech and its gender, information which is not included in the original terminology.

```
risk :Word_1
a ontolex :Word ;
```

```
dc :language “http://id.loc.gov/vocabulary/iso639-2/spa” ;
lexinfo :gender lexinfo :masculine ;
lexinfo :partOfSpeech lexinfo :noun ;
rdfs :label “riesgo”@es ;
ontolex :canonicalForm risk :Form_1 ;
ontolex :evokes risk :LexicalConcept_1 .
```

The next two code examples are encoding the morphological variants of the lexical entry. This is an important information as some terms are restricting the usage of some words to be either singular or plural. The above mentioned OntoLex-Lemon module “lexicog” is specifying the way such usage restrictions can be encoded.

```
risk :Form_1
a ontolex :Form ;
lexinfo :number lexinfo :singular ;
ontolex :writtenRep “riesgo[@es}” .
risk :Form_2
a ontolex :Form ;
ontolex :representation “riesgos”@es .
```

These code examples show only a simple term, consisting of only a unique word. But the OntoLex-Lemon supports also the linking of terms consisting of multi word expressions to a full description of the lexical units of the term.[9] We note that the mapping from TBX to OntoLex-Lemon is not just realising a format conversion, but it is also leading to a simplification of the the original terminological data, as it does not need to include lexical information any more but can link to a specialised lexical data set. Another important aspect is the fact that we are now able to express lexical restriction on the words used in the terms.

6. Conclusion

In this paper, we outlined the possibility to use a domain specific infrastructure for another domain, for which it was not initially created, and we identified the key points for the usage. As use case we have chosen a research infrastructure for lexicography, showing how it can be used in the field of terminology. In the described use case, the key points to be able to use the

9 (Tiberius and Declerck (2017)) describe how Dutch compounds can be represented in their constituent parts. See also for the relevant sections of https ://www.w3.org/2016/05/ontolex/.

infrastructure are common formats and standards or the possibility of mapping and converting them, and supporting interoperability between different types of language data : terminological and lexical data.

Acknowledgements. Contributions by the Austrian Centre for Digital Humanities and Cultural Heritage at the Austrian Academy of Sciences were supported in part by the H2020 project "ELEXIS" with Grant Agreement number 731015. Contributions by the German Research Center for Artificial Intelligence (DFKI GmbH) were supported in part by the H2020 project Prêt-à-LLOD with Grant Agreement number 825182.

References

Ahmadi, Sina, Arcan, Mihael, Declerck, Thierry, Kernerman, Ilan, Khan, Fahad, Krek, Simon, McCrae, John, Měchura, Michal, Monachini, Monica, Roche, Christophe, Tiberius, Carole, Troelsgård, Thomas, Zaytseva, Ksenia. 2019. D2.1.Interface for Interoperable Lexical Resources. Accessed September 19. https ://elex.is/wp-content/uploads/2019/02/ELEXIS_D2_1_Interface_for_Interoperable_Resources.pdf

Bański, Piotr, Bowers, Jack, Erjavec,Tomaz. TEI-Lex0 guidelines for the encoding of dictionary information on written and spoken forms. Electronic Lexicography in the 21st Century : Proceedings of eLex 2017 Conference, Sep 2017, Leiden, Netherlands. <hal-01757108>

Budin, Gerhard. 2011. Designing and Implementing Strategies of Global, Multilingual "Disaster Communication". In Alekseeva, Larissa (ed). The 18th European Symposium on Language for Special Purposes : Proceedings. Perm State National Research University, Perm, Russia. 11-26.

Cimiano, Philipp, McCrae, John and Paul Buitelaar. 2016. Lexicon Model for Ontologies : W3C Community Report.

Cimiano, Philipp, McCrae, John, Rodriguez-Doncel, Victor, Gornostaya, Tatiana, Gomez-Perez, Asuncion, Siemoneit, Benjamin, Lagzdins, Andis. 2015. Linked Terminology : Apply-ing Linked Data Principles to Terminological Resources. Proceedings of eLex 2015. 504-517.

DARIAH WG Lexical Resources. 2019. TEI Lex-0 — A baseline encoding for lexicographic data. Accessed September 19 https ://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html#index.xml-body.1_div.1_div.1

Declerck, Thierry, McCrae, John, Navigli, Roberto, Zaytseva, Ksenia, Wissik, Tanja . 2018. ELEXIS - European Lexicographic Infrastructure :

Contributions to and from the Linguistic Linked Open Data : Proceedings of the 2nd GLOBALEX Workshop, Japan.

Doel, Wim van den and Katrien Maes. 2012. Social Sciences and Humanities : Essential Fields for European Research and in Horizon 2020. League of European Research. Accessed November 14. https ://www.leru.org/files/Social-Sciences-and-Humanities-Essential-Fields-for-European-Research-and-Horizon-2020-Full-paper.pdf

ESFRI. 2018. "ESFRI roadmap 2018". Accessed June 19. http://roadmap2018.esfri.eu/media/1066/esfri-roadmap-2018.pdf

Illmayer, Klaus. 2017. Aufbau einer digitalen Infrastruktur für Theaterwissenschaft. Skizze einer digitalen Forschungsplattform. Presentation at DHA 2017, Innsbruck Austria. https ://zenodo.org/record/1123312#.XF1Q35rA9PY

Kallas, Jelena, Koeva, Svetla, Kosem,Iztok,Langemets, Margit, Tiberius, Carole. 2019. D1.1. Lexicographic Practices in Europe : A Survey of User Needs. Accessed September 19. https ://elex.is/wpcontent/uploads/2019/02/ELEXIS_D1_1_Lexicographic_Practices_in_Europe_A_Survey_of_User_Needs.pdf

Krek, Simon, Iztok Kosem, John P. McCrae, Roberto Navigli, Bolette S. Pedersen, Carole Tiberius, Tanja Wissik. 2018. European Lexicographic Infrastructure (ELEXIS). In Proceedings of the XVIII EURALEX International Congress Lexicography in Global Contexts. 881-891.

Lušicky, Vesna &Wissik. Tanja. 2017. Discovering Resources in the VLO : A Pilot Study with Students of Translation Studies. In Selected papers from the CLARIN Annual Conference 2016, Aix-en-Provence, 26-28 October 2016, CLARIN Common Language Resources and Technology Infrastructure, ed. L. Borin, 63-75. Linköping University Electronic Press, Linköpings universitet.

McCrae, John, Aguado-de Cea, Guadalupe, Buitelaar, Paul, Cimiano, Philipp, Declerck, Thierry, Gomez-Perez, Asuncion, Garcia, Jorge, Hollink, Laura, Montiel-Ponsoda, Elena, Spohr, Dennis and Wunner, Tobias. 2012. Interchanging Lexical Resources on the Semantic Web. Language Resources and Evaluation, 46(4):701-719.

McCrae, John P., Chiarcos, Christian, Bond, Francis, Cimiano, Philipp, Declerck, Thierry, Melo, Gerard de, Gracia, Jorge, Hellmann, Sebastian, Klimek, Bettina, Moran, Steven, Osenova, Petya, Pareja-Lora, Antonio, Pool, Jonathan. 2016. The Open Linguistics Working Group :: Developing the Linguistic Linked Open Data Cloud. Proceedings of the

Tenth International Conference on Language Resources and Evaluation (LREC'16).

McCrae, John P. Tiberius, Carole, Khan, Anas Fahad, Kernerman, Ilan, Declerck, Thierry, Krek, Simon, Monachini, Monica, Ahmadi, Sina. 2019. The ELEXIS Interface for Interoperable Lexical Resources. In Proceedings of the eLex 2019 conference. Biennial Conference on Electronic Lexicography (eLex-2019) Electronic lexicography in the 21st century. October 1-3 Sintra Portugal Pages 642-659 ISBN 2533-5626 Lexical Computing CZ, s.r.o Brno 10/2019.

Melby, Alan K. 2015. TBX : A terminological exchange format for the translation and localisation industry. In Kockaert, Hendrik J. and Frieda Steurs (eds.). Handbook of Terminology. Volume 1. Amsterdam, Philadelphy : John Benjamins Publishing.

Rodiguez-Doncel,Victor, Casanovas, Pompeu. 2018. A Linked Data Terminology for Copy-right Based on Ontolex-Lemon : AICOL. International Workshops 2015-2017. 410-423.

Romary, Laurent. 2015. TEI and LMF crosswalks. JLCL (30)

Romary, Laurent. 2014. "TBX goes TEI- Implementing a {TBX} basic extension for the Text Encoding Initiative guidelines", In *CoRR*. Accessed September 19. http://arxiv.org/abs/1403.0052

Terminorgs. 2014. TBX-BASIC. Version 3.1, Terminology for Large Organizations. Accessed September 19 http://www.terminorgs.net/downloads/TBX_Basic_Version_3.1.pdf

Tiberius, Carole and Declerck, Thierry. 2017. A lemon Model for the ANW Dictionary. In Proceedings of the eLex 2017 conference, Pages 237-251, Leiden, Netherlands, Lexical Computing CZ s.r.o., INT, Trojína and Lexical Computing, Brno, Czech Republic.

Résumé

Dans cette contribution, nous décrivons la (ré)utilisation de l'infrastructure de recherche ELEXIS pour la lexicographie afin de traiter des données terminologiques. Nous présentons les aspects centraux de l'infrastructure ELEXIS et les normes qu'elle applique et développe. Nous présentons également TBX, qui est la principale norme utilisée pour représenter les données terminologiques. Nous décrivons en détail les spécifications d'OntoLex-Lemon, qui résultent d'un "Community Group" du W3C et qui jouent un rôle central dans notre travail, notamment en décrivant les données terminologiques au sein d'une infrastructure pour la lexicographie, car ces spécifications permet-

tent de relier les systèmes d'organisation des connaissances à une description lexicale détaillée. Pour illustrer cette capacité, nous utilisons des données de terminologie multilingues, à l'origine codées en TBX, provenant du domaine de la gestion des risques.