

TOTh 2021

Terminologie & Ontologie: Théories et Applications

Terminologie & Ontologie: Théories et Applications

Actes de la conférence

TOTh 2021

Université Savoie Mont Blanc

3 & 4 juin 2021



Les ouvrages TOTh précédents sont disponibles :

- sur le site de l'Université Savoie Mont Blanc (btk.univ-smb.fr/livres)
- 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 : 978-2-37741-079-8
ISSN : 2607-5008
Dépôt légal : juillet 2022

Terminologie & Ontologie : Théories et Applications



Actes de la conférence

TOTH 2021

Université Savoie Mont Blanc

3 & 4 juin 2021

<http://toth.condillac.org>

avec le soutien de :

- Université Savoie Mont Blanc
- École d'ingénieurs Polytech Annecy Chambéry
- Ministère de la Culture. Ce projet est soutenu financièrement par le Ministère de la Culture - Délégation Générale à la Langue Française et aux Langues de France



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 2021

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>

Amparo Alcina	Universitat Jaume I – Spain
Xiaomi An	Renmin University – China
Albina Auksooriute	Institute of the Lithuanian Language – Lithuania
Jean-Paul Barthès	Université Technologie de Compiègne – France
Christopher Brewster	Maastricht University – Netherlands
Nicolleta Calzolari	Istituto di Linguistica Computazionale, CNR – Italy
Danielle Candell	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
Éric De La Clergerie	INRIA – France
Luc Damas	Université Savoie Mont-Blanc – France
Dardo De Vecchi	Kedge Business School – France
Thierry Declerck	DFKI – Germany
Valérie Delavigne	Université Paris 3 – France
Sylvie Desprès	Université Paris 13 – France
Juan Carlos Diaz Vasquez	EAFIT University – Colombia
Pamela Faber	Universidad de Granada – Spain
Christiane Fellbaum	Princeton University – USA
Cécile Frérot	Université Stendhal Grenoble 3 – France
Iolanda Galanes	Universidade de Vigo – Spain
Teodora Ghiviriga	Alexandru Ioan Cuza University – Romania
Rufus Gouws	University of Stellenbosch – South Africa
Jean-Yves Gresser	ancien Directeur à la Banque de France – France
John Humbley	Université Paris 7 – France
Yangli Jia	University of Liaocheng – China
Kyo Kageura	University of Tokyo – Japan
Barbara Karsch	BIK Terminology – USA
Hendrik Kockaert	University of Leuven – Belgium

Hélène Ledouble	Université de Toulon – France
Patrick Leroyer	Aarhus University – Denmark
Georg Löckinger	University of Applied Sciences Upper Austria – Austria
John McCrae	National University of Ireland – Ireland
Candida Jaci de Sousa Melo	Universidade Federal do Rio Grande do Norte – Brazil
Christine Michaux	Université de Mons – Belgium
Fidelma Ní Ghallchobhair	Foras na Gaeilge, Irish-Language Body – Ireland
António Pareja Lora	Universidad Complutense de Madrid – Spain
Silvia Piccini	Italian National Research Council – Italy
Suzanne Pinson	Université Paris Dauphine - France
Maria Pozzi	El colegio de México – Mexico
Bihua Qiu	China National Committee for Terms in Sciences and Technologies – China
Jean Quirion	Université d’Ottawa – Canada
Renato Reinau	Université de Genève – Switzerland
Christophe Roche	Université Savoie Mont-Blanc – France
Mathieu Roche	CIRAD – France
Laurent Romary	INRIA & HUB-ISDL – Germany
Micaela Rossi	Università degli studi di Genova – Italy
Klaus-Dirk Schmitz	Cologne University – Germany
Frieda Steurs	University of Leuven – Belgium
Toma Tasovac	Belgrade Center for Digital Humanities – Serbia
Philippe Thoiron	Université Lyon 2 – France
Kara Warburton	City University of Hong Kong – China
Maria Teresa Zanola	Università Cattolica del Sacro Cuore – Italy

Avant-propos



La pandémie de COVID-19 a durablement impacté nos façons de travailler, et en particulier l'organisation d'événements tels que les conférences. Si rien ne peut remplacer la richesse des contacts humains que procure le présentiel, il nous faut accepter d'autres modes de participation et d'échanges. La participation à distance, l'enregistrement des interventions en font partie. Au-delà d'une gestion différente des coûts et du temps, cela offre d'autres perspectives dont une plus grande diffusion des travaux menés et donc une meilleure visibilité.

Nous avons pu revenir en 2021 à la traditionnelle planification de TOTh la 1^{re} semaine de juin établie depuis 2007. La formation et la conférence se sont déroulées conjointement en présentiel et à distance, avec une très forte participation à distance, les restrictions sanitaires étant toujours en vigueur. L'organisation est certes plus compliquée et même si le présentiel devrait, nous l'espérons, revenir en force l'année prochaine, la participation à distance sera dorénavant proposée. L'Université de Savoie et l'équipe Condillac sur lesquelles reposent l'organisation de la conférence et la publication des actes seront plus fortement impliquées.

Avant de présenter les actes de cette année, j'aimerais remercier à nouveau les membres du Comité international de programme 2021 pour leur travail. Fortement mobilisés – les soumissions sont évaluées par au moins trois relecteurs – ils sont garants de la qualité des travaux menés à TOTh. Je rappelle que le Comité de programme est constitué chaque année à partir du Comité scientifique de TOTh en fonction des soumissions reçues. Le Comité scientifique est composé de 75 membres, experts internationalement reconnus du domaine, représentant 24 nationalités différentes.

La Conférence TOTh 2021 s'est ouverte avec la conférence invitée de notre collègue Nicola Guarino, bien connu des «ontologues», qui a dirigé le laboratoire d'ontologie appliquée de Trento rattaché au Conseil National de la Recherche italienne. Son intervention a porté sur «Events and their Names», un sujet aussi difficile qu'il est important. Nous présentons ici un résumé d'une communication que Nicola Guarino développera dans une communication ultérieure.

Notre collègue, François Gaudin, de l'Université de Rouen, a proposé cette année une Disputatio nous invitant à une lecture sociolinguistique de la référence chez Hilary Putnam.

Sur les 13 communications présentées, seules 10 ont été retenues pour publication. Elles ont abordé de nombreux sujets tant théoriques que pratiques portant sur des domaines aussi variés que les humanités numériques, la finance, les modèles de représentation, ou l'harmonisation de termes et de concepts.

Cette année nous avons eu le plaisir de décerner deux prix jeunes chercheurs. Cela est suffisamment exceptionnel pour que nous y consacrons quelques lignes. Instauré en 2011, ce prix n'a été décerné que deux fois, en 2011 et en 2018. Cette année ce sont deux jeunes chercheuses, toutes les deux italiennes, qui ont été récompensées. Cristina Farroni, de l'Università degli studi di Macerata, a présenté une contribution intitulée « Collaborative terminology management in a business environment : a case study in the field of wood paints and coatings ». Federica Vezzani, de l'Università degli studi di Padova, nous a présenté ses travaux en français sur le thème de « La gestion de (méta)données terminologiques « FAIR » : le répertoire de catégories de données de la ressource TriMED ».

Plus de 60 personnes ont suivi de manière assidue les présentations, ce qui correspond à la participation moyenne à la conférence. 21 pays étaient représentés : Afrique du Sud, Albanie, Allemagne, Autriche, Belgique, Chine, Espagne, États-Unis, France, Ghana, Grèce, Hongrie, Irlande, Italie, Lituanie, Luxembourg, Portugal, Roumanie, Royaume-Uni, Sénégal, Suisse.

Je vous invite à découvrir les communications que nous avons retenues à travers ces actes réalisés avec M^{me} Catherine Brun et publiés aux Presses Universitaires Savoie Mont-Blanc. Les actes des années précédentes sont accessibles à partir du site de la conférence (<http://toth.condillac.org/>) et des Presses Universitaires Savoie Mont Blanc (https://btk.univ-smb.fr/livres/?fwp_collections_revues=terminologica).

Avant de vous souhaiter bonne lecture, j'aimerais terminer en remerciant le Ministère de la Culture, et plus précisément la Délégation Générale à la Langue Française et aux Langues de France, l'Université Savoie Mont-Blanc, l'École Polytech Annecy-Chambéry et l'équipe Condillac pour leur support et leur aide financière à l'organisation de la conférence et à la publication des actes.

Christophe Roche
Président du Comité scientifique

SOMMAIRE

CONFÉRENCE D'OUVERTURE	13
Events and their Names	
Nicola Guarino	15
ARTICLES	19
Formalizations of Knowledge and/in Semiotic Models in Terminology Science	
Marija Ivanović, Thierry Declerck,	21
Creating a termino-ontological resource for translators in the domain of viral infectious diseases	
Alice Sanfilippo	39
Crisis in troubled ancient times: ontological modelling of textual evidence from Greek historians	
Maria Papadopoulou, Eleni-Melina Tamiolaki, Christophe Roche	57
Evolution of Modular Ontology: Application to Personalization	
Rahma Dandan, Sylvie Despres	83
The Harmonization of Terms and Concepts in Function of the Standardization of Computer Terminology in the Albanian Language	
Anila Çepani, Adelina Çerpja	105
Extending TBX2RDF	
Thierry Declerck, Patricia Martín Chozas, Tom Winter, Tanja Wissik	123
Transformation of Graphical Semiotic Models into a Graph-based Formal Representation	
Thierry Declerck, Marija Ivanović	137
Role of the Corpus in Ontoterminology: the Case of the Balance of Payments	
Stéphane Carsenty	153

**Collaborative terminology management in a business environment :
a case study in the field of wood paints and coatings**

Cristina Farroni

181

**La gestion de (méta)données terminologiques «FAIR»: le
répertoire de catégories de données de la ressource TriMED**

Federica Vezzani

199

Extending TBX2RDF

Thierry Declerck*/**, Patricia Martín Chozas***, Tom Winter****,
Tanja Wissik**

*DFKI GmbH, Saarland Informatics Campus
Multilinguality and Language Technology Lab
Stuhlsatzenhausweg, 3
D-66123 Saarbrücken, Germany
declerck@dfki.de

[https://www.dfk.de/~declerck /](https://www.dfk.de/~declerck/)

**Austrian Academy of Sciences

Austrian Centre for Digital Humanities and Cultural Heritage
Sonnenfelsgasse, 19
A-1010 Vienna, Austria
tanja.wissik@oeaw.ac.at

<https://www.oeaw.ac.at/acdh/team/current-team/tanja-wissik>

***Universidad Politécnica de Madrid

Ontology Engineering Group
Campus de Montegancedo, s/n. ETSIINF, 28660 Boadilla del Monte, Madrid, Spain
pmchozas@fi.upm.es

****Deutsche Bahn AG

Sprachenmanagement (IBS)

Gallusanlage, 8

D-60329 Frankfurt a. Main, Germany

Tom.Winter@deutschebahn.com

Abstract. Following the goal of transforming two multilingual terminologies to a native graph representation format, we propose an extension of an already existing initiative tackling this matter, named as TBX2RDF, which leads to the design of a new RDF-based modelling of traditional terminological data. The original multilingual terminological data handled in this work come from the terminology of the Deutsche Bahn AG and the Interactive Terminology for Europe (IATE).

1. Introduction

We present in this paper results from our work which consists in porting two multilingual terminologies from their TBX¹ representation onto an RDF² format. While the purpose of this exercise is not to change anything at the level of the content of the original terminologies, their modelling in a native graph-based representation offers the possibility of interlinking and merging them with other resources, being in the realm of terminologies or other types of resources, like for example detailed lexicographic resources, knowledge bases and even ontologies.

Within this context, we came into several modelling suggestions that go beyond the TBX2RDF converter³, an existing application to map TBX data to RDF vocabularies, which we also present in this paper. This eventually led to the design of a new RDF-based modelling approach for terminologies, which is now being discussed within the Ontolex W3C Community Group⁴. Through the following sections, we present the two multilingual terminological resources that are the core of our modelling work. In Section 4, we describe the LIDER project⁵ in which the TBX2RDF service was developed, alongside with the resulting guidelines and ontology. Section 5 presents some suggested extensions for the TBX2RDF model and finally, in Section 6, we conclude and suggest future work.

2. Deutsche Bahn Terminology

The Deutsche Bahn (DB)⁶ terminology aims, amongst others, at supporting the group-wide linguistic consistency, as a uniform use of language is gaining in importance with internationalization. The DB Language Portal helps with the translation of railway and company-specific words. Its underlying terminology covers 16 languages and can be accessed online⁷. The

1 TBX stands for “TermBase eXchange”. See <https://www.tbxinfo.net/> [accessed 2021-10-24], See also (Lommel *et al.*, 2014) and (Melby, 2015).

2 RDF stands for “Resource Description Framework”. See <https://www.w3.org/TR/rdf-primer/>.

3 An updated version of those guidelines is available at <https://github.com/bpmlod/report/blob/gh-pages/multilingual-terminologies/index.html> [accessed 2021-10-24].

4 <https://www.w3.org/community/ontolex/>

5 <http://lider-project.eu/lider-project.eu/index.html> [accessed 2021-10-24].

6 Deutsche Bahn AG is a large German railway company.

7 www.deutschebahn.com/dblanguageportal [accessed 2021-10-24].

data we are working on is a TBX file export for Multiterm⁸ friendly provided for this study by the Language Department of the Deutsche Bahn. In this export, values of the “type” attribute are written in German language (one our intention is to propose a way to support multilingualism for those values). Examples of this export, covering here only German and French terms, are given in the Annex.

3. Interactive Terminology for Europe (IATE)

IATE is the multilingual terminology database of the European Union, intended to support EU translators and as a means of standardisation of terminology across all institutions. It is available through an open access platform and constitutes the most important terminological reference for translators and language users in Europe. It collects terminological data from ten European Union’s partners, including the European Parliament, the European Commission, the Court of Justice of the EU, the European Central Bank, and the Translation Centre for the Bodies of the European Union (CdT), to mention but a few.

The database puts together terminologies that were previously generated by the above-mentioned partners, such as Euroterms from the CdT and CuriaTerm from the Court of Justice. This means that this resource is also multidisciplinary, including terms in the financial, environmental, and labour domain, amongst others. Overall, IATE contains more than 8 million terms in the 24 EU languages plus Latin.

4. LIDER Guidelines and TBX Ontology

The aim of the past LIDER project was “to provide the basis for the creation of a Linguistic Linked Data cloud that can support content analytics tasks of unstructured multilingual cross-media content”⁹, also including the mapping TBX to RDF. LIDER developed for this purpose a series of guidelines¹⁰

8 MultiTerm is a termbase management software, provided by SDL. See <https://docs.rws.com/785445/641133/sdl-multiterm-2019/welcome-to-----sdl-multiterm-2019> [accessed 2021-10-24].

9 <http://lider-project.eu/lider-project.eu/index.html> [accessed 2021-10-24].

10 An updated version of those guidelines is available at <https://github.com/bpmlod/report/blob/gh-pages/multilingual-terminologies/index.html> [accessed 2021-10-24].

and a TBX ontology¹¹, in which TBX elements are converted into OWL¹² and associated with other RDF vocabularies, while the basic vocabulary chosen as the backbone of the conversion was the *lemon* model¹³. (Cimiano *et al.*, 2015) and (McCrae *et al.*, 2015) describe the TBX2RDF approach and the resulting resources. (di Buono *et al.*, 2020) presents recent developments related to this TBX to RDF initiative, which consist in transforming and publishing terminologies as linked data, relying on a virtualization approach that is making use of containerization technologies.

While the LIDER TBX2RDF approach is representing the TBX terminological concepts as `skos:Concept` and the TIG/NTIG elements of TBX as `ontolex:LexicalEntry`, most of the other TBX elements are straightforwardly mapped onto RDF, meaning that they are encoding as URIs for representing a resource that can be associated with RDF predicates and objects. We note also that TBX2RDF does not represent the TBX `langSet` data as such, but instead is creating language specific lexicons in which all the data included in the original `langSet` element are encoded.

In our transformation work, we are investigating if all of the original TBX elements can be modelled also by reusing existing vocabularies, in order to take more advantage of the graph-based modelling facility that is supported by RDF. In doing so, we can introduce sub-class hierarchies and support the translation of all elements of TBX, beyond the sole translation of terms. Our model can still be mapped backward to a native TBX representation.

5. TBX2RDF Extensions

In our current work we make use of the most recent version of OntoLex-Lemon, which is effectively integrating the SKOS vocabulary¹⁴ for expressing conceptual units. This was not the case for the former *lemon* vocabulary, which was used in the LIDER project. But the difference in the modelling is minimal, as we can now use properties defined in OntoLex-Lemon for linking the concepts to lexical entries, while the LIDER TBX2RDF converter

11 <https://github.com/cimiano/tbx2rdf/blob/master/ontology/tbx.owl> [accessed 2021-10-24].

12 OWL stands for “Web Ontology Language”. See <https://www.w3.org/TR/owl2-primer/> [accessed 2021-02-13].

13 See (McCrae *et al.*, 2012) for more details.

14 SKOS stands for “Simple Knowledge Organization System”. It is a W3C recommendation as a “common data model for knowledge organization systems such as thesauri, classification schemes, subject heading systems and taxonomies” (<https://www.w3.org/TR/skos-reference/> [accessed 2021-10-24]).

was using a custom property for this purpose. We introduce a skos:ConceptScheme for encoding the whole conceptual organisation of the DB terminology, and within this scheme we allow for the definition of specific domain subsets, something which is not explicitly foreseen in the original terminology, and maybe also not possible to formulate in TBX¹⁵.

As the most recent version of OntoLex-Lemon is foreseeing a class `ontolex:LexicalConcept` for linking lexical entries to the conceptual part described in the SKOS vocabulary, we encode all the term IDs as instances of this class, as this can be seen in FIG. 3.

Another, and more significant, departure from the LIDER TBX2RDF approach is the fact that we model definitions and contexts as instances of classes, and no longer as literal values, as was done previously by applying the property `skos:definition`. In doing so, we can describe specific relations between the definitions within one language or across different languages. In the latter case, we can specify if the definitions given for concepts in two different languages are translations of each other, are multilingual equivalents or are just monolingual definitions included in the multilingual terminology.

We have encountered those issues for the definitions and notes in both the DB and the IATE terminology resources. Within their entries, we can find data about term definitions and notes that cannot be represented only as literal values of properties. Modern approaches to generate terminologies from diverse resources also have an imperative need to represent the provenance of each piece of data, a feature that current vocabularies do not support.

FIG. 1 and FIG. 2 below are showing how definitions and notes are included in both the DB and the IATE terminologies.

fast train path		Entry 23003
validiert		ja
Sachgebiet		Betrieb - Fahrplan
Definition_DE:		Trassenprodukt der DB Netz AG. Zügen mit dem Zusatz "Schnell" wird grundsätzlich Vorrang in der betrieblichen Durchführung vor allen Zügen gewährt, mit Ausnahme von dringlichen Hilfszügen und anderen Zügen mit dem Zusatz "Express" bzw. "Schnell".
Definition_EN		Train path product offered by DB Netz. Trains whose train path product has the suffix "Schnell" are generally given priority in traffic management over all trains with the exception of urgent rescue trains and other trains with the suffix "Express" or "Schnell".
Quelle_Def_DE		Trassenpreisystem 2018 der DB Netz AG, gültig vom 10. Dezember 2017 bis 8. Dezember 2018
Quelle_Def_EN		"The Track Access Charges 2018 of DB Netz AG", valid from 10 December 2017 through 8 December 2018

FIG. 1 – *Term excerpt from the DB Portal showing definitions and their sources.*

15 See (Reineke and Romary, 2019) for a discussion on the difference between the “subject-Field” in TBX and the conceptual hierarchy in SKOS.

Definition:	pharmacoepidemiological study or a clinical trial carried out in accordance with the terms of the marketing authorisation, conducted with the aim of identifying or quantifying a safety hazard relating to an authorised medicinal product
Definition reference:	Directive 2001/83/EC on the Community code relating to medicinal products for human use, CELEX:32001L0083/EN
Note:	A PASS may be a clinical trial or a non-interventional study. For more information, see: European Medicines Agency. <i>Guideline on good pharmacovigilance practices (GVP). Module VIII – Post-authorisation safety studies</i> . 2012 http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/02/WC500123204.pdf [28.3.2012]
Owner:	COM




FIG. 2 – Entry from the IATE portal showing the definition of the term “post-authorization safety study” with a note with several elements.

To overcome the above-mentioned gaps from the existing vocabularies, we have worked on a complementary module to model current terminological resources. We have published a first draft of this proposal as a wiki page within the Ontology Lexica Community Group. This extension proposal, temporarily named as Termlex, works with the `ontolex:LexicalConcept` as its key element. Therefore, one term ID is modelled as one `LexicalConcept`. This idea differs from the original intention of the Ontolex vocabulary where one `LexicalConcept` could have many `LexicalSenses` with several `LexicalEntries` associated. To avoid ambiguity, following traditional terminology theories, we restrict to one `LexicalSense` per `LexicalConcept`. Such a restriction can be represented with a functional property.

We present in this paper the aspects of Termlex dealing with definitions and notes, for which 3 classes are introduced:

Definition Class: a definition is a statement that explains the meaning of a concept. As mentioned earlier, we need the class `Definition` since this language data is much more than a literal. For instance, in FIG. 1, we see the terminology card of the term “fast train path”, with two definitions and each of them has a different source that must be represented. Also, we propose a distinction amongst lexicographic and terminological definitions since their scope is also different. We therefore suggest creating the subclasses `LexicographicDefinition` and `TerminologicalDefinition`.

Note Class: the same happens with notes. A note may contain much more data than a simple text. In FIG. 2, we identify four different types of information in the note attached to the term “post-authorization safety study”:

- The value (*A PASS may be a clinical trial...*), a literal that could be represented with the property `rdf:value`.
- A note for more information (*For more information, see...*), that could be modelled with a `skos:note` property, inside our Note class.

A pointer to the additional information (<http://ema.europa.eu...>), that could be modelled with a `rdfs:seeAlso` property.

The creation date of the note (28.3.2012) that could be represented with `dcterms:created` or simply `dcterms:date`.

Source Class: Likewise, sources can be much more than one piece of data: title, identifier, author, organisation, or application, to mention but a few. Therefore, we propose a class to collect them all. This class is especially needed when terminologies are generated from multiple resources, to maintain the traceability, since sources may even be chained, for instance, a definition that has been extracted with an application (source 1) from a given corpus (source 2). To model each element inside our Source class we also reuse other vocabularies such as DublinCore¹⁶ or the Prov Ontology¹⁷. Our proposed inclusion of those classes with elements of OntoLex-Lemon is displayed in FIG. 3.

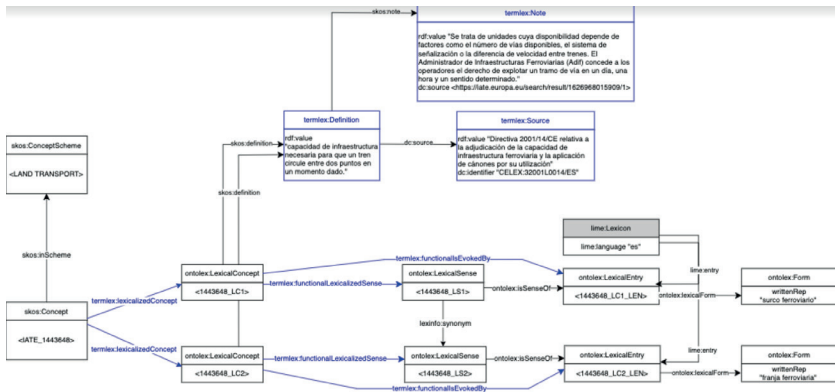


FIG. 3 – *Integration of the classes Definition, Note and Source and the OntoLex-Lemon classes as proposed in Termlex*

16 See for more details: <https://dublincore.org/> [accessed 2021-10-24]. We use for example DublinCore for marking the creation date (`dcterms:created` or `dcterms:date`).

17 See for more details: <https://www.w3.org/TR/prov-o/> [accessed 2021-10-24].

We have also reviewed other vocabularies to reuse their classes to represent other elements of terminological entries, such as the term contexts, for which we are reusing the class `lexicog:UsageExample`, from the Lexicog¹⁸ extension of Ontolex to model lexicographic resources; and the recommended usage of a term, for which we are reusing the class `lexinfo:NormativeAuthorization`, from the Lexinfo model¹⁹, that has a fixed list of values (preferredTerm, deprecatedTerm, admittedTerm...).

Additionally, we introduce an explicit class hierarchy for the subdomains of the original DB terminology. There, the hierarchical relations between subject fields are represented by the sole use of a “hyphen” sign, like:

- `terms:subject "\Recht und Regulierung\''"@de-DE`
- `terms:subject "\Recht und Regulierung - Verkehrsrecht\''"@de-DE`
- `terms:subject "\Recht und Regulierung - Verträge\''"@de-DE;`

FIG. 4 displays a screenshot from the ontology editor we are using and shows, in the left panel, the subclass hierarchy we introduced, in our OWL-based representation of the original DB terminology. This way, we might apply inference mechanisms to all the subclasses of a specific domain of the terminology.

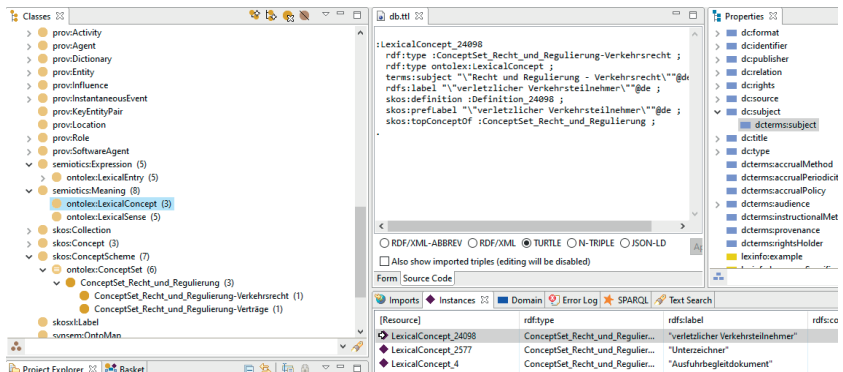


FIG. 4 – Representing hierarchies within domains of the DB terminology

18 See for more details: <https://www.w3.org/2019/09/lexicog/>

19 See for more details: <https://lexinfo.net/>

6. Conclusions and Future Work

We described on-going work in porting the multilingual terminologies of the Deutsche Bahn and of IATE onto a Linked Data compliant representation language. This work led us to the question if it would not be suitable to extend the modelling of TBX terminologies in RDF already proposed by the LIDER TBX2RDF converter. One aspect consists in considering definitions or contexts as full ontological elements that can thus be put explicitly in relation to each other.

Another aspect we are working on is dealing with the inclusion of explicit subclass hierarchies, supporting the application of inference mechanisms to the specialisations of domains of the terminology.

Current work is dealing with the formulation of lexical restrictions that can apply to the terms, along the lines of the approach described in (Declerck *et al.*, 2020).

Our proposal is now under discussion for admission as a new module extension to the OntoLex-Lemon framework.

Acknowledgements

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. 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. The work described in this paper was also pursued in part in the larger context of the COST Action CA18209 - NexusLinguarum “European network for Web-centred linguistic data science”.

References

- Philipp Cimiano, John P. McCrae, and Paul Buitelaar. (2016). Lexicon Model for Ontologies: Community Report. W3C Community Group Final Report
- Maria Pia di Buono, Philipp Cimiano, Mohammad Fazleh Elahi, Frank Grimm. (2020). Terme-à-LLOD: Simplifying the Conversion and Hosting of Terminological Resources as Linked Data. In Proceedings of the 7th Workshop on Linked Data in Linguistics (LDL-2020)

- Cimiano, P., McCrae, J. P., Rodríguez-Doncel, V., Gornostay, T., Gómez-Pérez, A., Siemoneit, B., and Lagzdins, A. (2015). Linked terminologies: applying linked data principles to terminological resources. In Proceedings of the eLex 2015 Conference, pages 504-517.
- John McCrae, Guadalupe Aguado-de-Cea, Paul Buitelaar, Philipp Cimiano, Thierry Declerck, Asunción Gómez-Pérez, Jorge Gracia, Laura Hollink, Elena Montiel-Ponsoda, Dennis Spohr and Tobias Wunner, *Language Resources and Evaluation*, 46(6), pp 701-709, (2012).
- McCrae, J. P., Cimiano, P., and Doncel, V.R. (2015). Guidelines for linguistic linked data generation: Multilingual terminologies (TBX) -- Final Community Group Report. <https://www.w3.org/2015/09/bpmlod-reports/multilingual-terminologies/>
- Declerck, Thierry, Racioppa, Stefania, Wissik, Tanja. 2020. Formal Representation of Linguistic Restrictions expressed in Terminologies. In Proceedings of TOTH 2020.
- Lommel, Arle, Melby, Alan, Glenn, Nathan, Hayes, James and Snow, Tyler. 2014. TBX-Min: A Simplified TBX-Based Approach to Representing Bilingual Glossaries. Proceedings of *Terminology and Knowledge Engineering 2014*, Jun 2014, Berlin, Germany. 10p.
- 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.
- Detlef Reineke, Laurent Romary. Bridging the gap between SKOS and TBX. *Die Fachzeitschrift für Terminologie*, Deutscher Terminologie-Tag e.V. (DTT), 2019, *Begriffssysteme und ihre Darstellung*, 19 (2).

Annex: Examples from the TBX Export of the Deutsche Bahn Terminology (German and French)

```

<termEntry id="4">
  <descripGrp>
    <descrip type="validiert">Ja</descrip>
  </descripGrp>
  <descripGrp>
    <descrip type="Sachgebiet">Recht und Regulierung</descrip>
  </descripGrp>
  <descripGrp>
    <descrip type="Definition">Ein Nachweis der Zollstelle über die Zulässigkeit der Ausfuhr. Das Ausfuhrbegleitdokument wird auch Ausfuhranmeldung oder früher Ausfuhrerklärung genannt.</descrip>
  </descripGrp>
  <descripGrp>
    <descrip type="Quelle der Definition">
      www.dashoefer.de/thema/ausfuhrbegleitdokument.html, 01.08.2017
    </descrip>
  </descripGrp>
</termEntry>

```

```

<langSet xml:lang="de-DE">
  <tig>
    <term id="18">Ausfuhrbegleitdokument</term>
    <descripGrp>
      <descrip type="Benennungstyp">Vollform</descrip>
    </descripGrp>
    <descripGrp>
      <descrip type="Gebrauch">bevorzugt</descrip>
    </descripGrp>
    <descripGrp>
      <descrip type="Genus">n</descrip>
    </descripGrp>
    <descripGrp>
      <descrip type="Datenquelle">Sprachenmanagement DB AG</descrip>
    </descripGrp>
    <descripGrp>
      <descrip type="Zuverlässigkeit">verifiziert</descrip>
    </descripGrp>
    <descripGrp>
      <descrip type="Kontext">Das Ausfuhrbegleitdokument ersetzt die Ausfuhranmeldung in Papierform. Auf dem Ausfuhrbegleitdokument sind die Bezugsnummer der Sendung (Moving Reference Number, MRN) und ein Barcode enthalten, der bei Verlassen des Gemeinschaftsgebiets gescannt wird und einen Ausgangsvermerk (AgV) erzeugt. Dieser ersetzt den Stempelaufdruck beim Ausfuhrbegleitdokument in Papierform.
    </descrip>
  </descripGrp>
  <descripGrp>
    <descrip type="Quelle für Kontext">
      www.dashoefer.de/thema/ausfuhrbegleitdokument.html, 01.08.2017
    </descrip>
  </descripGrp>
</tig>

```

```

<tig>
  <term id="19">ABD</term>
  <descripGrp>
    <descrip type="Benennungstyp">Abkürzung</descrip>
  </descripGrp>
  <descripGrp>
    <descrip type="Gebrauch">erlaubt</descrip>
  </descripGrp>
  <descripGrp>
    <descrip type="Datenquelle">Sprachenmanagement DB AG</descrip>
  </descripGrp>
  <descripGrp>
    <descrip type="Zuverlässigkeit">verifiziert</descrip>
  </descripGrp>
</tig>
</langSet>

```

```

<langSet xml:lang="fr-FR">
  <tig>
    <term id="22">document d'accompagnement d'exportation</term>
    <descripGrp>
      <descrip type="Benennungstyp">Vollform</descrip>
    </descripGrp>
    <descripGrp>
      <descrip type="Gebrauch">bevorzugt</descrip>
    </descripGrp>
    <descripGrp>
      <descrip type="Datenquelle">Sprachenmanagement DB AG</descrip>
    </descripGrp>
    <descripGrp>
      <descrip type="Zuverlässigkeit">verifiziert</descrip>
    </descripGrp>
  </tig>

```

```

<tig>
  <term id="23">DAE</term>
  <descripGrp>
    <descrip type="Benennungstyp">Abkürzung</descrip>
  </descripGrp>
  <descripGrp>
    <descrip type="Gebrauch">erlaubt</descrip>
  </descripGrp>
  <descripGrp>
    <descrip type="Datenquelle">Sprachenmanagement DB AG</descrip>
  </descripGrp>
  <descripGrp>
    <descrip type="Zuverlässigkeit">verifiziert</descrip>
  </descripGrp>
</tig>

```

```
<tig>
  <term id="24">document d'accompagnement export</term>
  <descripGrp>
    <descrip type="Benennungstyp">Vollform</descrip>
  </descripGrp>
  <descripGrp>
    <descrip type="Gebrauch">erlaubt</descrip>
  </descripGrp>
  <descripGrp>
    <descrip type="Datenquelle">Sprachenmanagement DB AG</descrip>
  </descripGrp>
  <descripGrp>
    <descrip type="Zuverlässigkeit">verifiziert</descrip>
  </descripGrp>
</tig>
</langSet>
</termEntry>
```

