



Building the Legal Knowledge Graph for Smart Compliance Services in Multilingual Europe

D2.5 Report on Lynx Acquired Vocabularies

| | |
|---------------------------------|--|
| PROJECT ACRONYM | Lynx |
| PROJECT TITLE | Building the Legal Knowledge Graph for Smart Compliance Services in Multilingual Europe |
| GRANT AGREEMENT | H2020-780602 |
| FUNDING SCHEME | ICT-14-2017 - Innovation Action (IA) |
| STARTING DATE (DURATION) | 01/12/2017 (36 months) |
| PROJECT WEBSITE | http://lynx-project.eu |
| COORDINATOR | Elena Montiel-Ponsoda (UPM) |
| RESPONSIBLE AUTHORS | Ilan Kernerman (KD), Patricia Martín Chozas (UPM), Andis Lagzdinš (Tilde), Jorge Gracia (Unizar) |
| CONTRIBUTORS | Dorielle Lonke (KD), Raya Abu Ahmad (KD), Sotiris Karampatakis (SWC) |
| REVIEWERS | Elena Montiel-Ponsoda (UPM), Stefanie Hegele (DFKI), Julian Moreno-Schneider (DFKI) |
| VERSION STATUS | V1.0 Final |
| NATURE | Report |
| DISSEMINATION LEVEL | Public |
| DOCUMENT DOI | 10.5281/zenodo.3558710 |
| DATE | 30/11/2019 (M24) |



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780602

| VERSION | MODIFICATION(S) | DATE | AUTHOR(S) |
|---------|--|------------|-------------------------------------|
| 01 | First version | 1/11/2019 | Ilan Kernerman (KD) |
| 01 | Initial review | 4/11/2019 | Stefanie Hegele (DFFKI) |
| 02 | Update of 2.1.2 and 3 | 11/11/2019 | Patricia Martín Chozas (UPM) |
| 03 | Additional section for RDF conversion of DIV | 11/11/2019 | Jorge Gracia (UNIZAR) |
| 03 | Revision accepting changes, answering comments | 12/11/2019 | Ilan Kernerman (KD) |
| 03 | Update of ToC, tables and figure numbers, style unification and reformatting | 14/11/2019 | Patricia Martin Chozas (UPM) |
| 04 | Rearrangement of sections and ToC, document styling and reformatting | 15/11/2019 | Dorielle Lonke (KD) |
| 04 | Multiple (re)adjustments, re-introduction of changes and reply to comments, reformat | 15/11/2019 | Ilan Kernerman (KD) |
| 05 | Second review | 15/11/2019 | Julian Moreno Schneider (DFKI) |
| 06 | Final draft for review | 17/11/2019 | Ilan Kernerman (KD) |
| 06 | Minor corrections | 19/11/2019 | Jorge Gracia (UNIZAR) |
| 07 | Reference to comments | 20/11/2019 | Julian Moreno Schnieder (DFKI) |
| 07 | Updates and reformatting | 20/11/2019 | Ilan Kernerman, Raya Abu-Ahmad (KD) |
| 07 | Comments | 26/11/2019 | Elena Montiel Ponsoda (UPM) |
| 08 | Update of section 3 | 26/11/2019 | Patricia Martin Chozas (UPM) |
| 09 | Updates and reformatting | 27/11/2019 | Ilan Kernerman, Raya Abu-Ahmad (KD) |
| 10 | Minor adjustments | 28/11/2019 | Ilan Kernerman (KD) |

DISCLAIMER

This document does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of its content. Neither the Lynx consortium as a whole, nor a certain party of the Lynx consortium warrant that the information contained in this document is capable of use, nor that use of the information is free from risk, and does not accept any liability for loss or damage suffered by any person using this information.

EXECUTIVE SUMMARY

This deliverable constitutes the final report on the vocabulary resources collected and generated in the Lynx project, covering both general language and legal terminology. It is part of Work Package 2, which is concerned with the acquisition and management of linguistic data for Lynx, and includes the process followed for (a) the identification of existing language resources (corpora, dictionaries, glossaries, thesauri, terminologies, ontologies) of interest for the domains covered in the project; (b) the adaptation of resources coming from Lynx partners; and (c) the creation of language resources to cover the specific needs of Lynx use cases (when such data sources did not exist or were not freely available). Both types of legal and general vocabularies play a vital role in the Legal Knowledge Graph that is at the heart of Lynx, to enable and facilitate its multilingual services.

The existing vocabularies ensue mainly from two Lynx partners (TILDE and KD), along with some open access terminological resources in particular, and are complemented by resources specifically created to meet business case needs. This process has met with considerable success so far, as shown by the number and quality of vocabularies reported upon in this document, and will be enhanced, if necessary, in the context of WP3 (service development) and WP5 (pilots), where language resources are going to be used. The aim is to further reinforce, enhance and improve the legal knowledge services provided by Lynx, and align more closely with its three use case partners. The Lynx partners would thus be able to offer and assure yet higher standards for its linguistic infrastructure regarding both general language and legal terminology.

TABLE OF CONTENTS

| | | |
|-------|---|----|
| 1 | INTRODUCTION | 7 |
| 2 | VOCABULARY IDENTIFICATION: DOMAIN DEPENDENT AND DOMAIN INDEPENDENT VOCABULARIES | 9 |
| 2.1 | DOMAIN DEPENDENT VOCABULARIES (DDV)..... | 9 |
| 2.1.1 | Existing DDV | 9 |
| 2.1.2 | Harvesting DDV | 9 |
| 2.2 | DOMAIN INDEPENDENT VOCABULARIES (DIV)..... | 13 |
| 2.2.1 | Existing DIV | 13 |
| 2.2.2 | Harvesting DIV..... | 15 |
| 3 | NEW DDV GENERATION | 16 |
| 3.1 | CREATING VOCABULARIES | 16 |
| 3.2 | DOCUMENTING VOCABULARIES..... | 18 |
| 4 | XML TO RDF CONVERSION OF DIV | 19 |
| 4.1 | INTRODUCTION | 19 |
| 4.2 | REPRESENTATION MODEL | 20 |
| 4.3 | METHODOLOGY | 22 |
| 5 | SUMMARY AND OUTLOOK..... | 24 |

TABLE OF FIGURES

| | |
|---|----|
| Figure 1. Workflow of processing different data formats | 11 |
| Figure 2. Global language cores and pairs | 14 |
| Figure 3. Password & MultiGloss cores and pairs | 15 |
| Figure 4. Automatic Generation of Legal Linked Terminologies | 16 |
| Figure 5. SKOS template applied | 17 |
| Figure 6. Sample of a linked terminological entry | 18 |
| Figure 7. Language pairs in the original, isolated, dictionaries. | 21 |
| Figure 8. Language pairs in the resulting RDF graph | 22 |
| Figure 9. Scheme of the lexicography module | 23 |
| Figure 10. Conversion flow of data format from XML to JSON to JSON-LD to enabling SPARQL queries .. | 24 |

TABLE OF TABLES

| | |
|---|----|
| Table 1. Data Portal Licence Analysis | 13 |
| Table 2. Basic statistics per language in the Global series | 14 |
| Table 3. Basic statistics per language in Password & MultiGloss | 14 |
| Table 4. Numbers of lemmas & inflected forms..... | 15 |

APPENDICES

| | |
|---|----|
| Appendix 1. Relevant language resources documented in Lynx Data Portal..... | 25 |
| Appendix 2. Existing Domain Independent vocabularies from KD..... | 32 |
| Appendix 3. Sample data of Domain Independent Vocabularies from KD | 33 |
| Appendix 4. JSON-LD conversion of DIV | 33 |

ACRONYMS

| | |
|------------|--|
| API | Application Program Interface |
| CEF | Common European Framework of Reference for Language |
| CKAN | Web-based management system for the storage and distribution of data |
| CSV | Comma-separated values (file format) |
| DDV | Domain-Dependent Vocabularies |
| DIV | Domain-Independent Vocabularies |
| DOC, DOCX | Filenames extension for document files |
| ELRA | European Language Resources Association |
| ELRC-SHARE | European Language Resource Coordination |
| eTTB | e-Translation Termbank |
| EU | European Union |
| FAIR | Findable, Accessible, Interoperable and Reusable |
| HTML | Hypertext Markup Language |
| IPR | Intellectual Property Rights |
| JPG/JPEG | Joint Photographic (Experts) Group (compression for digital images) |
| JSON(-LD) | Java Script Object Notation (-Linked Data) |
| KD | K Dictionaries Ltd |
| LD | Linked Data |
| LKG | Legal Knowledge Graph |
| LLOD | Linguistic Linked Open data |
| LOD | Linked Open Data |
| LRs | Language resources |
| MT | Machine Translation |
| OCR | Optical character recognition |
| PDF | Portable Document Format |
| PNG | Portable Network Graphics (image file format) |
| RDF | Resource Description Framework |
| SKOS | Simple Knowledge Organization System |
| SPARQL | Semantic (RDF) query language for databases |
| TBX | TermBase eXchange |

| | |
|-----------|---|
| TMX | Translation Memory eXchange (file format) |
| TR | Terminology resources |
| TSV | Tab-separated values |
| TXT | A filename extension for text files |
| UPM | Universidad Politécnica de Madrid |
| URI | Uniform Resource Identifier |
| W3C | World Wide Web Consortium |
| WP2, WP3 | (Lynx) Working Package 2, Working Package 3 |
| XLS, XLSX | Filenames extension for Microsoft Excel sheet format |
| XML | Extended Markup Language |
| XSL | Extensible Stylesheet Language |
| XSLT | Extensible Stylesheet Language Transformations (for transforming XML) |

1 INTRODUCTION

This final report on Lynx Acquired Vocabularies provides an account of the various language resources (LRs) that are made available as part of the Legal Knowledge Graph (LKG) for use in the Lynx services. The report is part of Work Package 2 (WP2), dealing with the acquisition and management of vocabularies used in Lynx. In addition to classifying and describing these LRs, we present some of the techniques and methodologies for their processing to prepare the common services infrastructure needed for Work Package 3 (WP3).

Work in WP2 includes the collection of data from existent multiple vocabularies and terminologies relevant to the Lynx goals, as well as the creation of new ones extracted from corpora of the domains covered in the project. Corpora of relevant documents are still being collected, not only for vocabulary extraction but for their later semantic annotation and integration (as part of Task 2.4), and translation corpora are also created to enable the later translation mechanisms in Lynx (as part of Task X).

The LRs for Lynx are divided into two main groups: Domain Dependent Vocabularies (DDV), which consist of terminologies (monolingual, bilingual and multilingual), as well as legal corpora of documents, directives, industry standards and norms; and Domain Independent Vocabularies (DIV), which consist of dictionaries (monolingual, bilingual and multilingual), along with morphological lists of inflected word forms.

DDV include the reuse of open terminology data, as well as the creation of terminologies from scratch. For this aim, an exhaustive search of existing DDV has been performed in a first stage. As a result, resources of potential interest for Lynx have been documented in the CKAN-based Lynx data portal (<http://data.lynx-project.eu>) to be easily accessed and updated. The main source of open terminological data to be reused in Lynx comes from the new EuroTermBank terminology repository (<http://www.eurotermbank.com/>). Resources identified here will be managed through TILDE's Terminology service, which also allows the management of private term collections. A workflow has been defined to convert resources in heterogeneous formats to RDF (Resource Description Framework), the Web of Data format, for their future inclusion in the Lynx LKG and the Linguistic Linked Open Data cloud (LLOD, <http://linguistic-lod.org/>).

DDV are also supported by the creation of Lynx use case specific terminologies. In this sense, a parallel workflow has been designed for the creation, documentation, conversion, and linking of such terminologies. The creation of new vocabularies has been semi-automatically performed starting from corpora provided by Lynx consortium members. Resulting terminological data has been transformed into RDF and linked to popular resources in the LLOD cloud.

TILDE'S Machine Translation (MT) systems, which are integrated in Lynx, can benefit from DDV by improving automated translation quality and adjusting MT systems to specific domains and use cases.

DIV include the use of KD's Lexicala API (<https://api.lexicala.com/>), which enables access to lexicographic data in RDF format, including dictionary entries with semantic and syntactic information and translation equivalents. The LRs of KD are available in JSON format, the main part of which is already available also in RDF (JSON-LD) and more such conversion of KD resources is still underway in the context of WP3.

Lynx focuses on four European languages, namely *English*, *German*, *Spanish* and *Dutch*, to best accommodate main working languages of our three use case partners (Cuatrecasas, DNV GL, openlaws) and increase the project's practicality and efficiency. Both main vocabulary providers (TILDE and KD) are well equipped with appropriate resources for these languages.

This document rounds up the interim report of D2.2 in M12 (November 2018), accounting for the relevant work that has continued and progressed in the second year of the Lynx project. This concerns mainly the addition of more DDV and the conversion of DIV data into RDF in conjunction with the W3C's new state-of-the-art *Lexicog* standard.

Following this Introduction, Section 2 presents in detail the different types of DDV and DIV. Then, Sections 3 and 4 describe the generation of new DDV and the RDF conversion of DIV, respectively. Finally, Section 5 concludes with a summary of Lynx LRs and our plans for updating or adapting new ones when required by the business cases in the context of WP3 and WP5 tasks. Such updates will be included in the new version of the Lynx Data Management Plan (DMP)..

The report is supplemented by three appendices: (1) containing a table of all the datasets harvested in the Lynx Data Portal and describing the current legal-domain resources that have been converted to RDF, (2) listing the existing DIV of KD, and (3) providing data samples of KD resources in different formats.

2 VOCABULARY IDENTIFICATION: DOMAIN DEPENDENT AND DOMAIN INDEPENDENT VOCABULARIES

The identification of DDV and DIV, as well as of corpora-derived LR, is part of the work carried out in WP2 with the aim of enabling optimal processing of multilingual data as part of Lynx services. It includes the collection of existent LR and the creation of new ones extracted from corpora in the legal domain, and their subsequent conversion to LD formats and linking with data sources in the Linked Open Data (LOD) cloud (<https://lod-cloud.net/>). In this process, the data extraction is followed by its semantic annotation and integration. In addition, translation corpora are created to facilitate the Lynx translation mechanisms. Overall, the WP2 objectives are listed as follows:

1. Provide and continuously update the Lynx Data Management Plan.
2. Provide a framework of unique identifiers for legal documents.
3. Harvest Lynx relevant data and information, including:
 - a. Domain Dependent Vocabularies (DDV);
 - b. Domain Independent Vocabularies (DIV);
 - c. Corpora of documents of legal documents, directives, industry standards and norms.
4. Create translation corpora to be used for machine translation in WP3.

The first two objectives are covered in the deliverables regarding the Data Management Plan, namely D2.1, D2.4 and D2.8. In this report, we focus on the third objective of DDV and DIV. The fourth objective is handled in D2.3 and D2.6.

2.1 DOMAIN DEPENDENT VOCABULARIES (DDV)

2.1.1 Existing DDV

The workflow of existing DDV includes the use of ready terminology services. The main resource in this sense is EuroTermBank portal (www.eurotermbank.com), a network of stakeholders for publishing and hosting EU-related open terminology data. Another platform is Tilde Terminology platform which provides services and tools (e.g. term annotation, automatic extraction) as well as a platform for managing private term collections. Thus, Lynx can benefit from EuroTermBank by getting public and open resources, as well as from Tilde Terminology by using its services (automatic term extraction, translation lookup).

EuroTermBank was an initial EU project whose purpose was to gather information related to EU-linked terminology. As the EuroTermBank project focus was on harmonization and consolidation of terminology work in EU member states, transferring experience from other EU terminology networks and accumulating competencies and efforts of the accessed countries. EuroTermBank currently provides terminology in 34 languages and wide domain coverage, and it is perfectly suitable to addressing the harvesting process of DDV in Lynx. Conversely, Lynx will also contribute to this project, since all existing DDV now have the option of being converted into RDF format, as explained in the next section.

2.1.2 Harvesting DDV

To acquire new domain vocabularies, the consortium looked at different data sources to harvest terminology resources that deal with legal vocabulary. Various paths were followed to identify and explore relevant resources for Lynx, including:

- General web search
- Lookup of resources described in papers from specialized literature
- Search in data portals specialized in language resources

The latter includes the ELRC-SHARE repository, a repository created by the European Language Resource Coordination (<http://www.lr-coordination.eu/>) to improve the quality of automated translation solutions in the EU, and used for documenting LRs in Europe; the ReTeLe catalogue created by the network of excellence for language technologies in Spain (<http://catalogo.retele.linkeddata.es/>); and the ELRA Language Resource Catalogue (<http://catalogue.elra.info/en-us/>), a unified portal for LRs.

The main focus was in identifying vocabularies containing terms of the legal domain, but some other resources that gather terms of related domains (social sciences, economics, for instance) were also taken into account. Some resources were already published as Linked Data as part of the LLOD cloud. These are especially interesting for the generation of what we have termed in D2.1 the Linguistic *Legal* Linked Open Data cloud, that is, a sub-cloud of linguistic resources in the legal domain that are already a tangible result of the Lynx project.

In this sense, it is worth mentioning TILDE's E-Translation Termbank (eTTB, <https://www.tilde.com/news/european-commission-entrusts-tilde-ettranslation-integration>), one of TILDE's successfully completed terminology-acquiring projects, which provides now over 150 new terminological resources to the language resource pipeline of CEF eTranslation (<https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eTranslation>), covering all official languages of the EU plus Icelandic and Norwegian for three sector-specific domains – health, legal (business legislation) and customer protection. The eTTB project ended in February 2019, so the resources gathered during this project were made available also for Lynx using the EuroTermBank platform.

Within the eTTB project, a new workflow was created to harvest new resources, which has also been adopted for Lynx. The workflow is illustrated in Figure 1 and is explained below. Basically, it allows resources in heterogeneous format to be analysed with regard to IPR, cleared and converted into TBX and RDF. The RDF conversion is done by integrating TBX to RDF conversion service directly in EuroTermBank platform allowing to get the most actual RDF version of any interesting terminology.

Some data come in a structured form (XML or other easy-to-parse formats), whereas others are not fully organized and need to undergo a structuration process before the extraction of terms is possible. Best practice is to collect and provide terminology resources in already **structured formats**, e.g. TBX, XLS, XLSX, TMX, CSV, TSV. Some new terminology resources are available only via Application Program Interface (API), and such requests are also considered. Terminology resources in **semi-structured** and **unstructured formats**, e.g. DOC, DOCX, TXT, PDF, HTML, are also considered and undergo a three-stage conversion process that involves being converted to TSV, then to TBX, and finally to RDF. Finally, they may be enriched with further linguistic data and translation equivalents before being linked to other resources in the Lynx Legal Knowledge Graph (LKG).

In the following we refer to the general processes followed for the different types of resources depending on the input format, that is, structured, semi-structured and unstructured format.

- For **structured formats**, e.g. XML, we use scripts, which allow easy and fast processing of structured format and are customised according to the required processing of content. Some formats can be processed manually first, e.g. XLSX might have columns for different terminological data categories, which should be defined according to the TBX standard and can then be converted automatically into TBX or saved as CSV or TSV.
- To process **semi-structured formats**, e.g. DOCX, we use advanced scripts developed in Microsoft PowerShell and customised according to the necessary processing of semi-structured content. Terminological data in semi-structured formats need additional, sometimes considerable, human processing, e.g. terms might be formatted in bold and/or italic within term entries or separated from other terminological data, e.g. definitions, in another paragraph, or might be given within

the same paragraph and separated with hyphens or colons. In such case, terminological data should be structured first with advanced parsers and human processing.

- The most challenging and resource-consuming part of the extraction task is to process **unstructured terminological data**. To process terminology resources in unstructured formats, e.g. PDF containing scanned content, optical character recognition (OCR) is used first. Then additional processing – human or machine – is needed. The quality of the OCR result depends on the software and human knowledge. For this task professional OCR software is being used, Adobe Acrobat Pro, which is the most reliable programme nowadays, also used for image processing, e.g. JPG and PNG. On top of machine processing, human processing is required to validate the result given by the OCR, e.g. to correct unrecognised characters and spelling errors, structure content and apply appropriate formatting. For this task, linguistic knowledge is needed. Applying linguistic knowledge might take vast amounts of time (even several days) to process a single document.

DOCX format saved in PDF is less challenging to process and structure, as it is semi-structured and therefore requires less human processing and linguistic knowledge. However, it might contain errors, e.g. due to encoding issues or characters that are easy to process by humans but are difficult for machines. Such content should be validated by linguistic experts.

Figure 1 presents the full workflow of processing different data formats.

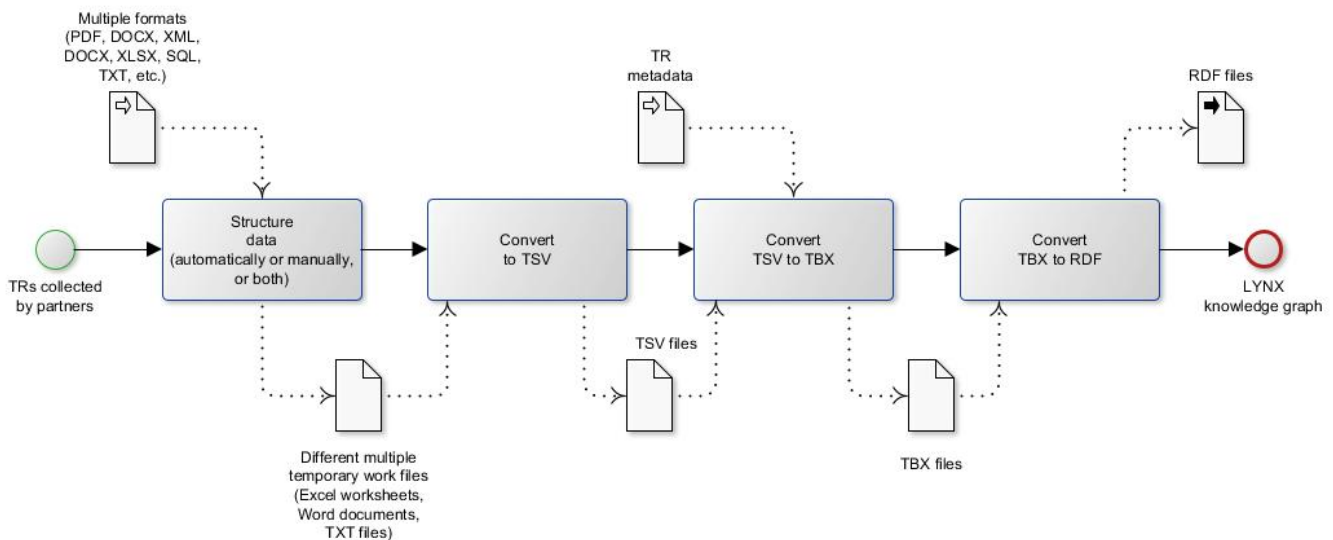


Figure 1. Workflow of processing different data formats

A separate part of the workflow is devoted to finding out IPRs and describing those where the status of intellectual property rights is not clear or where the art work is copyright protected (not under Creative Commons conditions).

Altogether, consortium partners have identified up to 1200 sources as potential DDVs. More than 200 resources have clear licencing status and data structure, and are thus processed into TBX. Around 70 of these have already been processed and registered to the public Lynx data portal. Appendix 1 gathers all the datasets collected during this harvesting stage.

During this process, other potentially useful resources have been identified, but they have not been archived since they are either published in formats that cannot be handled, are no longer available, or are not supported by their publishers.

Following this resource harvesting stage, some conclusions about the status of the legal vocabularies can be drawn:

- Legal language resources are often generated for one single sub-domain of law (e.g. economic law) so they may not directly cover the domains we are interested in for Lynx.
- Usually, vocabularies represent terms in major languages, which make it more complex to find resources for other languages.
- Datasets are heterogeneous in terms of format, language and lexical information.
- Resources generated by small organisations or individuals are often in obsolete formats that are no longer available or useful, and thus must be converted to an up-to-date format.

In addition, we have also performed an analysis of the data resources harvested, evaluating their usefulness taking into account different parameters:

- Language: giving priority to resources in Dutch, English, German and Spanish
- Format: giving priority to resources in RDF and machine readable formats
- Domain: giving priority to resources from the legal domain

After discarding the great part of the resources in the Data Portal as per the previous parameters, we have analysed the remaining 23%, considering useful only those resources with open licences (see Table 1).

| <i>Resource Name</i> | <i>Licence Name</i> | <i>Open?</i> | <i>Comments</i> |
|--|--|--------------|-----------------|
| e-Compliance multilingual thesaurus | - | - | Unreachable |
| DGT Acquis | EUPL | Yes | Corpus |
| Court Thesaurus | - | - | Unreachable |
| BabelNet | Non Commercial | - | |
| JRC Acquis | EUPL | Yes | Corpus |
| ECB corpus | - | - | Unreachable |
| Jurivoc | - | - | Unreachable |
| EuroVoc | EUPL | - | |
| UNESCO Thesaurus | CC 3.0 | Yes | |
| EurLex | Non Commercial | Yes | Corpus |
| EUGO Glossary | Non Commercial | Yes | |
| Termcat Glossaries | - | Yes | |
| Copyright Termbank | CC BY | Yes | |
| IATE | Open for commercial and non commercial | Yes | |
| STW Thesaurus for economics | OBDL | Yes | |
| TheSoz | CC BY 2.0 | Yes | |
| Finnish Legislation | CC BY 2.0 | Yes | LKG |
| EU Cases | - | - | Unreachable |
| Connecticut Legal Glossary | Under Request | | |

| | | | |
|---|----------------|-----|-------------|
| SAIJ Legal Thesaurus | ODbL | Yes | |
| EUGDPR Glossary | - | - | Unreachable |
| Glossary of Pension Terms for Consumers | CC-BY-4.0 | Yes | |
| Glossary "Austrian Administration. German - English" | open Under-PSI | Yes | |

Table 1. Data Portal Licence Analysis

This means that we could safely reuse 3 corpora and 14 terminological resources: corpora can be annotated and indexed in the LKG; and terminological resources can be used to enrich our own glossaries and continue the generation of the Legal Linguistic Linked Open Data cloud.

These reasons have impeded us to generate more legal vocabularies to ease the annotation, classification and translation activities that are required in the Lynx project. The creation “from scratch” of DDV that cover the terminological needs of use cases has been reported in section 3.

2.2 DOMAIN INDEPENDENT VOCABULARIES (DIV)

General language resources that are not specifically oriented to the Lynx domain (and thus not part of DDV) form the domain independent vocabularies (DIV) and complement DDV in the aim of offering wholesome linguistic information. To begin with they derive from existing resources of KD, and they are due to expand in the context of WP3 through linking to open resources that will be harvested on the Web, primarily from the Linguistic Linked Open Data cloud.

2.2.1 Existing DIV

The KD data sets cover the four principle languages of Lynx – Dutch (NL), English (EN), German (DE) and Spanish (ES) – and extend to others too. They are based on deep lexicographic analysis and mapping of each language, and include sense disambiguation, translation, morphological inflection, and other features and information.

These resources are all available in XML, and to various extents in RDF, JSON and JSON-LD formats. The model chosen to represent the LD version of such data was lemon-Ontolex, with RDF serialisation in JSON-LD, and, particularly, its *lexicog* module (<https://www.w3.org/2019/09/lexicog/>). In the following, we describe the different families of KD data sets that serve as DIV in Lynx.

(a) Global series

A monolingual lexical core of each of the languages involved, with translation equivalents in some of the other languages, and options to link to each other and to more languages. It includes semantic and syntactic details, representation of pronunciation and alternative scripts, inflected forms, grammatical categorization, sense division, attributes and definitions, phrases and examples, cross-references, frequency and other information, as well as translation equivalents of word meanings, multiword units and examples of usage. Every language core thus constitutes a multilayer crosslingual lexical network that can also link to others. The Global sets are originally in XML format and are nearly fully converted to RDF and JSON-LD in line with the Ontolex-*lemon* model. The main figures per language are listed in Table 1, including division into different levels according to the size of the core of each language.

| language | level | words & phrases | senses | examples of usage | translated to |
|--------------|-------|-----------------|--------|-------------------|--------------------------|
| Dutch (NL) | 1 | 17,000 | 21,000 | 21,000 | English, German, Spanish |
| | 2 | 31,000 | 36,000 | 37,000 | English, German |
| English (EN) | 1 | 17,000 | 27,000 | 30,000 | Spanish |
| German (DE) | 1 | 19,000 | 30,000 | 32,000 | English, Dutch |
| | 2 | 35,000 | 50,000 | 57,000 | English, Dutch |
| Spanish (ES) | 1 | 15,000 | 28,000 | 57,000 | English, Dutch |
| | 2 | 28,000 | 47,000 | 75,000 | English, Dutch |
| | 3 | 57,000 | 81,000 | 107,000 | |

Table 2. Basic statistics per language in the Global series

Figure 2 presents the existing Global dictionary monolingual cores (in green) and their bilingual pairs.

| | DE | EN | ES | NL |
|----|----|----|----|----|
| DE | ● | ● | | ● |
| EN | | ● | ● | |
| ES | | ● | ● | ● |
| NL | ● | ● | ● | ● |

Figure 2. Global language cores and pairs

(b) Password and MultiGloss

The Password multilingual set of semi-bilingual English learner's dictionaries includes nearly fifty language versions, consisting of an English-English core with full dictionary entries including brief translations for each sense, so the Dutch, German and Spanish equivalents can be juxtaposed against the English term.

Dutch, German and Spanish have their own lists of simple entries consisting of words and phrases that are linked to the corresponding Password English entries and senses (excluding examples of usage). This forms human edited bilingual glossaries that are multilingualised automatically by associating the translations from the other languages to that sense of the Password entry.

The data are in XML format. Main figures per language are shown in Table 2.

| language | source | words & phrases | senses | examples of usage | translated to |
|----------|------------|-----------------|--------|-------------------|----------------------------------|
| English | Password | 29,000 | 39,000 | 37,000 | Dutch, German, Spanish |
| Dutch | MultiGloss | 21,000 | 35,000 | — | English > German, Spanish |
| German | MultiGloss | 25,000 | 39,000 | — | English > Dutch, Spanish |
| Spanish | MultiGloss | 22,000 | 45,000 | — | English > Dutch, German, Italian |

Table 3. Basic statistics per language in Password & MultiGloss

Figure 3 shows the monolingual cores (in green), manually created pairs (blue for English>L2 and yellow for L2>English) and auto-generated pairs (orange).

| | DE | EN | ES | NL |
|----|----|----|----|----|
| DE | | | | |
| EN | | | | |
| ES | | | | |
| NL | | | | |

Figure 3. Password & MultiGloss cores and pairs

(c) Random House Webster’s College Dictionary

A comprehensive English monolingual dictionary that was originally conceived for American university students and English native general users, including over 130,000 entries and 190,000 senses. The data is in XML format.

(d) Wordforms

Morphology lists of inflected forms linked to their lemmas, the figures are shown in Table 3.

| language | lemmas | inflected forms |
|----------|---------|-----------------|
| Dutch | 157,000 | 206,000 |
| English | 69,000 | 160,000 |
| German | 95,000 | 456,000 |
| Spanish | 125,000 | 499,000 |

Table 4. Numbers of lemmas & inflected forms

KD data samples for DIV are included in Appendix 2.

2.2.2 Harvesting DIV

External DIV resources are planned to be harvested mainly from the Linguistic Linked Open Data cloud (LLOD, <http://linguistic-lod.org/>). The major LLOD resources for this purpose are BabelNet and DBpedia. Initial experiments are carried out to link the RDF version of the Global KD data set to such external sources. However, the final outcomes of this activity will ultimately depend on the needs of such enriched data by the Lynx services (WP3) and on its use by the project’s pilots (WP5), and will be reported in the corresponding deliverables.

3 NEW DDV GENERATION

3.1 CREATING VOCABULARIES

In previous stages of the project, we followed a semi-automatic approach for the generation of new DDV. This preliminary semi-automatic approach, proposed in D2.2, involved a lot of manual effort (managing terms, formats, etc.), and has now evolved into an automatic methodology for the creation of legal linked terminologies, as described in Figure 4.

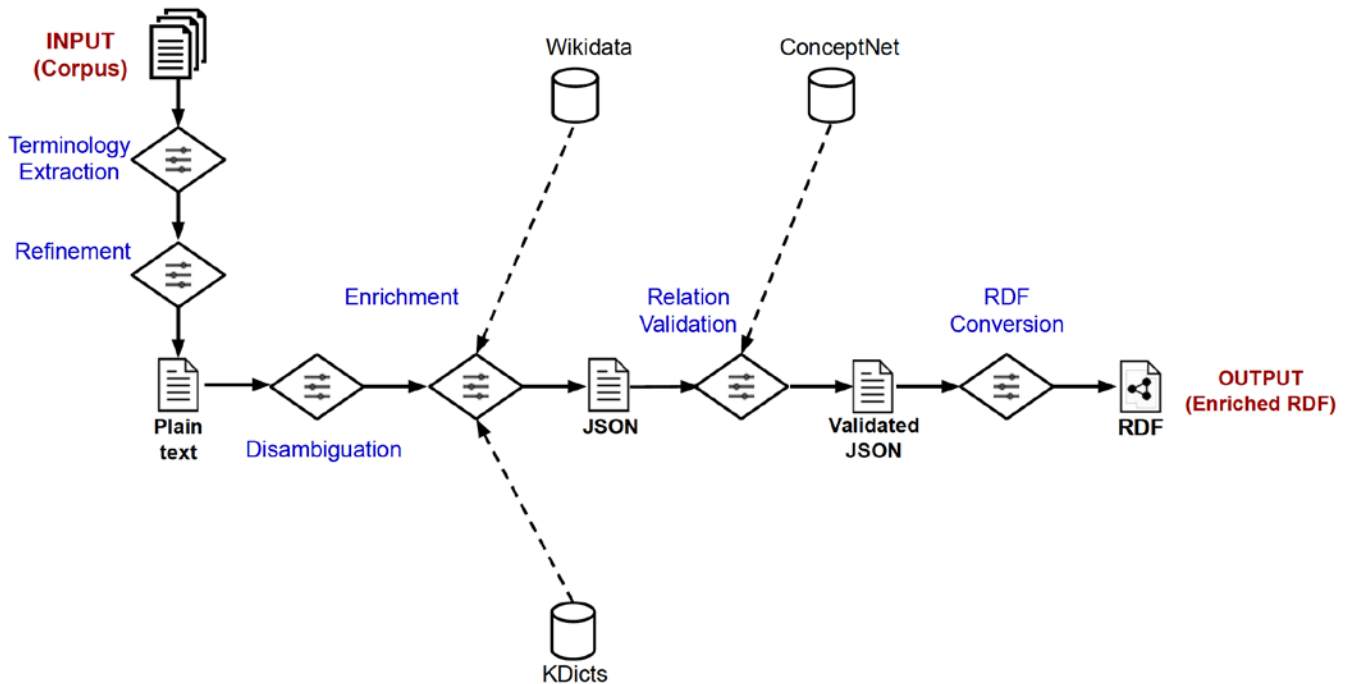


Figure 4. Automatic Generation of Legal Linked Terminologies

The several tasks in this methodology are spelled out in the following:

Task 1: Terminology Extraction. The base input is a domain specific corpus. This corpus is processed through TILDE’s Terminology Extraction service (TermEx). The output of the task is a list of domain specific terms and the contexts in which those terms appear (to be used in a later disambiguation stage).

Task 2: Refinement. The output of the previous task is a list of terms with duplicates, inconsistencies, or wrongly identified terms that needs to be reviewed before it is used in the next task. For this reason, we propose a series of refinement suggestions that include lemmatization, removing wrongly identified terms (such as, “appropriate resolution”, “different branch”), removing duplicates, unifying caps, creating top concepts (such as “business” as the top concept of “business partner”, “business unit”, etc.) and removing Named Entities (such as “Ms Robertson” or even “Ikea”). These issues are automatically corrected applying diverse NLP techniques, regular expressions and linguistic patterns.

Tasks 2 and 3: Disambiguation and Enrichment. After the refinement task, the next step is to enrich the refined terms with additional information by querying available external knowledge bases that contain the terms. For the time being, the enrichment task has been achieved by querying Wikidata, an open knowledge base that contains thesaurus-like linguistic information related to concepts, that is, broader, narrower and related terms, definitions and translations. To correctly map source term to target term in

Wikidata, a disambiguation process is needed. For this purpose, the Word Sense Induction and Disambiguation service developed by Semantic Web Company (SWC) is used.

Task 4: Relation Validation. When accessing Wikidata as external knowledge base, we noticed many issues concerning the data collected under the *also known as* property. These data should be aliases (spelling variants, scientific names and nicknames) and should be categorised as *synonyms* of the source term. However, in many occasions we found broader, narrower and related terms contained under this property, so we have developed a series of axioms to verify each type. In this step, we also rely on a Linguistic Knowledge Base as a *synonym goldstandard*, to verify that this kind of retrieved content is correct.

Task 5: RDF Conversion: Considering the content of the terminologies and their objective (annotate legal documents), we decided to use SKOS (Simple Knowledge Organization System) as it seemed an appropriate model due to its simplicity and intuitive representation. In our new approach, we continue applying this vocabulary by using a simple Python script and a SKOS template to transform the intermediate JSON data into RDF (**Figure 5**)

```
<http://lkg.lynx-project.eu/kos/labourlaw/terms/SCTMID> a skos:Concept;
skos:closeMatch <https://www.wikidata.org/wiki/WDTMID>;
skos:prefLabel "TERM"@LANG;
skos:altLabel "ALTERM"@LANG;
skos:description "DESC"@LANG;

skos:broader <https://www.wikidata.org/wiki/WDBRTMID>;
skos:narrower <https://www.wikidata.org/wiki/WDNRTMID>;
skos:related <https://www.wikidata.org/wiki/WDRLTMID>.

<http://lkg.lynx-project.eu/kos/labourlaw/LTBRTMID> a skos:Concept;
skos:prefLabel "BRTERM"@LANG;
skos:exactMatch <https://www.wikidata.org/wiki/WDBRTMID>.

<http://lkg.lynx-project.eu/kos/labourlaw/LTNRTMID> a skos:Concept;
skos:prefLabel "NRTERM"@LANG;
skos:exactMatch <https://www.wikidata.org/wiki/WDNRTMID>.

<http://lkg.lynx-project.eu/kos/labourlaw/LTRLTMID> a skos:Concept;
skos:prefLabel "RLTERM"@LANG;
skos:exactMatch <https://www.wikidata.org/wiki/WDRLTMID>.
```

Figure 6. SKOS template applied.

Our most recent efforts were devoted to generate a script to automatically link resources in RDF by means of the `skos:closeMatch` property. As mentioned before, we can now link our terminologies with Wikidata, although the same approach can be used with other knowledge bases. The following figure is a graphical representation of a linked terminological entry.

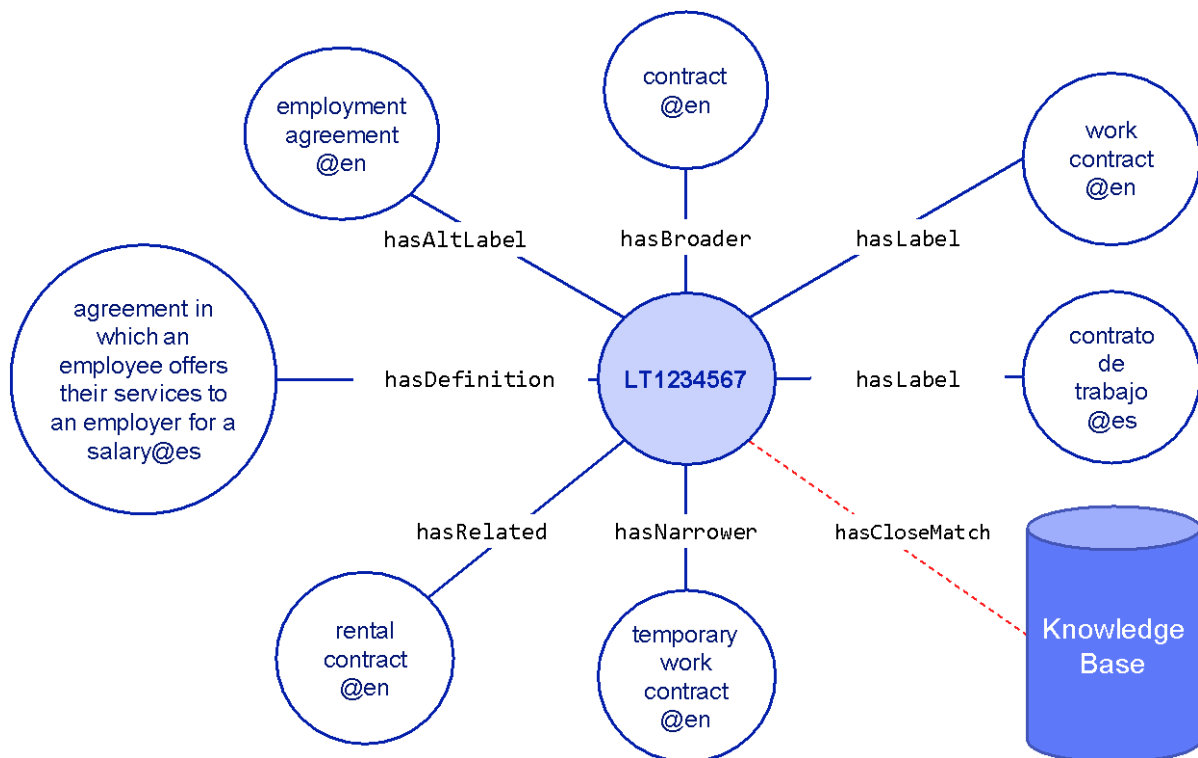


Figure 7. Sample of a linked terminological entry

3.2 DOCUMENTING VOCABULARIES

An issue spotted during the harvesting of terminological resources is the fact that the greatest part of existing resources is not properly documented. This means that they are difficult to find, handle and update. To solve this situation, a common repository has been created through a CKAN platform, where the metadata of the harvested and created resources are collected. The platform can be accessed through this URL: <http://data.lynx-project.eu/>.

The resources are described according to a set of metadata descriptors that have been collected in two main blocks: information about the dataset and information about the resource. This distinction is quite significant: within this context, *dataset* makes reference to the whole asset, while *resource* defines each of the different formats in which the dataset is published. (An example of a terminological record of this portal appears in Section 3.1.2 of D2.1.)

While at the time that D2.1 was generated, 26 datasets were collected in CKAN, now there are 65 datasets documented in the portal. All the partners have contributed to their identification and documentation. Most of them are already in RDF format, but many others still need to be transformed in order to be used by semantic web technologies.

4 XML TO RDF CONVERSION OF DIV

4.1 INTRODUCTION

As pointed out in Section 2, the Global series of KD was chosen for its representation in RDF in the context of Lynx. The motivation is to move towards well-structured, detailed and extensive lexicographic data rather than single dictionary products, by reinforcing and further expanding existing data, and improving interoperability between content from the Global series and other multilingual data on the Web, attaining reciprocal enrichment of the Global series by external resources. The new data will enable richer and with better coverage results when used as basis of the WP3 Lynx services (e.g, for the purposes of semantic annotation, entity extraction, or word sense disambiguation, for instance).

The conversion of such dictionary data into RDF has led to several advantages:

- Reuse of URIs across dictionaries, which leads to a **less redundant model**, less prone to inconsistencies
- Enhanced access to data: since every element is treated as “first-class citizen” in the RDF graph (i.e., not only dictionary entries but their synonyms, translations, etc.), information is more **easily accessible** through simple SPARQL queries
- Use of Semantic Web standards, thus making data ready to be **linked** to and **consumed** by other LD resources
- Use of standard models and access means, which leads to an enhanced **interoperability**

In the following pictures, we illustrate how the initial dictionaries, originally built in a disconnected manner to each other, are connected in a single unified graph in the RDF version of KD Global.

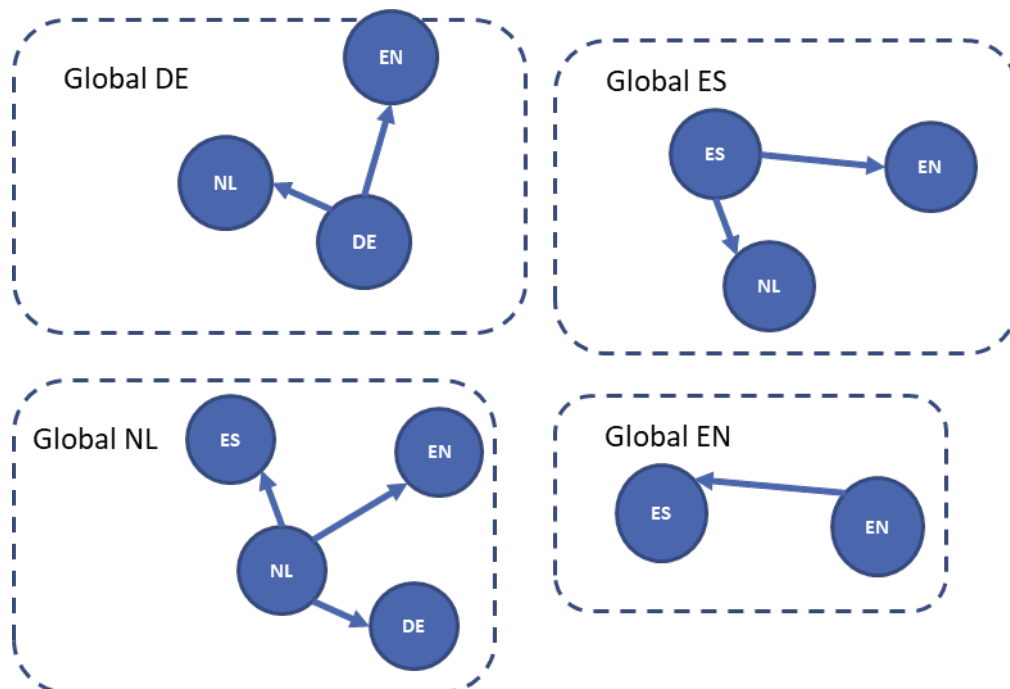


Figure 7. Language pairs in the original, isolated, dictionaries. Arrows represent the directionality of the translations

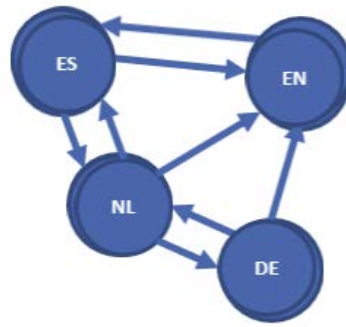


Figure 8. Language pairs in the resulting RDF graph

4.2 REPRESENTATION MODEL

In order to represent the lexicographic content in KD, the *'lexicog'* module of Ontolex-lemon has been chosen. The model was recently released by the W3C Ontology Lexica community group (<https://www.w3.org/2019/09/lexicog/>) as an extension of the Ontolex-lemon specification tailored to lexicographic data. Several members of the Lynx project have actively contributed to the definition of *lexicog*. Actually, the Global KD data representation in RDF constitutes a pioneering experience in validating and adopting the *lexicog* module.

Lexicog aims at capturing the underlying original structure and annotations of the lexicographic entry in a way that keeps the purely lexical content separated from the lexicographic one, minimizing information loss and allowing queries restricted to the lexical layer. By being able to keep record of the original dictionary arrangement as RDF, the module does not impose a certain view on the lexicon and thus becomes agnostic to the standpoint of the lexicographer. For that purpose, new ontology elements have been added that reflect the dictionary structure (e.g., sense ordering, entry hierarchies, etc.) and complement the OntoLex-lemon lexicon. Figure 16 describes the main classes and properties of the module.

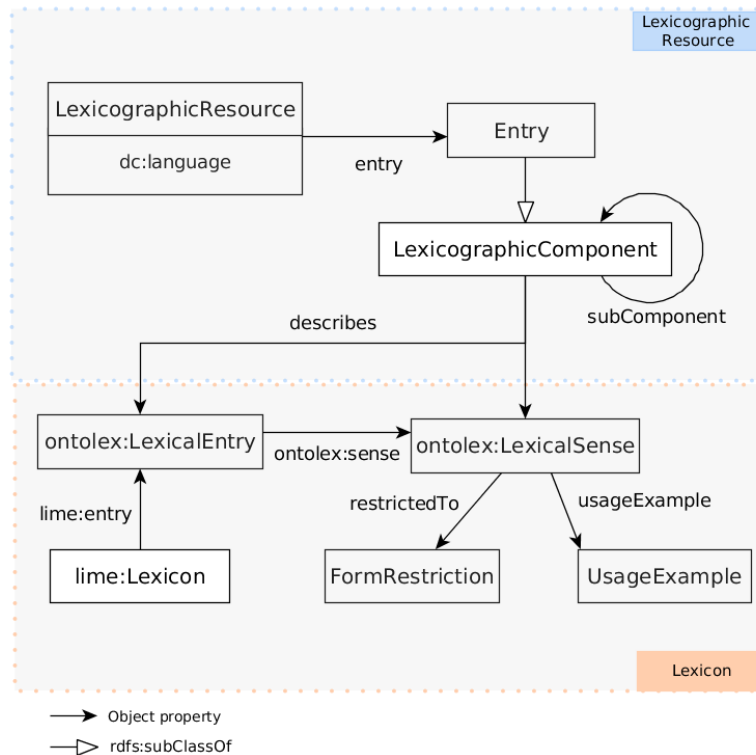


Figure 9. Scheme of the lexicography module (taken from the “OntoLex-lemon Lexicography Module” W3C community group final report).

We refer to the specification document for more details, but we give here an overview of its main modelling ingredients:

- *LexicographicResource*, which represents a collection of lexicographic entries in accord with the lexicographic criteria followed in the development of that resource.
- *Entry*, a structural element that represents a lexicographic article or record as it is arranged in a source lexicographic resource.
- *LexicographicComponent*, which is a structural element aimed at representing the (sub)structures of lexicographic articles providing information about entries, senses or subentries. Lexicographic components can be arranged in a specific order and/or hierarchy.

The lexicographic components only reflect the structure of the dictionary and do not encode any lexical content themselves. To associate them to their corresponding lexical information (e.g. lexical entries or lexical senses), the property “describes” is used. Such elements belonging to a lexicon are taken from OntoLex, in particular:

- *ontolex:LexicalEntry*, which consists of a set of forms that are grammatically related and a set of base meanings that are associated with all of these forms.
- *ontolex:LexicalSense*, which represents the lexical meaning of a lexical entry when interpreted as referring to the corresponding ontology element.
- *lime:Lexicon*, or a collection of lexical entries for a particular language or domain.

4.3 METHODOLOGY

The conversion methodology, based on well-established methodologies for multilingual linked data generation¹, is an iterative approach in which every iteration will convert a portion of the data model, from simpler to more complex data, following these steps:

1. Identify a subset of elements to convert. A manually created example of an RDF entry with real KD data was created for every element.
2. Define the XML – RDF mapping table for the selected elements.
3. Define a URI naming strategy for the chosen elements. Such strategy was constantly validated and revised in every iteration.
4. Implement and run the corresponding conversion scripts. The manually constructed example had a vital part in determining the conversion script.
5. Validate the resulting RDF data. The method of validation selected to this end is twofold, consisting of the JSON Schema as an initial means of validation, and a SPARQL endpoint and query service for querying the RDF output.
6. Repeat.

The RDF conversion pipeline relied on an already existing conversion of XML data into JSON, adding LD elements and restructuring the JSON document to comply with the triple relations encompassed in the JSON-LD structure. In each iteration, the conversion was applied to all of the resources of the Global series, resulting in a collection of JSON-LD documents, with each dictionary entry represented by its own JSON document and reflecting an RDF graph introducing only the components that were the focus of the current iteration, on top of previously covered components. Visually, the conversion pipeline went through these different formats:



Figure 10. Conversion flow of data format from XML to JSON to JSON-LD to enabling SPARQL queries

The result of the whole process was a unified graph of cross-lingual linked KD resources, which is serialized in JSON-LD and accessible through a SPARQL endpoint.

This is an example of a simple SPARQL query to retrieve all the lexical entries that evoke the concept of "derecho laboral" (the Spanish expression for "labour law"):

```

PREFIX lexicog: <http://www.w3.org/ns/lemon/lexicog#>
PREFIX ontolex: <http://www.w3.org/ns/lemon/ontolex#>
PREFIX vartrans: <http://www.w3.org/ns/lemon/vartrans#>
PREFIX lime: <http://www.w3.org/ns/lemon/lime#>
PREFIX lexinfo: <http://www.lexinfo.net/ontology/2.0/lexinfo#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX skos: <http://www.w3.org/2004/02/skos#>
  
```

¹ See D. Vila-Suero, A. Gómez-Pérez, E. Montiel-Ponsoda, J. Gracia, and G. Aguado-de Cea, "Publishing Linked Data: the multilingual dimension," in *Towards the Multilingual Semantic Web*, P. Cimiano and P. Buitelaar, Eds. Springer Berlin Heidelberg, 2014, pp. 101–118.

```

SELECT DISTINCT ?entry ?sense ?def {
  ?concept a ontolex:LexicalConcept;
    skos:definition ?def .
  ?sense ontolex:isLexicalizedSenseOf ?concept .
  ?entry ontolex:sense ?sense .
  FILTER regex (?def, "derecho laboral")
}
    
```

This is the result of such a query (a fragment):

Note: the links can be accessed only with a lexicala API account.

| | entry | sense | def |
|---|---|---|--|
| 1 | http://lexicala.com/id/global/LexiconBR/trabalhista-adj | http://lexicala.com/id/global/LexiconBR/trabalhista-adj-laboralista-adj-SE00044624-sense | "que se especializa en derecho laboral"@es |
| 2 | http://lexicala.com/id/global/LexiconEN/labor_relations-adj | http://lexicala.com/id/global/LexiconEN/labor_relations-adj-laboralista-adj-SE00044624-sense | "que se especializa en derecho laboral"@es |
| 3 | http://lexicala.com/id/global/LexiconES/laboralista-adj | http://lexicala.com/id/global/LexiconES/laboralista-adj-SE00044624-sense | "que se especializa en derecho laboral"@es |
| 4 | http://lexicala.com/id/global/LexiconBR/trabalhista-n | http://lexicala.com/id/global/LexiconBR/trabalhista-n-laboralista-n-SE00044625-sense | "abogado que se especializa en derecho laboral"@es |
| 5 | http://lexicala.com/id/global/LexiconEN/labor_relations_lawyer-n | http://lexicala.com/id/global/LexiconEN/labor_relations_lawyer-n-laboralista-n-SE00044625-sense | "abogado que se especializa en derecho laboral"@es |

A more complete technical description of the Global KD series conversion process can be found in the following publication:

J. Bosque-Gil, D. Lonke, J. Gracia and I. Kernerman, "Validating the OntoLex-lemon Lexicography Module with K Dictionaries' Multilingual Data," in Proceedings of 6th biennial conference on electronic lexicography, eLex 2019, 2019, pp. 726–746.

More details of the conversion to JSON-LD of DIV can be found in Appendix 4.

5 SUMMARY AND OUTLOOK

The main vocabulary achievements in the first two years of the Lynx project can be summarized in the processes of identifying the relevant resources, preparing to collect those already available from the consortium partners and for their smooth delivery to suit the LKG, and planning further linked data resource enrichment that will continue on a regular basis during the project's last year and thereafter.

With the aim of reinforcing, enhancing and improving the legal knowledge services to be provided by Lynx, and in accordance with the main languages practiced by the three use case partners, it has been decided to replace Italian by Dutch, in addition to English, French and German. The Lynx partners are thus able to assure a more solid and efficient linguistic infrastructure for the LKG regarding both general language and legal terminology.

Terminology resources gathered and selected for the Lynx project are available via the EuroTermBank portal and the Lynx data portal. The primary location for those resources harvested in synergy with the eTTB project is EuroTermBank portal, which enables the management of terms, facilitates licencing, provides search and export functionalities, and ensures all terminology in a single format. While the Lynx data portal CKAN acts as an electronic catalogue for related resources, it can also lead to external platforms and contain any data and in any format. The Lynx data portal CKAN functions as a guide for new use cases and where to get Lynx domain dependent terminologies, as well as to document terminologies that are used by Lynx services.

At the end of this task, the consortium has documented all Lynx related terminology resources in the CKAN official data portal. The Lynx data portal CKAN will be public and available for everyone. Thus, the documentation of all Lynx-related resources will be part of the future work as well continuing to gather new existing resources.

With regard to data linking tools, in the third year of the project, other tools will be considered for application, such as VocBench, maintained by the Publications Office of the European Union, or PoolParty, developed by Lynx partner SWC.

Appendix 1. Relevant language resources documented in Lynx Data Portal

| name | title | domain | publisher | jurisdiction | language | notes |
|---------------------------|--|-------------------|--|----------------|----------------|--|
| abc-of-oil | ABC of oil | industry | Norwegian Petroleum Directorate | Norway | en, no | ABC of oil contains terminology and abbreviations used in oil industry. |
| agrovoc | Agrovoc | | | | 29 languages | Controlled general vocabulary |
| babelnet | BabelNet | General | | | 270 languages | BabelNet is regarded as a multilingual encyclopaedic dictionary but also as a semantic network and a knowledge base that combines general data with lexical information that comes from WordNet. |
| biocides | Biocides | agriculture | Health and Safety Executive | European Union | en | This glossary contains terminology related to biocides and their regulation. |
| bpr | BPR - Bibliography of the Italian Parliament | Legal | Italian Parliament | | it | The BPR - Bibliografia del Parlamento italiano e degli studi elettorali (Bibliography of the Italian Parliament and Electoral Studies) is a database of bibliographic references of books and articles in periodical journals addressing the history of the Italian Parliament and the history of elections. |
| calathe | Cadastral Vocabulary (CaLaThe) | Environmental Law | University of Denmark and University of Turkey | | en | Monolingual thesaurus for the domain of cadastre and land administration that provides a controlled vocabulary. It contains 142 terms related to previous vocabularies, such as AGROVOC, GEMET and STW thesaurus for economics. Terms are arranged in graphical, tabular and alphabetical order. |
| cdisc-glossary | CDISC Glossary | Clinical Data | CDISC | | en | CDISC develops data standards to streamline clinical research and enable connections to healthcare. This glossary contains definitions of terms and abbreviations. It is published as PDF and XLS formats, and it has also been converted into RDF. |
| conneticut-legal-glossary | Conneticut Legal Glossary | Legal | State of Conneticut Judicial Branch | | en, es | Bilingual glossary from English into Spanish organised in alphabetical order that can be downloaded in PDF. It is published by the Conneticut Judicial Branch. Terms contained in this glossary cover general law area, including labour law and economic terminology. |
| copyrighttermbank | Copyright Termbank | Copyright | OEG | European Union | en, es, fr, pt | A multilingual termbank that contains copyright-related terms from WIPO definitions, IATE and Creative Commons licenses. This termbank is also connected to external resources such as DBpedia and Lexvo. Terms have been hierarchically organised, and they are useful to annotate licenses. |
| court-thesaurus | Court Thesaurus | legal | Wolters Kluwer | | de | A monolingual thesaurus in German containing names of German and international courts. |

| | | | | | | |
|------------------------|-------------------------------|--|--------------------------------|----------------|--|--|
| dataprotectionglossary | Data Protection Lynx Glossary | Legal | OEG | | en | Monolingual termlist from the Data Protection domain in RDF linked with BabelNet, DBpedia and EuroVoc. |
| dbpedia | DBpedia | | | | >100 languages | Extensive general knowledge base providing information about approx. 4 million entities in more than one hundred languages. |
| ecb | European Central Bank Corpus | financial, legal | OPUS, the open parallel corpus | European Union | cs, da, de, el, en, es, et, fi, fr, hu, it, lt, lv, mt, nl, pl, pt, sk, sl | ECB Corpus is a multilingual corpus that contains financial vocabulary. It has been extracted from the website and documentation of the European Central Bank, and it is aligned among 19 European languages. |
| edp | European Data Portal | Several topics including: government, justice and legal fields | European Commission | European Union | bg, cs, da, de, el, en, es, et, fi, fr, ga, hr, hu, it, lt, lv, mt, nl, pl, pt, ro, sk, sl, sv | The European Data Portal harvests the metadata of Public Sector Information available on public data portals across European countries. Not only linguistic information is presented here, but any kind of data related to diverse domains: environment, government, law, etc. Information regarding the provision of data and the benefits of re-using data is also included. Also, datasets can be filtered by country, language, format, domain, etc. |
| eige | EIGE Glossary | Gender Equality | EIGE | | en | Glossary of gender-related terms in English published by the European Institute of Gender Equality. It contains over 400 terms in English that have been extracted from 92 resources. Each entry includes a link to its related source. |
| eucases | EUCases | | | eu, uk, bg | en, bg, de | This is the result of the EUCases Research Project which developed a European case law Linking Platform, transforming multilingual legal open data into linked open data. The EUCases Linking Platform links EU law with legislative acts and court decisions of six EU member states: Austria, Bulgaria, France, Germany, Italy and United Kingdom. |
| eugdpr-glossary | EUGDPR Glossary | Legal | EUGDPR | European Union | en | A Glossary of Terms and Definitions as used in relation to the GDPR in the EU. |
| eugo | EUGO Glossary | Business | EUGO | Spain | es | The glossary of EUGO is addressed to companies and entrepreneurs that need to comply with administrative or professional requirements to perform a remunerated economic activity in Spain. |
| eurlex | EUR-Lex | Law | Publications Office | European Union | bg, cs, da, de, el, en, es, et, fi, fr, ga, hr, hu, it, lt, lv, mt, nl, pl, pt, ro, sk, sl, sv | EUR-Lex gives access to EU Law, the jurisprudence of the EU Court of Justice, other EU public documents and the electronic edition of the Official Journal of the EU, in 24 languages. |

| | | | | | | |
|------------------------------------|--|--|---|----------------|--|--|
| eurovoc | EuroVoc Thesaurus | Politics, international relations, european communities, law | European Union | n.a. | bg, cs, da, de, el, en, es, et, fi, fr, hr, hu, it, lt, lv, mt, nl, pl, pt, ro, sk, sl, sv | EuroVoc is a multilingual, multidisciplinary thesaurus that covers the activities of the EU. It is available in 23 languages of the European Union. It is intended to be used as a disambiguation tool by contextualising each term, offering univocal meanings. Also, the tool provides related terms and preferred and non preferred designations to guide the user. |
| eurovoc-agrovoc-semantic-alignment | EuroVoc-AgroVoc Semantic Alignment | Agricultural, General | Publications Office | | en | Semantic Alignment of AgroVoc Thesaurus with EuroVoc Thesaurus. |
| eurovoc-eige | EuroVoc-EIGE Semantic Alignment | Gender Equality, Legal | Publications Office | European Union | en | Semantic Alignment of Gender Equality glossary with EuroVoc Thesaurus. |
| eurovoc-gemet-semantic-alignment | EuroVoc-GEMET Semantic Alignment | Environmental | Publications Office | | Several languages including: en, de, es, it | Semantic Alignment of GEMET Thesaurus with EuroVoc Thesaurus. |
| eurovoc-inspire-alignment | EuroVoc-Inspire Alignment | General | Publications Office | European Union | en | Semantic Alignment of Inspire glossary with EuroVoc Thesaurus. |
| eurovoc-lcsh-semantic-alignment | EuroVoc-LCSH Semantic Alignment | General | Publications Office | | en | Semantic Alignment of the Library of Congress Subject Headings with EuroVoc Thesaurus. |
| eurovoc-stw | EuroVoc-STW Semantic Alignment | Economics | Publications Office | | en | Semantic Alignment of STW Thesaurus for economics with EuroVoc Thesaurus. |
| eurovoc-umthes-semantic-alignment | EuroVoc-Umthes Semantic Alignment | Environmental | Publications Office | | de, en | Semantic Alignment of the Umthes (German Environmental Thesaurus) with EuroVoc Thesaurus. |
| eurovoc-unbis | EuroVoc-UNBIS Semantic Alignment | General | Publications Office | | ar, cn, en, es, fr, ru | Semantic Alignment of UNBIS Thesaurus with EuroVoc Thesaurus. |
| eurovoc-unesco-semantic-alignment | EuroVoc-UNESCO Semantic Alignment | General | Publications Office | European Union | en, es, fr, ru | Semantic Alignment of the UNESCO Thesaurus with EuroVoc Thesaurus. |
| evroterm | Evroterm | legislation | General Secretariat of the Government of the Republic of Slovenia | European Union | en, sl | The Evroterm terminology collection contains terms from EU legal regulations and other documents related to the EU. |
| finnish-legislation | Finnish legislation | | | Finland | | Finnish legislation as linked data. This dataset covers the following legal subdomains: legislation, case law (supreme court), case law (supreme administrative) and court. http://data.finlex.fi |
| gemet | GEMET (General Multilingual Environmental Thesaurus) | General | EIONET (European Environment Information and Observation Network) | | Several languages including: en, de, es, it | GEMET is a compilation of the following resources. - "Umwelt Thesaurus" that has more than 2.000 descriptors out of 8.500 in German and English. - "Thesaurus Italiano per l'Ambiente (TIA)" with more than 4.000 descriptors in Italian, English, Dutch and German. - "Multilingual Environment Thesaurus (MET)" with more than 2.300 descriptors in Dutch, Danish, English, French, German, Italian, Norwegian and |

| | | | | | | |
|--|---|--|-----------------------------------|----------------|--|---|
| | | | | | | Spanish. - “EnVoc Thesaurus”, of UNEP Infoterra, 1997 edition, with about 2.000 descriptors in English, French and Spanish, with possibility of access to Arabic, Chinese and Russian. - “Thesaurus de Medio Ambiente” with more than 2.600 descriptors in Spanish, English, French, German. - “Lexique environnement - Planète” with more than 5.000 descriptors in French and English. - Descriptors of relevant documents of the EEA, namely “Europe’s Environment, The Dobris Assessment”, the “DPSIR Data Flow Scheme”, as well as terminology of ETCs and Eionet, in English. - Descriptors of the “Thesaurus Eurovoc” in French, English, Dutch, German, Italian, and Spanish, with possibility of access to Danish, Greek, and Portuguese. The result are 6562 terms arranged in 3 super-groups, 30 groups plus 5 accessory, instrumental groups, hierarchically organised. |
| gllir | Glossary of labour law and industrial relations | employment | International Labour Organization | European Union | en | |
| gllt | German labour law thesaurus | Labor Law | Wolters Kluwer Deutschland GmbH | Germany | de | Labour law thesaurus covers all main areas of labor law, including the roles of employee and employer; legal aspects around labour contracts and dismissal; also co-determination and industrial action. |
| gowers-review-of-intellectual-property | Gowers Review of Intellectual Property | laws | Crown | | en | Glossary about Intellectual Property can be found on pages 121 to 127. |
| iate | IATE | Law, information technology, agriculture, etc. | European Union | n.a. | bg, cs, da, de, el, en, es, et, fi, fr, ga, hr, hu, it, la, lt, lv, mt, nl, pl, pt, ro, sk, sl, sv | IATE is a terminological database developed by the European Union that contains around 8,5 million terms in the 24 official languages of the EU. It can also be downloaded both in RDF and TBX format. IATE uses EuroVoc Thesaurus to classify its entries by domain. |
| ilo-taxonomy | ILO Taxonomy | employment | International Labour Organization | | en, fr, es | ILO Taxonomy contains terms related to the world of work in English, French, and Spanish. |
| imf | International Monetary Fund Terminology | | IMF | | de, en, es | This terminology list contains over 150,000 terms useful to translators working with IMF material. It provides English terms with their equivalents in a number of languages. This list includes words, phrases, and institutional titles commonly encountered in IMF documents in areas such as money and banking, public finance, balance of payments, and economic growth. |
| industrialstandardsglossary | Industrial Standards Lynx Glossary | Legal | OEG | | en | Monolingual termlist from the Industrial Standards domain in RDF linked with BabelNet and DBpedia. |

| | | | | | | |
|----------------------|--|---------------------|--|-----------------------|--|--|
| informea | InforMEA Glossary (UNESCO) | Environmental Law | UNESCO | | en | The glossary contains terms, definitions and related information on Multilateral Environmental Agreements. Such terms are classified in 6 different domains: Water, Chemicals and Wastes, Biodiversity, Air and Climate, Land and Environmental Law and Governance. Terms are also organised hierarchically, grouped in small sets of terms that are all dependent on a top term. This glossary also establishes broader and narrower relations between terms. Moreover, it provides sources where the terms have been applied to check their right usage context. |
| inspire | INSPIRE Glossary (EU) | Spatial information | European Union | | en, es | Glossary developed by the INSPIRE Knowledge Base of the European Union. The INSPIRE Glossary contains 195 general terms and definitions that specify the common terminology used in the INSPIRE Directive and in the INSPIRE Implementing Rules documents. This glossary contains definitions both in English and Spanish. It also provides information about the previous versions of each entry and it marks if terms are valid or not. |
| jrcaquis | JRC-Acquis | Legal | Ralf Steinberger - European Commission - Joint Research Centre (JRC) | | bg, cs, da, de, el, en, es, et, fi, fr, hu, it, lt, lv, mt, nl, pl, pt, ro, sk, sl, sv | JRC-Acquis is a collection of legislative texts from the European Union generated between the years 1958 and 2006. They are available in xml format and in 22 languages of the EU. |
| jurivoc | Jurivoc | Law | Bundesgericht (Switzerland) | Switzerland | de, it, fr (and some terms in EU languages) | JURIVOC, is the juridical thesaurus of the Federal Supreme Court and the former Federal Insurance of Switzerland. It contains information in German, French and Italian. Terms are arranged as per the following relations: - LT: lead term - SN: scope note - UF: used for - UFA: used for... and... - NT: narrower term - SA: see also - BT: broader term |
| labour-law-corpus | Labour Law Corpus | Labour Law | Lynx Project | National and European | en, es, de, it | Corpus of legal documents from different countries: Austria, Germany, Ireland, Italy, Spain and UK. And also from the European Union. |
| labourlawglossary | Labour Law Lynx Glossary | Legal | OEG | | es | Monolingual glossary containing terms in Spanish from the labour law domain. Terms have been automatically extracted from legal corpora and manually reviewed by language and law professionals. The glossary was first generated in TBX and eventually converted into RDF to be linked with BabelNet, Eurovoc and DBpedia. |
| legislative-glossary | Legislative Glossary (Glossário Legislativo) | legislation | Câmara Municipal de Bento Gonçalves | | pt | Monolingual glossary about legislation in Portuguese |

| | | | | | | |
|-----------------------------|---|---------------------------------------|------------------------------------|----------------|------------|--|
| libraryofcongress | Library of Congress | General: books, agreements, documents | Library of Congress | Global | en | The Library of Congress Linked Data Service enables both humans and machines to programmatically access authority data of the Library of Congress. |
| llicorporuses | Spanish Labour Law Corpus | Legal | | Spanish | es | Labour law corpus composed by 20 agreements in Spanish provided by Cuatrecasas Lynx partners. |
| myndigheternas-foreskrifter | Government Regulations (Myndigheternas föreskrifter) | legislation | Sveriges regering | Sweden | sv | |
| nacerev2 | NACE, Rev. 2 | statistics | Eurostat | European Union | en | Statistical Classification of Economic Activities in the European Community, Rev. 2 |
| nomothesia | NOMOTHESIA | | | Greece | el | Greek legislation modelled as per Metalex OWL ontology. It can be accessed through a SPARQL endpoint or downloaded as RDF file. The content of this resource can also be requested through a websearch application. |
| osh-thesaurus | OSH Thesaurus | employment | International Labour Organization | United Nations | en, fr, es | OSH Thesaurus contains over 15,000 multilingual terms and synonyms on occupational safety and health. |
| psi-glossary | PSI Glossary | information | European Commission | | en | Contains terminology related to Public Sector Information |
| quality-glossary | Quality Glossary (Терминологичен речник на качеството) | statistics | Национален статистически институт. | European Union | bg, en | Bilingual glossary about quality in English and Bulgarian with definitions. |
| saij | SAIJ Legal Thesaurus (Argentine Juridical Information System) | Law | Government of Argentine | | es | SAIJ Thesaurus organises legal knowledge through a list, modelled with SKOS, of controlled terms which represent concepts. It is used to ease users' access information related to the Argentinian legal system that can be found in a file or in a documentation centre. Terms are also hierarchically organised with broader and narrower relations. |
| stw | STW Thesaurus for Economics | Economics | ZBW | | en | The thesaurus provides vocabulary on any economic subject: almost 6,000 standardized subject headings and about 20,000 additional entry terms to support individual keywords. Terms used in law, sociology, or politics can also be found. This thesaurus is provided by the Leibniz Information Centre and it has been published in RDFa format to boost the reuse of such resources in the Semantic Web. It is one of the first resources in the Linked Open Data Cloud since it is mapped with a great number of related resources. |

| | | | | | | |
|------------------|--|---|---|----------------|--|--|
| tcnen | TERMCAT Collective Negotiation Glossary EN | Legal | TERMCAT | dbpedia | en | Monolingual termlist from the Collective Negotiation (Labour Law) domain in RDF linked with DBpedia and EuroVoc. (English) |
| tcnes | TERMCAT Collective Negotiation Glossary ES | Legal | TERMCAT | | es | Monolingual termlist from the Collective Negotiation (Labour Law) domain in RDF linked with DBpedia and EuroVoc. (Spanish) |
| temcoord | Temcoord Glossaries | Several topics | DG TRAD | European Union | bg, cs, da, de, el, en, es, et, fi, fr, ga, hr, hu, it, lt, lv, mt, nl, pl, pt, ro, sk, sl, sv | Repository that contains around 300 hundred glossaries developed by European institutions and bodies in the 23 languages of the European Union. Most of them are available as PDF files, while others have been published in HTML format. The content of this repository is heterogeneous in terms of domain and information exposed in each glossary: some glossaries provide with definitions of the terms while others establish equivalences in different languages, etc. |
| termcat | Termcat Terminological Database | General | Termcat | | ca, en, es, de, fr, it | TERMCAT's mission is to ensure the development and integration of Catalan terminology into both specialist sectors and society in general. |
| thesoz | TheSoz - Thesaurus for Social Sciences | Social Sciences | Leibniz Institut für Sozialwissenschaften | | en, de, fr, ru | Thesaurus about social sciences organised according SKOS vocabulary containing terms in English, German, French and Russian. |
| uk-legislation | UK legislation | | | | | UK legislation as linked data |
| umwelt-thesaurus | Umwelt-Thesaurus | Environmental | Federal Environment Agency, Germany | | de, en | German Thesaurus containing terms on environmental protection. It includes 13500 descriptors and 40000 linked German-language synonyms with 35000 English translations and around 11000 definitions. |
| unbis | UNBIS Thesaurus | General | UN Library | | ar, cn, en, es, fr, ru | The UNBIS Thesaurus is a multilingual database of the controlled vocabulary used to describe UN documents, and terms are extracted from these documents. |
| unesco | UNESCO Thesaurus | Education, Science, Culture, Politics, Countries, Information | UNESCO | | en, es, fr, ru | The UNESCO Thesaurus is a controlled and structured list of terms used in subject analysis and retrieval of documents and publications in several fields. The thesaurus is divided into seven greater domains and other smaller thesauri, where terms are hierarchically organised as per the following properties: SN - Scope Note MT - Microthesaurus UF - Used For BT - Broader Term NT - Narrower Term RT - Related Term The UNESCO Thesaurus can be downloaded as RDF file, accessed through a SPARQL endpoint and in the website: search by domain, by alphabetical order and by hierarchical order. |

Appendix 2. Existing Domain Independent vocabularies from KD

| name | title | domain | publisher | description | language | format |
|---------|---|--------|----------------|---|-------------------------|---------------|
| api | Lexicala API | DIV | K Dictionaries | data transfer facility | 40+ languags | JSON, JSON-LD |
| mdls-de | Global German | DIV | K Dictionaries | German monolingual dictionary core, with translation to English, Dutch (+ other languages) | de, en, nl + 7 | XML, RDF |
| mdls-en | Global English | DIV | K Dictionaries | English monolingual dictionary core, with translation to Spanish (+ other languages) | en, es + 6 | XML, RDF |
| mdls-es | Global Spanish | DIV | K Dictionaries | Spanish monolingual dictionary core, with translation to English, Dutch (+ other languages) | es, en, nl + 6 | XML, RDF |
| mdls-it | Global Italian | DIV | K Dictionaries | Italian monolingual dictionary core, with translation to English (+ other languages) | it, en + 2 | XML |
| mdls-nl | Global Dutch | DIV | K Dictionaries | Dutch monolingual dictionary core, with translation to German, English, Spanish | nl, de, en, es | XML |
| pw | Password Multilingual Dictionary | DIV | K Dictionaries | English multilingual dictionary (translation 40+ languages) | en, de, es, it, nl +40 | XML |
| kmt-de | MultiGloss German | DIV | K Dictionaries | German-English semi-automated multilingual glossary | de, en, es, it, nl + 40 | XML |
| kmt-es | MultiGloss Spanish | DIV | K Dictionaries | Spanish-English semi-automated multilingual glossary | es, en, de, it, nl + 40 | XML |
| kmt-it | MultiGloss Italian | DIV | K Dictionaries | Italian-English semi-automated multilingual glossary | it, en, de, es, nl + 40 | XML |
| kmt-nl | MultiGloss Dutch | DIV | K Dictionaries | Dutch-English semi-automated multilingual glossary | nl, en, de, es, nl + 40 | XML |
| rhwcd | Random House Webster's College Dictionary | DIV | K Dictionaries | English monlingual glossary | en | XML |
| wfl | Word Form Lists | DIV | K Dictionaries | morphol;ogical lists of inflected forms and keywords | de, en, es, it, nl | XLSX |

Appendix 3. Sample data of Domain Independent Vocabularies from KD

1. GLOBAL – Dutch, German, English, Spanish
 - i. HTML
 - ii. XML
 - iii. JSON
 - iv. JSON-LD
-
2. PASSWORD – English, German, English, Spanish
 - i. HTML
 - ii. XML
-
3. KMT – Spanish, German, English, Dutch
 - i. HTML
 - ii. XML
-
4. RANDOM HOUSE
 - i. HTML
 - ii. XML

GLOBAL - NL-DE,EN,ES

HTML

grond [xɾɔnt] *nm* (*pl -en* []) **1** =bodem; oppervlakte van de aarde of vloer van een ruimte

{de} - Boden *m*, Erdboden *m*

{en} - ground, land

{es} - tierra *f*, suelo *m*

◊ *De gevangenen moesten op de grond gaan liggen.*

{de} - *Die Gefangenen mussten sich auf den Boden legen.*

{en} - *The prisoners had to get down on the ground.*

{es} - *Los presos tuvieron que acostarse en el suelo.*

◆ **te gronde gaan** verloren gaan

{de} - zugrunde gehen -

{en} - to be ruined; to go to the dogs

{es} - venirse abajo

◆ **met de grond gelijkmaken** (iets) helemaal afbreken

{de} - dem Erdboden gleichmachen -

{en} - to wipe (something) off the map

{es} - derribar

◆ **met de grond gelijkmaken** scherpe kritiek leveren op (iemand)

{de} - jemanden zugrunde richten -

{en} - to run (someone) down

{es} - poner (a alguien) por los suelos

◆ **de grond in boren** ernstige kritiek leveren op (iets of iemand)

{de} - jemanden zugrunde richten -, jemanden am Boden zerstören -

{en} - to grind (someone) down

{es} - no dejar títere con cabeza

◆ **van de grond komen** ontstaan, vorm krijgen

{de} - vom Fleck kommen -

{en} - to get off the ground

{es} - arrancar

◊ *Dat onderzoek komt maar niet van de grond.*

{de} - *Diese Untersuchung kommt nicht vom Fleck.*

{en} - *That research just won't get off the ground.*

{es} - *Esa investigación no acaba de arrancar.*

◆ **uit de grond stampen** in korte tijd maken of doen ontstaan

{de} - aus dem Boden stampfen -

{en} - to whip up (something)

{es} - construir en menos de nada, sacar de la nada

◊ *projecten uit de grond stampen*

{de} - *Projekte aus dem Boden stampfen*

{en} - *to whip up some projects*

{es} - *realizar proyectos de la nada*

◆ **begane grond** verdieping van een gebouw op het niveau van de straat

{de} - Erdgeschoss *nt*

{en} - ground floor

{es} - planta baja

2 =aarde; bovenste laag van het aardoppervlak

{de} - Erde *f*, Erdboden *m*, Boden *nt*

{en} - ground, land

{es} - tierra *f*

◊ Door de vruchtbare grond kun je hier goed groenten kweken.

{de} - Durch den fruchtbaren Boden kann man hier gut Gemüse anbauen.

{en} - You can easily grow vegetables here, because the soil is quite fertile.

{es} - Por la tierra fértil es fácil cultivar verduras aquí.

3 =argument; reden

{de} - Grund *m*

{en} - ground, foundation, reason

{es} - motivo *m*, razón *f*

◊ op goede gronden iets kiezen

{de} - etwas aus guten Gründen wählen

{en} - to choose something with good reason

{es} - elegir algo por motivos fundados

◆ op grond van=vanwege; wegens

{de} - aufgrund -, wegen -

{en} - on the strength of, due to

{es} - por causa de

◊ Het bedrijf nam hem in dienst op grond van zijn specialisme.

{de} - Das Unternehmen stellte ihn aufgrund seiner Spezialisierung ein.

{en} - The company took him on payroll due to his expertise.

{es} - La empresa lo contrató por causa de su especialidad.

XML

```

<Entry identifier="EN00008029" hw="grond" HomNum="" pos="noun">
  <DictionaryEntry identifier="NL_DE00008062" version="1">
    <HeadwordBlock>
      <HeadwordCtn>
        <Headword>grond</Headword>
        <Display>grond</Display>
        <PronunciationBlock>
          <PronunciationCtn>
            <Pronunciation>xrɔnt</Pronunciation>
          </PronunciationCtn>
        </PronunciationBlock>
        <PartOfSpeech value="noun" />
        <GrammaticalGender value="masculine" />
        <InflectionBlock>
          <InflectionCtn>
            <GrammaticalNumber value="plural" />
            <Inflection>grond|en</Inflection>
            <Display>-en</Display>
            <PronunciationCtn>
              <Pronunciation>'xrɔndə(n)</Pronunciation>
              <Display />
            </PronunciationCtn>
          </InflectionCtn>
        </InflectionBlock>
      </HeadwordCtn>
    </HeadwordBlock>
    <SenseBlock>
      <SenseGrp identifier="NL_SE00009937" version="1">
        <Synonym>bodem</Synonym>
        <SidCtn identifier="SI00009582" version="1">
          <SenseIndicator />
        </SidCtn>
        <DefinitionCtn>
          <Definition>oppervlakte van de aarde of vloer van een ruimte</Definition>
        </DefinitionCtn>
        <TranslationCluster identifier="TC00024561" text="oppervlakte van de aarde of vloer van een ruimte"
type="def">
          <Locale lang="de">
            <TranslationBlock>
              <TranslationCtn>
                <Translation>Boden</Translation>
                <GrammaticalGender value="masculine" />
              </TranslationCtn>
              <TranslationCtn>
                <Translation>Erdboden</Translation>
                <GrammaticalGender value="masculine" />
              </TranslationCtn>
            </TranslationBlock>
          </Locale>
          <Locale lang="en">
            <TranslationBlock>

```

```

    <TranslationCtn>
      <Translation>ground, land</Translation>
    </TranslationCtn>
  </TranslationBlock>
</Locale>
<Locale lang="es">
  <TranslationBlock>
    <TranslationCtn>
      <Translation>tierra</Translation>
      <GrammaticalGender value="feminine" />
    </TranslationCtn>
    <TranslationCtn>
      <Translation>suelo</Translation>
      <GrammaticalGender value="masculine" />
    </TranslationCtn>
  </TranslationBlock>
</Locale>
</TranslationCluster>
<ExampleCtn type="sid" version="1" id="EX00012749">
  <Example>De gevangenen moesten op de grond gaan liggen.</Example>
  <TranslationCluster identifier="TC00024562" text="De gevangenen moesten op de grond gaan liggen."
type="exmp">
  <Locale lang="de">
    <TranslationBlock>
      <TranslationCtn>
        <Translation>Die Gefangenen mussten sich auf den Boden legen.</Translation>
      </TranslationCtn>
    </TranslationBlock>
  </Locale>
  <Locale lang="en">
    <TranslationBlock>
      <TranslationCtn>
        <Translation>The prisoners had to get down on the
ground.</Translation>
      </TranslationCtn>
    </TranslationBlock>
  </Locale>
  <Locale lang="es">
    <TranslationBlock>
      <TranslationCtn>
        <Translation>Los presos tuvieron que acostarse en el suelo.</Translation>
      </TranslationCtn>
    </TranslationBlock>
  </Locale>
</TranslationCluster>
</ExampleCtn>
<CompositionalPhraseCtn version="1" id="CP00002201">
  <CompositionalPhrase>te gronde gaan</CompositionalPhrase>
  <DefinitionCtn>
    <Definition>verloren gaan</Definition>
  </DefinitionCtn>
  <TranslationCluster identifier="TC00024563" text="verloren gaan" type="phrasedef">
    <Locale lang="de">
      <TranslationBlock>
        <TranslationCtn>
          <Translation>zugrunde gehen</Translation>
          <GrammaticalGender value="-" />
        </TranslationCtn>
      </TranslationBlock>
    </Locale>
    <Locale lang="en">
      <TranslationBlock>
        <TranslationCtn>
          <Translation>to be ruined; to go to the dogs</Translation>
        </TranslationCtn>
      </TranslationBlock>
    </Locale>
    <Locale lang="es">
      <TranslationBlock>
        <TranslationCtn>
          <Translation>venirse abajo</Translation>
        </TranslationCtn>
      </TranslationBlock>
    </Locale>
  </TranslationCluster>
</CompositionalPhraseCtn>
<CompositionalPhraseCtn version="1" id="CP00002202">

```

```

<CompositionalPhrase>met de grond gelijkmaken</CompositionalPhrase>
<DefinitionCtn>
  <Definition>(iets) helemaal afbreken</Definition>
</DefinitionCtn>
<TranslationCluster identifier="TC00024564" text="(iets) helemaal afbreken" type="phrasedef">
  <Locale lang="de">
    <TranslationBlock>
      <TranslationCtn>
        <Translation>dem Erdboden gleichmachen</Translation>
        <GrammaticalGender value="-" />
      </TranslationCtn>
    </TranslationBlock>
  </Locale>
  <Locale lang="en">
    <TranslationBlock>
      <TranslationCtn>
        <Translation>to wipe (something) off the map</Translation>
      </TranslationCtn>
    </TranslationBlock>
  </Locale>
  <Locale lang="es">
    <TranslationBlock>
      <TranslationCtn>
        <Translation>derribar</Translation>
      </TranslationCtn>
    </TranslationBlock>
  </Locale>
</TranslationCluster>
</CompositionalPhraseCtn>
<CompositionalPhraseCtn version="1" id="CP0002203">
  <CompositionalPhrase>met de grond gelijkmaken</CompositionalPhrase>
  <DefinitionCtn>
    <Definition>scherpe kritiek leveren op (iemand)</Definition>
  </DefinitionCtn>
  <TranslationCluster identifier="TC00024565" text="scherpe kritiek leveren op (iemand)"
type="phrasedef">
    <Locale lang="de">
      <TranslationBlock>
        <TranslationCtn>
          <Translation>jemanden zugrunde richten</Translation>
          <GrammaticalGender value="-" />
        </TranslationCtn>
      </TranslationBlock>
    </Locale>
    <Locale lang="en">
      <TranslationBlock>
        <TranslationCtn>
          <Translation>to run (someone) down</Translation>
        </TranslationCtn>
      </TranslationBlock>
    </Locale>
    <Locale lang="es">
      <TranslationBlock>
        <TranslationCtn>
          <Translation>poner (a alguien) por los suelos</Translation>
        </TranslationCtn>
      </TranslationBlock>
    </Locale>
  </TranslationCluster>
</CompositionalPhraseCtn>
<CompositionalPhraseCtn version="1" id="CP0002204">
  <CompositionalPhrase>de grond in boren</CompositionalPhrase>
  <DefinitionCtn>
    <Definition>ernstige kritiek leveren op (iets of iemand)</Definition>
  </DefinitionCtn>
  <TranslationCluster identifier="TC00024566" text="ernstige kritiek leveren op (iets of iemand)"
type="phrasedef">
    <Locale lang="de">
      <TranslationBlock>
        <TranslationCtn>
          <Translation>jemanden zugrunde richten</Translation>
          <GrammaticalGender value="-" />
        </TranslationCtn>
        <TranslationCtn>
          <Translation>jemanden am Boden zerstören</Translation>
          <GrammaticalGender value="-" />
        </TranslationCtn>
      </TranslationBlock>
    </Locale>
  </TranslationCluster>
</CompositionalPhraseCtn>

```

```

    </TranslationBlock>
  </Locale>
  <Locale lang="en">
    <TranslationBlock>
      <TranslationCtn>
        <Translation>to grind (someone) down</Translation>
      </TranslationCtn>
    </TranslationBlock>
  </Locale>
  <Locale lang="es">
    <TranslationBlock>
      <TranslationCtn>
        <Translation>no dejar títere con cabeza</Translation>
      </TranslationCtn>
    </TranslationBlock>
  </Locale>
</TranslationCluster>
</CompositionalPhraseCtn>
<CompositionalPhraseCtn version="1" id="CP00002205">
  <CompositionalPhrase>van de grond komen</CompositionalPhrase>
  <DefinitionCtn>
    <Definition>ontstaan, vorm krijgen</Definition>
  </DefinitionCtn>
  <TranslationCluster identifier="TC00024567" text="ontstaan, vorm krijgen" type="phrasedef">
    <Locale lang="de">
      <TranslationBlock>
        <TranslationCtn>
          <Translation>vom Fleck kommen</Translation>
          <GrammaticalGender value="-" />
        </TranslationCtn>
      </TranslationBlock>
    </Locale>
    <Locale lang="en">
      <TranslationBlock>
        <TranslationCtn>
          <Translation>to get off the ground</Translation>
        </TranslationCtn>
      </TranslationBlock>
    </Locale>
    <Locale lang="es">
      <TranslationBlock>
        <TranslationCtn>
          <Translation>arrancar</Translation>
        </TranslationCtn>
      </TranslationBlock>
    </Locale>
  </TranslationCluster>
  <ExampleCtn type="sid" version="1" id="EX00012750">
    <Example>Dat onderzoek komt maar niet van de grond.</Example>
    <TranslationCluster identifier="TC00024568" text="Dat onderzoek komt maar niet van de grond."
type="phraseexmp">
      <Locale lang="de">
        <TranslationBlock>
          <TranslationCtn>
            <Translation>Diese Untersuchung kommt nicht vom Fleck.</Translation>
          </TranslationCtn>
        </TranslationBlock>
      </Locale>
      <Locale lang="en">
        <TranslationBlock>
          <TranslationCtn>
            <Translation>That research just won't get off the
ground.</Translation>
          </TranslationCtn>
        </TranslationBlock>
      </Locale>
      <Locale lang="es">
        <TranslationBlock>
          <TranslationCtn>
            <Translation>Esa investigación no acaba de arrancar.</Translation>
          </TranslationCtn>
        </TranslationBlock>
      </Locale>
    </TranslationCluster>
  </ExampleCtn>
</CompositionalPhraseCtn>
<CompositionalPhraseCtn version="1" id="CP00002206">

```

```

<CompositionalPhrase>uit de grond stampen</CompositionalPhrase>
<DefinitionCtn>
  <Definition>in korte tijd maken of doen ontstaan</Definition>
</DefinitionCtn>
<TranslationCluster identifier="TC00024569" text="in korte tijd maken of doen ontstaan"
type="phrasedef">
  <Locale lang="de">
    <TranslationBlock>
      <TranslationCtn>
        <Translation>aus dem Boden stampfen</Translation>
        <GrammaticalGender value="-" />
      </TranslationCtn>
    </TranslationBlock>
  </Locale>
  <Locale lang="en">
    <TranslationBlock>
      <TranslationCtn>
        <Translation>to whip up (something)</Translation>
      </TranslationCtn>
    </TranslationBlock>
  </Locale>
  <Locale lang="es">
    <TranslationBlock>
      <TranslationCtn>
        <Translation>construir en menos de nada</Translation>
      </TranslationCtn>
      <TranslationCtn>
        <Translation>sacar de la nada</Translation>
      </TranslationCtn>
    </TranslationBlock>
  </Locale>
</TranslationCluster>
<ExampleCtn type="sid" version="1" id="EX00012751">
  <Example>projecten uit de grond stampen</Example>
  <TranslationCluster identifier="TC00024570" text="projecten uit de grond stampen" type="phraseexmp">
    <Locale lang="de">
      <TranslationBlock>
        <TranslationCtn>
          <Translation>Projekte aus dem Boden stampfen</Translation>
        </TranslationCtn>
      </TranslationBlock>
    </Locale>
    <Locale lang="en">
      <TranslationBlock>
        <TranslationCtn>
          <Translation>to whip up some projects</Translation>
        </TranslationCtn>
      </TranslationBlock>
    </Locale>
    <Locale lang="es">
      <TranslationBlock>
        <TranslationCtn>
          <Translation>realizar proyectos de la nada</Translation>
        </TranslationCtn>
      </TranslationBlock>
    </Locale>
  </TranslationCluster>
</ExampleCtn>
</CompositionalPhraseCtn>
<CompositionalPhraseCtn version="1" id="CP00002207">
  <CompositionalPhrase>begane grond</CompositionalPhrase>
  <DefinitionCtn>
    <Definition>verdieping van een gebouw op het niveau van de straat</Definition>
  </DefinitionCtn>
  <TranslationCluster identifier="TC00024571" text="verdieping van een gebouw op het niveau van de
straat" type="phrasedef">
    <Locale lang="de">
      <TranslationBlock>
        <TranslationCtn>
          <Translation>Erdgeschoss</Translation>
          <GrammaticalGender value="neuter" />
        </TranslationCtn>
      </TranslationBlock>
    </Locale>
    <Locale lang="en">
      <TranslationBlock>
        <TranslationCtn>

```

```

        <Translation>ground floor</Translation>
    </TranslationCtn>
</TranslationBlock>
</Locale>
<Locale lang="es">
    <TranslationBlock>
        <TranslationCtn>
            <Translation>planta baja</Translation>
        </TranslationCtn>
    </TranslationBlock>
</Locale>
</TranslationCluster>
</CompositionalPhraseCtn>
</SenseGrp>
<SenseGrp identifier="NL_SE00009938" version="1">
    <Synonym>aarde</Synonym>
    <SidCtn identifier="SI00009583" version="1">
        <SenseIndicator />
    </SidCtn>
    <DefinitionCtn>
        <Definition>bovenste laag van het aardoppervlak</Definition>
    </DefinitionCtn>
    <TranslationCluster identifier="TC00024572" text="bovenste laag van het aardoppervlak" type="def">
        <Locale lang="de">
            <TranslationBlock>
                <TranslationCtn>
                    <Translation>Erde</Translation>
                    <GrammaticalGender value="feminine" />
                </TranslationCtn>
                <TranslationCtn>
                    <Translation>Erdboden</Translation>
                    <GrammaticalGender value="masculine" />
                </TranslationCtn>
                <TranslationCtn>
                    <Translation>Boden</Translation>
                    <GrammaticalGender value="neuter" />
                </TranslationCtn>
            </TranslationBlock>
        </Locale>
        <Locale lang="en">
            <TranslationBlock>
                <TranslationCtn>
                    <Translation>ground, land</Translation>
                </TranslationCtn>
            </TranslationBlock>
        </Locale>
        <Locale lang="es">
            <TranslationBlock>
                <TranslationCtn>
                    <Translation>tierra</Translation>
                    <GrammaticalGender value="feminine" />
                </TranslationCtn>
            </TranslationBlock>
        </Locale>
    </TranslationCluster>
    <ExampleCtn type="sid" version="1" id="EX00012752">
        <Example>Door de vruchtbare grond kun je hier goed groenten kweken.</Example>
        <TranslationCluster identifier="TC00024573" text="Door de vruchtbare grond kun je hier goed groenten kweken." type="exmp">
            <Locale lang="de">
                <TranslationBlock>
                    <TranslationCtn>
                        <Translation>Durch den fruchtbaren Boden kann man hier gut Gemüse anbauen.</Translation>
                    </TranslationCtn>
                </TranslationBlock>
            </Locale>
            <Locale lang="en">
                <TranslationBlock>
                    <TranslationCtn>
                        <Translation>You can easily grow vegetables here, because the soil is quite fertile.</Translation>
                    </TranslationCtn>
                </TranslationBlock>
            </Locale>
            <Locale lang="es">
                <TranslationBlock>
                    <TranslationCtn>

```



```

        <Translation>Por la tierra fértil es fácil cultivar verduras aquí.</Translation>
    </TranslationCtn>
</TranslationBlock>
</Locale>
</TranslationCluster>
</ExampleCtn>
</SenseGrp>
<SenseGrp identifier="NL_SE0009939" version="1">
    <Synonym>argument</Synonym>
    <SidCtn identifier="SI0009584" version="1">
        <SenseIndicator />
    </SidCtn>
    <DefinitionCtn>
        <Definition>reden</Definition>
    </DefinitionCtn>
    <TranslationCluster identifier="TC00024574" text="reden" type="def">
        <Locale lang="de">
            <TranslationBlock>
                <TranslationCtn>
                    <Translation>Grund</Translation>
                    <GrammaticalGender value="masculine" />
                </TranslationCtn>
            </TranslationBlock>
        </Locale>
        <Locale lang="en">
            <TranslationBlock>
                <TranslationCtn>
                    <Translation>ground, foundation, reason</Translation>
                </TranslationCtn>
            </TranslationBlock>
        </Locale>
        <Locale lang="es">
            <TranslationBlock>
                <TranslationCtn>
                    <Translation>motivo</Translation>
                    <GrammaticalGender value="masculine" />
                </TranslationCtn>
                <TranslationCtn>
                    <Translation>razón</Translation>
                    <GrammaticalGender value="feminine" />
                </TranslationCtn>
            </TranslationBlock>
        </Locale>
    </TranslationCluster>
</ExampleCtn type="sid" version="1" id="EX00012753">
    <Example>op goede gronden iets kiezen</Example>
    <TranslationCluster identifier="TC00024575" text="op goede gronden iets kiezen" type="exmp">
        <Locale lang="de">
            <TranslationBlock>
                <TranslationCtn>
                    <Translation>etwas aus guten Gründen wählen</Translation>
                </TranslationCtn>
            </TranslationBlock>
        </Locale>
        <Locale lang="en">
            <TranslationBlock>
                <TranslationCtn>
                    <Translation>to choose something with good
                    reason</Translation>
                </TranslationCtn>
            </TranslationBlock>
        </Locale>
        <Locale lang="es">
            <TranslationBlock>
                <TranslationCtn>
                    <Translation>elegir algo por motivos fundados</Translation>
                </TranslationCtn>
            </TranslationBlock>
        </Locale>
    </TranslationCluster>
</ExampleCtn>
<CompositionalPhraseCtn version="1" id="CP0002208">
    <CompositionalPhrase>op grond van</CompositionalPhrase>
    <Synonym>vanwege</Synonym>
    <DefinitionCtn>
        <Definition>wegens</Definition>
    </DefinitionCtn>

```

```

<TranslationCluster identifier="TC00024576" text="wegens" type="phrasedef">
  <Locale lang="de">
    <TranslationBlock>
      <TranslationCtn>
        <Translation>aufgrund</Translation>
        <GrammaticalGender value="-" />
      </TranslationCtn>
      <TranslationCtn>
        <Translation>wegen</Translation>
        <GrammaticalGender value="-" />
      </TranslationCtn>
    </TranslationBlock>
  </Locale>
  <Locale lang="en">
    <TranslationBlock>
      <TranslationCtn>
        <Translation>on the strength of, due to</Translation>
      </TranslationCtn>
    </TranslationBlock>
  </Locale>
  <Locale lang="es">
    <TranslationBlock>
      <TranslationCtn>
        <Translation>por causa de</Translation>
      </TranslationCtn>
    </TranslationBlock>
  </Locale>
</TranslationCluster>
<ExampleCtn type="sid" version="1" id="EX00012754">
  <Example>Het bedrijf nam hem in dienst op grond van zijn specialisme.</Example>
  <TranslationCluster identifier="TC00024577" text="Het bedrijf nam hem in dienst op grond van zijn
specialisme." type="phraseexmp">
    <Locale lang="de">
      <TranslationBlock>
        <TranslationCtn>
          <Translation>Das Unternehmen stellte ihn aufgrund seiner Spezialisierung ein.</Translation>
        </TranslationCtn>
      </TranslationBlock>
    </Locale>
    <Locale lang="en">
      <TranslationBlock>
        <TranslationCtn>
          <Translation>The company took him on payroll due to his
expertise.</Translation>
        </TranslationCtn>
      </TranslationBlock>
    </Locale>
    <Locale lang="es">
      <TranslationBlock>
        <TranslationCtn>
          <Translation>La empresa lo contrató por causa de su especialidad.</Translation>
        </TranslationCtn>
      </TranslationBlock>
    </Locale>
  </TranslationCluster>
</ExampleCtn>
</CompositionalPhraseCtn>
</SenseGrp>
</SenseBlock>
</DictionaryEntry>
</Entry>

```

JSON

```

{
  "id": "NL_DE0008062",
  "source": "global",
  "language": "nl",
  "version": 1,
  "headword": {
    "text": "grond",
    "display": "grond",
    "pronunciation": {
      "value": "xɾɔnt"
    }
  },
  "pos": "noun",
  "gender": "masculine",
  "inflections": [

```

```

    {
      "text": "grond|en",
      "display": "-en",
      "number": "plural"
    }
  ]
},
"senses": [
  {
    "id": "NL_SE0009937",
    "definition": "oppervlakte van de aarde of vloer van een ruimte",
    "synonyms": [
      "bodem"
    ],
    "translations": {
      "de": [
        {
          "text": "Boden",
          "gender": "masculine"
        },
        {
          "text": "Erdboden",
          "gender": "masculine"
        }
      ],
      "en": [
        {
          "text": "ground"
        },
        {
          "text": "floor"
        }
      ],
      "es": [
        {
          "text": "tierra",
          "gender": "feminine"
        },
        {
          "text": "suelo",
          "gender": "masculine"
        }
      ]
    },
    "examples": [
      {
        "text": "De gevangenen moesten op de grond gaan liggen.",
        "translations": {
          "de": {
            "text": "Die Gefangenen mussten sich auf den Boden legen."
          },
          "en": {
            "text": "The prisoners had to get down on the ground."
          },
          "es": {
            "text": "Los presos tuvieron que acostarse en el suelo."
          }
        }
      }
    ]
  },
  "compositional_phrases": [
    {
      "text": "te gronde gaan",
      "definition": "verloren gaan",
      "translations": {
        "de": {
          "text": "zugrunde gehen",
          "gender": "-"
        },
        "en": [
          {
            "text": "to be ruined"
          },
          {
            "text": "to go to the dogs"
          }
        ]
      }
    }
  ],

```

```

    "es": {
      "text": "venirse abajo"
    }
  },
  {
    "text": "met de grond gelijkmaken",
    "definition": "(iets) helemaal afbreken",
    "translations": {
      "de": {
        "text": "dem Erdboden gleichmachen",
        "gender": "-"
      },
      "en": {
        "text": "to wipe (something) off the map"
      },
      "es": {
        "text": "derribar"
      }
    }
  },
  {
    "text": "met de grond gelijkmaken",
    "definition": "scherpe kritiek leveren op (iemand)",
    "translations": {
      "de": {
        "text": "jemanden zugrunde richten",
        "gender": "-"
      },
      "en": {
        "text": "to run (someone) down"
      },
      "es": {
        "text": "poner (a alguien) por los suelos"
      }
    }
  },
  {
    "text": "de grond in boren",
    "definition": "ernstige kritiek leveren op (iets of iemand)",
    "translations": {
      "de": [
        {
          "text": "jemanden zugrunde richten",
          "gender": "-"
        },
        {
          "text": "jemanden am Boden zerstören",
          "gender": "-"
        }
      ],
      "en": {
        "text": "to grind (someone) down"
      },
      "es": {
        "text": "no dejar títere con cabeza"
      }
    }
  },
  {
    "text": "van de grond komen",
    "definition": "ontstaan, vorm krijgen",
    "translations": {
      "de": {
        "text": "vom Fleck kommen",
        "gender": "-"
      },
      "en": {
        "text": "to get off the ground"
      },
      "es": {
        "text": "arrancar"
      }
    }
  },
  "examples": [
    {
      "text": "Dat onderzoek komt maar niet van de grond.",
      "translations": {

```

```

    "de": {
      "text": "Diese Untersuchung kommt nicht vom Fleck."
    },
    "en": {
      "text": "That research just won't get off the ground."
    },
    "es": {
      "text": "Esa investigación no acaba de arrancar."
    }
  }
}
],
},
{
  "text": "uit de grond stampen",
  "definition": "in korte tijd maken of doen ontstaan",
  "translations": {
    "de": {
      "text": "aus dem Boden stampfen",
      "gender": "-"
    },
    "en": {
      "text": "to whip up (something)"
    },
    "es": [
      {
        "text": "construir en menos de nada"
      },
      {
        "text": "sacar de la nada"
      }
    ]
  }
],
},
"examples": [
  {
    "text": "projecten uit de grond stampen",
    "translations": {
      "de": {
        "text": "Projekte aus dem Boden stampfen"
      },
      "en": {
        "text": "to whip up some projects"
      },
      "es": {
        "text": "realizar proyectos de la nada"
      }
    }
  }
]
},
{
  "text": "begane grond",
  "definition": "verdieping van een gebouw op het niveau van de straat",
  "translations": {
    "de": {
      "text": "Erdgeschoss",
      "gender": "neuter"
    },
    "en": {
      "text": "ground floor"
    },
    "es": {
      "text": "planta baja"
    }
  }
}
],
},
{
  "id": "NL_SE00009938",
  "definition": "bovenste laag van het aardoppervlak",
  "synonyms": [
    "aarde"
  ],
  "translations": {
    "de": [
      {

```

```

    "text": "Erde",
    "gender": "feminine"
  },
  {
    "text": "Erdboden",
    "gender": "masculine"
  },
  {
    "text": "Boden",
    "gender": "neuter"
  }
],
"en": [
  {
    "text": "ground"
  },
  {
    "text": "land"
  }
],
"es": {
  "text": "tierra",
  "gender": "feminine"
}
},
"examples": [
  {
    "text": "Door de vruchtbare grond kun je hier goed groenten kweken.",
    "translations": {
      "de": {
        "text": "Durch den fruchtbaren Boden kann man hier gut Gemüse anbauen."
      },
      "en": {
        "text": "You can easily grow vegetables here, because the soil is quite fertile."
      },
      "es": {
        "text": "Por la tierra fértil es fácil cultivar verduras aquí."
      }
    }
  }
]
},
{
  "id": "NL_SE0009939",
  "definition": "reden",
  "synonyms": [
    "argument"
  ],
  "translations": {
    "de": {
      "text": "Grund",
      "gender": "masculine"
    },
    "en": [
      {
        "text": "ground"
      },
      {
        "text": "reason"
      }
    ],
    "es": [
      {
        "text": "motivo",
        "gender": "masculine"
      },
      {
        "text": "razón",
        "gender": "feminine"
      }
    ]
  }
},
"examples": [
  {
    "text": "op goede gronden iets kiezen",
    "translations": {
      "de": {

```



```

{
  "@id": "kd-lex:NL/grond-n-m-form",
  "@type": "Ontolex:Form",
  "text": {
    "nl": "grond"
  },
  "pronunciation": {
    "nl-fonipa": "xɾɔnt"
  },
  "gender": "lexinfo:masculine",
  "display": {
    "@id": "kd-lex:NL/grond-n-m-form-display",
    "@type": "kd:Display",
    "text": {
      "nl": "grond"
    }
  }
}
],
"sense": [
  {
    "@id": "kd-lex:NL/grond-n-NL_SE00009937-sense",
    "@type": "ontolex:LexicalSense",
    "reference": {
      "@id": "kd-base:NL_SE00009937-concept",
      "@type": "skos:Concept",
      "definition": {
        "nl": "oppervlakte van de aarde of vloer van een ruimte"
      }
    },
    "relation": {
      "@id": "kd-lex:NL/grond-n-NL_SE00009937-sense-bodem-grond-n-NL_SE00009937-sense-sr",
      "@type": "ontolex:SenseRelation",
      "category": "lexinfo:synonym",
      "relates": {
        "@id": "kd-lex:NL/bodem-grond-n-NL_SE00009937-sense",
        "@type": "ontolex:LexicalSense",
        "reference": {
          "@id": "kd-base:NL_SE00009937-concept",
          "@type": "skos:Concept"
        }
      },
      "sense_entry": {
        "@id": "kd-lex:NL/bodem",
        "@type": "ontolex:LexicalEntry",
        "form": {
          "@id": "kd-lex:NL/bodem-form",
          "@type": "Ontolex:Form",
          "text": {
            "nl": "bodem"
          }
        }
      }
    }
  }
],
"example": [
  {
    "@id": "kd-lex:NL/grond-n-NL_SE00009937-sense-TC00024562-ex",
    "@type": "kd:UsageExample",
    "value": {
      "nl": "De gevangenen moesten op de grond gaan liggen."
    },
    "relation": {
      "@id": "TC00024562-trans-ex-cl",
      "@type": "kd:TranslationExampleCluster",
      "relates": [
        {
          "@type": "kd:UsageExample",
          "value": {
            "de": "Die Gefangenen mussten sich auf den Boden legen."
          }
        },
        {
          "@type": "kd:UsageExample",
          "value": {
            "en": "The prisoners had to get down on the ground."
          }
        }
      ]
    }
  }
],

```



```

    {
      "@type": "kd:UsageExample",
      "value": {
        "es": "Los presos tuvieron que acostarse en el suelo."
      }
    }
  ]
}
],
"translation": [
  {
    "@id": "kd-trans:NL-ID/grond-n-NL_SE00009937-sense--grond-n-NL_SE00009937-sense-TC00024561-trans",
    "@type": "ontolex:Translation",
    "target": {
      "@id": "kd-lex:ID/-grond-n-NL_SE00009937-sense",
      "@type": "ontolex:LexicalSense",
      "reference": {
        "@id": "kd-base:NL_SE00009937-concept",
        "@type": "skos:Concept"
      },
      "sense_entry": {
        "@id": "kd-lex:ID/",
        "@type": "ontolex:LexicalEntry",
        "form": {
          "@id": "kd-lex:ID/-form",
          "@type": "ontolex:Form"
        }
      }
    }
  },
  {
    "@id": "kd-trans:NL-DE/grond-n-NL_SE00009937-sense-Boden-grond-n-NL_SE00009937-sense-TC00024561-trans",
    "@type": "ontolex:Translation",
    "target": {
      "@id": "kd-lex:DE/Boden-grond-n-NL_SE00009937-sense",
      "@type": "ontolex:LexicalSense",
      "reference": {
        "@id": "kd-base:NL_SE00009937-concept",
        "@type": "skos:Concept"
      },
      "sense_entry": {
        "@id": "kd-lex:DE/Boden",
        "@type": "ontolex:LexicalEntry",
        "form": {
          "@id": "kd-lex:DE/Boden-form",
          "@type": "ontolex:Form",
          "text": {
            "de": "Boden"
          }
        }
      }
    }
  },
  {
    "@id": "kd-trans:NL-DE/grond-n-NL_SE00009937-sense-Erdboden-grond-n-NL_SE00009937-sense-TC00024561-trans",
    "@type": "ontolex:Translation",
    "target": {
      "@id": "kd-lex:DE/Erdboden-grond-n-NL_SE00009937-sense",
      "@type": "ontolex:LexicalSense",
      "reference": {
        "@id": "kd-base:NL_SE00009937-concept",
        "@type": "skos:Concept"
      },
      "sense_entry": {
        "@id": "kd-lex:DE/Erdboden",
        "@type": "ontolex:LexicalEntry",
        "form": {
          "@id": "kd-lex:DE/Erdboden-form",
          "@type": "ontolex:Form",
          "text": {
            "de": "Erdboden"
          }
        }
      }
    }
  }
]
}

```

```

},
{
  "@id": "kd-trans:NL-EN/grond-n-NL_SE00009937-sense-ground-grond-n-NL_SE00009937-sense-TC00024561-trans",
  "@type": "ontolex:Translation",
  "target": {
    "@id": "kd-lex:EN/ground-grond-n-NL_SE00009937-sense",
    "@type": "ontolex:LexicalSense",
    "reference": {
      "@id": "kd-base:NL_SE00009937-concept",
      "@type": "skos:Concept"
    },
    "sense_entry": {
      "@id": "kd-lex:EN/ground",
      "@type": "ontolex:LexicalEntry",
      "form": {
        "@id": "kd-lex:EN/ground-form",
        "@type": "ontolex:Form",
        "text": {
          "en": "ground"
        }
      }
    }
  }
},
{
  "@id": "kd-trans:NL-EN/grond-n-NL_SE00009937-sense-floor-grond-n-NL_SE00009937-sense-TC00024561-trans",
  "@type": "ontolex:Translation",
  "target": {
    "@id": "kd-lex:EN/floor-grond-n-NL_SE00009937-sense",
    "@type": "ontolex:LexicalSense",
    "reference": {
      "@id": "kd-base:NL_SE00009937-concept",
      "@type": "skos:Concept"
    },
    "sense_entry": {
      "@id": "kd-lex:EN/floor",
      "@type": "ontolex:LexicalEntry",
      "form": {
        "@id": "kd-lex:EN/floor-form",
        "@type": "ontolex:Form",
        "text": {
          "en": "floor"
        }
      }
    }
  }
},
{
  "@id": "kd-trans:NL-ES/grond-n-NL_SE00009937-sense-tierra-grond-n-NL_SE00009937-sense-TC00024561-trans",
  "@type": "ontolex:Translation",
  "target": {
    "@id": "kd-lex:ES/tierra-grond-n-NL_SE00009937-sense",
    "@type": "ontolex:LexicalSense",
    "reference": {
      "@id": "kd-base:NL_SE00009937-concept",
      "@type": "skos:Concept"
    },
    "sense_entry": {
      "@id": "kd-lex:ES/tierra",
      "@type": "ontolex:LexicalEntry",
      "form": {
        "@id": "kd-lex:ES/tierra-form",
        "@type": "ontolex:Form",
        "text": {
          "es": "tierra"
        }
      }
    }
  }
},
{
  "@id": "kd-trans:NL-ES/grond-n-NL_SE00009937-sense-suelo-grond-n-NL_SE00009937-sense-TC00024561-trans",
  "@type": "ontolex:Translation",
  "target": {
    "@id": "kd-lex:ES/suelo-grond-n-NL_SE00009937-sense",
    "@type": "ontolex:LexicalSense",
    "reference": {

```

```

      "@id": "kd-base:NL_SE00009937-concept",
      "@type": "skos:Concept"
    },
    "sense_entry": {
      "@id": "kd-lex:ES/suelo",
      "@type": "ontolex:LexicalEntry",
      "form": {
        "@id": "kd-lex:ES/suelo-form",
        "@type": "ontolex:Form",
        "text": {
          "es": "suelo"
        }
      }
    }
  }
},
],
{
  "@id": "kd-lex:NL/grond-n-NL_SE00009938-sense",
  "@type": "ontolex:LexicalSense",
  "reference": {
    "@id": "kd-base:NL_SE00009938-concept",
    "@type": "skos:Concept",
    "definition": {
      "nl": "bovenste laag van het aardoppervlak"
    }
  },
  "relation": {
    "@id": "kd-lex:NL/grond-n-NL_SE00009938-sense-aarde-grond-n-NL_SE00009938-sense-sr",
    "@type": "ontolex:SenseRelation",
    "category": "lexinfo:synonym",
    "relates": {
      "@id": "kd-lex:NL/aarde-grond-n-NL_SE00009938-sense",
      "@type": "ontolex:LexicalSense",
      "reference": {
        "@id": "kd-base:NL_SE00009938-concept",
        "@type": "skos:Concept"
      },
      "sense_entry": {
        "@id": "kd-lex:NL/aarde",
        "@type": "ontolex:LexicalEntry",
        "form": {
          "@id": "kd-lex:NL/aarde-form",
          "@type": "ontolex:Form",
          "text": {
            "nl": "aarde"
          }
        }
      }
    }
  }
},
"example": [
  {
    "@id": "kd-lex:NL/grond-n-NL_SE00009938-sense-TC00024573-ex",
    "@type": "kd:UsageExample",
    "value": {
      "nl": "Door de vruchtbare grond kun je hier goed groenten kweken."
    },
    "relation": {
      "@id": "TC00024573-trans-ex-cl",
      "@type": "kd:TranslationExampleCluster",
      "relates": [
        {
          "@type": "kd:UsageExample",
          "value": {
            "de": "Durch den fruchtbaren Boden kann man hier gut Gemüse anbauen."
          }
        },
        {
          "@type": "kd:UsageExample",
          "value": {
            "en": "You can easily grow vegetables here, because the soil is quite fertile."
          }
        }
      ]
    }
  },
  {
    "@type": "kd:UsageExample",

```

```

    "value": {
      "es": "Por la tierra fértil es fácil cultivar verduras aquí."
    }
  ]
}
],
"translation": [
  {
    "@id": "kd-trans:NL-ID/grond-n-NL_SE00009938-sense--grond-n-NL_SE00009938-sense-TC00024572-trans",
    "@type": "ontolex:Translation",
    "target": {
      "@id": "kd-lex:ID/-grond-n-NL_SE00009938-sense",
      "@type": "ontolex:LexicalSense",
      "reference": {
        "@id": "kd-base:NL_SE00009938-concept",
        "@type": "skos:Concept"
      },
      "sense_entry": {
        "@id": "kd-lex:ID/",
        "@type": "ontolex:LexicalEntry",
        "form": {
          "@id": "kd-lex:ID/-form",
          "@type": "ontolex:Form"
        }
      }
    }
  },
  {
    "@id": "kd-trans:NL-DE/grond-n-NL_SE00009938-sense-Erde-grond-n-NL_SE00009938-sense-TC00024572-trans",
    "@type": "ontolex:Translation",
    "target": {
      "@id": "kd-lex:DE/Erde-grond-n-NL_SE00009938-sense",
      "@type": "ontolex:LexicalSense",
      "reference": {
        "@id": "kd-base:NL_SE00009938-concept",
        "@type": "skos:Concept"
      },
      "sense_entry": {
        "@id": "kd-lex:DE/Erde",
        "@type": "ontolex:LexicalEntry",
        "form": {
          "@id": "kd-lex:DE/Erde-form",
          "@type": "ontolex:Form",
          "text": {
            "de": "Erde"
          }
        }
      }
    }
  },
  {
    "@id": "kd-trans:NL-DE/grond-n-NL_SE00009938-sense-Erdboden-grond-n-NL_SE00009938-sense-TC00024572-trans",
    "@type": "ontolex:Translation",
    "target": {
      "@id": "kd-lex:DE/Erdboden-grond-n-NL_SE00009938-sense",
      "@type": "ontolex:LexicalSense",
      "reference": {
        "@id": "kd-base:NL_SE00009938-concept",
        "@type": "skos:Concept"
      },
      "sense_entry": {
        "@id": "kd-lex:DE/Erdboden",
        "@type": "ontolex:LexicalEntry",
        "form": {
          "@id": "kd-lex:DE/Erdboden-form",
          "@type": "ontolex:Form",
          "text": {
            "de": "Erdboden"
          }
        }
      }
    }
  }
],
]

```

```

"@id": "kd-trans:NL-DE/grond-n-NL_SE00009938-sense-Boden-grond-n-NL_SE00009938-sense-TC00024572-trans",
"@type": "ontolex:Translation",
"target": {
  "@id": "kd-lex:DE/Boden-grond-n-NL_SE00009938-sense",
  "@type": "ontolex:LexicalSense",
  "reference": {
    "@id": "kd-base:NL_SE00009938-concept",
    "@type": "skos:Concept"
  },
  "sense_entry": {
    "@id": "kd-lex:DE/Boden",
    "@type": "ontolex:LexicalEntry",
    "form": {
      "@id": "kd-lex:DE/Boden-form",
      "@type": "ontolex:Form",
      "text": {
        "de": "Boden"
      }
    }
  }
},
},
{
"@id": "kd-trans:NL-EN/grond-n-NL_SE00009938-sense-ground-grond-n-NL_SE00009938-sense-TC00024572-trans",
"@type": "ontolex:Translation",
"target": {
  "@id": "kd-lex:EN/ground-grond-n-NL_SE00009938-sense",
  "@type": "ontolex:LexicalSense",
  "reference": {
    "@id": "kd-base:NL_SE00009938-concept",
    "@type": "skos:Concept"
  },
  "sense_entry": {
    "@id": "kd-lex:EN/ground",
    "@type": "ontolex:LexicalEntry",
    "form": {
      "@id": "kd-lex:EN/ground-form",
      "@type": "ontolex:Form",
      "text": {
        "en": "ground"
      }
    }
  }
},
},
{
"@id": "kd-trans:NL-EN/grond-n-NL_SE00009938-sense-land-grond-n-NL_SE00009938-sense-TC00024572-trans",
"@type": "ontolex:Translation",
"target": {
  "@id": "kd-lex:EN/land-grond-n-NL_SE00009938-sense",
  "@type": "ontolex:LexicalSense",
  "reference": {
    "@id": "kd-base:NL_SE00009938-concept",
    "@type": "skos:Concept"
  },
  "sense_entry": {
    "@id": "kd-lex:EN/land",
    "@type": "ontolex:LexicalEntry",
    "form": {
      "@id": "kd-lex:EN/land-form",
      "@type": "ontolex:Form",
      "text": {
        "en": "land"
      }
    }
  }
},
},
{
"@id": "kd-trans:NL-ES/grond-n-NL_SE00009938-sense-tierra-grond-n-NL_SE00009938-sense-TC00024572-trans",
"@type": "ontolex:Translation",
"target": {
  "@id": "kd-lex:ES/tierra-grond-n-NL_SE00009938-sense",
  "@type": "ontolex:LexicalSense",
  "reference": {
    "@id": "kd-base:NL_SE00009938-concept",
    "@type": "skos:Concept"
  }
}
}

```

```

    },
    "sense_entry": {
      "@id": "kd-lex:ES/tierra",
      "@type": "ontolex:LexicalEntry",
      "form": {
        "@id": "kd-lex:ES/tierra-form",
        "@type": "ontolex:Form",
        "text": {
          "es": "tierra"
        }
      }
    }
  }
}
]
},
{
  "@id": "kd-lex:NL/grond-n-NL_SE00009939-sense",
  "@type": "ontolex:LexicalSense",
  "reference": {
    "@id": "kd-base:NL_SE00009939-concept",
    "@type": "skos:Concept",
    "definition": {
      "nl": "reden"
    }
  },
  "relation": {
    "@id": "kd-lex:NL/grond-n-NL_SE00009939-sense-argument-grond-n-NL_SE00009939-sense-sr",
    "@type": "ontolex:SenseRelation",
    "category": "lexinfo:synonym",
    "relates": {
      "@id": "kd-lex:NL/argument-grond-n-NL_SE00009939-sense",
      "@type": "ontolex:LexicalSense",
      "reference": {
        "@id": "kd-base:NL_SE00009939-concept",
        "@type": "skos:Concept"
      }
    },
    "sense_entry": {
      "@id": "kd-lex:NL/argument",
      "@type": "ontolex:LexicalEntry",
      "form": {
        "@id": "kd-lex:NL/argument-form",
        "@type": "Ontolex:Form",
        "text": {
          "nl": "argument"
        }
      }
    }
  }
}
},
"example": [
  {
    "@id": "kd-lex:NL/grond-n-NL_SE00009939-sense-TC00024575-ex",
    "@type": "kd:UsageExample",
    "value": {
      "nl": "op goede gronden iets kiezen"
    },
    "relation": {
      "@id": "TC00024575-trans-ex-cl",
      "@type": "kd:TranslationExampleCluster",
      "relates": [
        {
          "@type": "kd:UsageExample",
          "value": {
            "de": "etwas aus guten Gründen wählen"
          }
        },
        {
          "@type": "kd:UsageExample",
          "value": {
            "en": "to choose something with good reason"
          }
        }
      ],
      {
        "@type": "kd:UsageExample",
        "value": {
          "es": "elegir algo por motivos fundados"
        }
      }
    }
  }
]

```

```

    }
  }
]
},
"translation": [
  {
    "@id": "kd-trans:NL-ID/grond-n-NL_SE00009939-sense--grond-n-NL_SE00009939-sense-TC00024574-trans",
    "@type": "ontolex:Translation",
    "target": {
      "@id": "kd-lex:ID/-grond-n-NL_SE00009939-sense",
      "@type": "ontolex:LexicalSense",
      "reference": {
        "@id": "kd-base:NL_SE00009939-concept",
        "@type": "skos:Concept"
      },
    },
    "sense_entry": {
      "@id": "kd-lex:ID/",
      "@type": "ontolex:LexicalEntry",
      "form": {
        "@id": "kd-lex:ID/-form",
        "@type": "ontolex:Form"
      }
    }
  }
},
  {
    "@id": "kd-trans:NL-DE/grond-n-NL_SE00009939-sense-Grund-grond-n-NL_SE00009939-sense-TC00024574-trans",
    "@type": "ontolex:Translation",
    "target": {
      "@id": "kd-lex:DE/Grund-grond-n-NL_SE00009939-sense",
      "@type": "ontolex:LexicalSense",
      "reference": {
        "@id": "kd-base:NL_SE00009939-concept",
        "@type": "skos:Concept"
      },
    },
    "sense_entry": {
      "@id": "kd-lex:DE/Grund",
      "@type": "ontolex:LexicalEntry",
      "form": {
        "@id": "kd-lex:DE/Grund-form",
        "@type": "ontolex:Form",
        "text": {
          "de": "Grund"
        }
      }
    }
  }
},
  {
    "@id": "kd-trans:NL-EN/grond-n-NL_SE00009939-sense-ground-grond-n-NL_SE00009939-sense-TC00024574-trans",
    "@type": "ontolex:Translation",
    "target": {
      "@id": "kd-lex:EN/ground-grond-n-NL_SE00009939-sense",
      "@type": "ontolex:LexicalSense",
      "reference": {
        "@id": "kd-base:NL_SE00009939-concept",
        "@type": "skos:Concept"
      },
    },
    "sense_entry": {
      "@id": "kd-lex:EN/ground",
      "@type": "ontolex:LexicalEntry",
      "form": {
        "@id": "kd-lex:EN/ground-form",
        "@type": "ontolex:Form",
        "text": {
          "en": "ground"
        }
      }
    }
  }
},
  {
    "@id": "kd-trans:NL-EN/grond-n-NL_SE00009939-sense-reason-grond-n-NL_SE00009939-sense-TC00024574-trans",
    "@type": "ontolex:Translation",
    "target": {

```


de **treten**

es **dar un puntapié, golpear con el pie, dar una patada a**

nl **schoppen**

2 (of a gun) to jerk or spring back violently when fired.

de **zurückstoßen**

es **dar un culatazo**

nl **terugslag hebben**

◆ **noun**

1 a blow with the foot: *The boy gave him a kick on the ankle; He was injured by a kick from a horse.*

de **der Tritt**

es **puntapié, patada**

nl **schop**

2 the springing back of a gun after it has been fired.

de **der Rückstoß**

es **culatazo**

nl **terugslag**

3 a pleasant thrill: *She gets a kick out of entertaining people.*

de **mächtiger Spaß**

es **diversión, emoción**

nl **stimulans, kick**

kick about/around *phrasal verb*

to treat badly or bully: *The bigger boys are always kicking him around.*

de **schikanieren**

es **maltratar**

nl **ruw behandelen**

kick off *phrasal verb*

to start a football game by kicking the ball: *We kick off at 2.30; (noun 'kick-off': The kick-off is at 2.30)*

de **anstößen; der Anstoß**

es **sacar, hacer el saque inicial**

nl **afrappen**

kick up *phrasal verb*

to cause or start off (a fuss etc): *Don't start kicking up a fuss.*

de **verursachen**

es **armar follón, armar bronca**

nl **veroorzaken**

XML

```
<DictionaryEntry actionValue="unspecified">
  <ElementGrp identifier="PW00014096">
    <HeadwordCtn>
      <Headword from="11760015" soundfile="ePASSWORD.wav" to="11760450">kick</Headword>
      <PronunciationCtn>
        <Pronunciation index="PR00013417">kik</Pronunciation>
      </PronunciationCtn>
      <PartOfSpeechCtn>
        <PartOfSpeech>verb</PartOfSpeech>
      </PartOfSpeechCtn>
    </HeadwordCtn>
    <SenseCtn actionValue="unspecified">
      <SenseNumber>1</SenseNumber>
      <Sense identifier="PS00018306">
        <DefinitionCtn>
          <Definition>to hit or strike out with the foot</Definition>
          <ExampleGrp>
            <Example>The child kicked his brother</Example>
            <Example>He kicked the ball into the next garden</Example>
            <Example>He kicked at the locked door</Example>
            <Example>She kicked open the gate.</Example>
          </ExampleGrp>
        </DefinitionCtn>
      </Sense>
    </SenseCtn>
  </ElementGrp>
</DictionaryEntry>
```

```

    <TranslationCtn index="TC00018269" oldID="16698">
      <Translation lang="de" checkTrans="no">treten</Translation>
      <Translation lang="es" checkTrans="no">dar un puntapié, golpear con el pie, dar una patada
a</Translation>
      <Translation lang="nl" checkTrans="no">schoppen</Translation>
    </TranslationCtn>
  </Sense>
</SenseCtn>
<SenseCtn actionValue="unspecified">
  <SenseNumber>2</SenseNumber>
  <Sense identifier="PS00018307">
    <DefinitionCtn>
      <Definition>(of a gun) to jerk or spring back violently when
      fired.</Definition>
    </DefinitionCtn>
    <TranslationCtn index="TC00018270" oldID="16699">
      <Translation lang="de" checkTrans="no">zurückstoßen</Translation>
      <Translation lang="es" checkTrans="no">dar un culatazo</Translation>
      <Translation lang="nl" checkTrans="no">terugslag hebben</Translation>
    </TranslationCtn>
  </Sense>
</SenseCtn>
</ElementGrp>
<ElementGrp identifier="PW00014097">
  <HomonymsCtn>
    <PartOfSpeechCtn>
      <PartOfSpeech>noun</PartOfSpeech>
    </PartOfSpeechCtn>
    <SenseCtn actionValue="unspecified">
      <SenseNumber>1</SenseNumber>
      <Sense identifier="PS00018308">
        <DefinitionCtn>
          <Definition>a blow with the foot</Definition>
          <ExampleGrp>
            <Example>The boy gave him a kick on the ankle</Example>
            <Example>He was injured by a kick from a horse.</Example>
          </ExampleGrp>
        </DefinitionCtn>
        <TranslationCtn index="TC00018271" oldID="16700">
          <Translation lang="de" checkTrans="no">der Tritt</Translation>
          <Translation lang="es" checkTrans="no">puntapié, patada</Translation>
          <Translation lang="nl" checkTrans="no">schop</Translation>
        </TranslationCtn>
      </Sense>
    </SenseCtn>
  </HomonymsCtn>
  <SenseCtn actionValue="unspecified">
    <SenseNumber>2</SenseNumber>
    <Sense identifier="PS00018309">
      <DefinitionCtn>
        <Definition>the springing back of a gun after it has been
        fired.</Definition>
      </DefinitionCtn>
      <TranslationCtn index="TC00018272" oldID="16701">
        <Translation lang="de" checkTrans="no">der Rückstoß</Translation>
        <Translation lang="es" checkTrans="no">culatazo</Translation>
        <Translation lang="nl" checkTrans="no">terugslag</Translation>
      </TranslationCtn>
    </Sense>
  </SenseCtn>
  <SenseCtn actionValue="unspecified">
    <SenseNumber>3</SenseNumber>
    <Sense identifier="PS00018310">
      <DefinitionCtn>
        <Definition>a pleasant thrill</Definition>
        <ExampleGrp>
          <Example>She gets a kick out of entertaining people.</Example>
        </ExampleGrp>
      </DefinitionCtn>
      <TranslationCtn index="TC00018273" oldID="16702">
        <Translation lang="de" checkTrans="no">mächtiger Spaß</Translation>
        <Translation lang="es" checkTrans="no">diversión, emoción</Translation>
        <Translation lang="nl" checkTrans="no">stimulans, kick</Translation>
      </TranslationCtn>
    </Sense>
  </SenseCtn>
</ElementGrp>

```

```

<ElementGrp identifier="PW00014098">
  <SubHeadword>
    <HeadwordCtn>
      <Headword from="11760550" soundfile="ePASSWORD.wav" to="11762152">kick about/around</Headword>
      <PartOfSpeechCtn>
        <PartOfSpeech>phrasal verb</PartOfSpeech>
      </PartOfSpeechCtn>
    </HeadwordCtn>
  </SubHeadword>
  <SenseCtn actionValue="unspecified">
    <Sense identifier="PS00018311">
      <DefinitionCtn>
        <Definition>to treat badly or bully</Definition>
        <ExampleGrp>
          <Example>The bigger boys are always kicking him around.</Example>
        </ExampleGrp>
      </DefinitionCtn>
      <TranslationCtn index="TC00018274" oldID="16703">
        <Translation lang="de" checkTrans="no">schikanieren</Translation>
        <Translation lang="es" checkTrans="no">maltratar</Translation>
        <Translation lang="nl" checkTrans="no">ruw behandelen</Translation>
      </TranslationCtn>
    </Sense>
  </SenseCtn>
</ElementGrp>
<ElementGrp identifier="PW00014099">
  <SubHeadword>
    <HeadwordCtn>
      <Headword from="11762252" soundfile="ePASSWORD.wav" to="11762890">kick off</Headword>
      <PartOfSpeechCtn>
        <PartOfSpeech>phrasal verb</PartOfSpeech>
      </PartOfSpeechCtn>
    </HeadwordCtn>
  </SubHeadword>
  <SenseCtn actionValue="unspecified">
    <Sense identifier="PS00018312">
      <DefinitionCtn>
        <Definition>to start a football game by kicking the ball</Definition>
        <ExampleGrp>
          <Example>We kick off at 2.30</Example>
          <Example>(<POS>noun</POS><b2>'kick-off</b2>: The kick-off is at 2.30)</Example>
        </ExampleGrp>
      </DefinitionCtn>
      <TranslationCtn index="TC00018275" oldID="16704">
        <Translation lang="de" checkTrans="no">anstoßen; der Anstoß</Translation>
        <Translation lang="es" checkTrans="no">sacar, hacer el saque inicial</Translation>
        <Translation lang="nl" checkTrans="no">aftrappen</Translation>
      </TranslationCtn>
    </Sense>
  </SenseCtn>
</ElementGrp>
<ElementGrp identifier="PW00014100">
  <SubHeadword>
    <HeadwordCtn>
      <Headword from="11765127" soundfile="ePASSWORD.wav" to="11765632">kick up</Headword>
      <PartOfSpeechCtn>
        <PartOfSpeech>phrasal verb</PartOfSpeech>
      </PartOfSpeechCtn>
    </HeadwordCtn>
  </SubHeadword>
  <SenseCtn actionValue="unspecified">
    <Sense identifier="PS00018313">
      <DefinitionCtn>
        <Definition>to cause or start off (a fuss  
<i0>etc</i0)</Definition>
        <ExampleGrp>
          <Example>Don't start kicking up a fuss.</Example>
        </ExampleGrp>
      </DefinitionCtn>
      <TranslationCtn index="TC00018276" oldID="16705">
        <Translation lang="de" checkTrans="no">verursachen</Translation>
        <Translation lang="es" checkTrans="no">armar follón, armar bronca</Translation>
        <Translation lang="nl" checkTrans="no">veroorzaken</Translation>
      </TranslationCtn>
    </Sense>
  </SenseCtn>
</ElementGrp>

```

```

</Sense>
</SenseCtn>
</ElementGrp>
</DictionaryEntry>

```

KMT : ES-EN - DE, NL

HTML

juntar *verb*

1. gather *verb*

to (cause to) come together in one place

◊ *A crowd of people gathered near the accident.*

{de} - versammeln

{nl} - (doen) bijeenkomen

2. gather together *phrasal verb*

to come or bring together, in a group

◊ *He gathered his books and papers together.*

{de} - zusammensammeln

{nl} - bijeenkomen, bijeenzoeken

3. join *verb*

(often with up, onetc) to put together or connect

◊ *The electrician joined the wires (up) wrongly* ◻ *You must join this piece (on) to that piece* ◻ *He joined the two stories together to make a play* ◻ *The island is joined to the mainland by a sandbank at low tide.*

{de} - verbinden

{nl} - verbinden

4. rally *verb*

to come or bring together for a joint action or effort

◊ *The supporters rallied to save the club from collapse* ◻ *The politician asked his supporters to rally to the cause.*

{de} - sich zusammentun

{nl} - zich verzamelen

5. lump *verb*

(usually with together) to treat or think of as (all) alike

◊ *She lumped all men together, and didn't trust any of them.*

{de} - in einen Topf werfen

{nl} - op een hoop gooien

6. pool *verb*

to put together for general use

◊ *We pooled our money and bought a caravan that we could all use.*

{de} - zusammenlegen

{nl} - bij elkaar leggen

XML

```

<DictionaryEntry identifier="EN00012407">
  <HeadwordCtn>
    <Headword>juntar</Headword>
    <PartOfSpeech>verb</PartOfSpeech>
  </HeadwordCtn>
  <SenseBlock>
    <SenseCtn id="SE00025240" num="">
      <EnCtn>
        <EnHeadword>gather</EnHeadword>
        <EnPOS>verb</EnPOS>
        <DefinitionCtn>
          <Definition>to (cause to) come together in one place</Definition>
        </DefinitionCtn>
        <ExampleGrp>
          <Example>A crowd of people gathered near the accident.</Example>
        </ExampleGrp>
      </EnCtn>
      <TranslationCtn index="TR00025240">
        <Translation lang="de">versammeln</Translation>
        <Translation lang="nl">(doen) bijeenkomen</Translation>
      </TranslationCtn>
    </SenseCtn>
  </SenseBlock>
</DictionaryEntry>

```

```

</TranslationCtn>
</SenseCtn>
<SenseCtn id="SE00025241" num="">
  <EnCtn>
    <EnHeadword>gather together</EnHeadword>
    <EnPOS>phrasal verb</EnPOS>
    <DefinitionCtn>
      <Definition>to come or bring together, in a group</Definition>
    </DefinitionCtn>
    <ExampleGrp>
      <Example>He gathered his books and papers together.</Example>
    </ExampleGrp>
  </EnCtn>
  <TranslationCtn index="TR00025241">
    <Translation lang="de">zusammensammeln</Translation>
    <Translation lang="nl">bijeenkomen, bijeenzoeken</Translation>
  </TranslationCtn>
</SenseCtn>
<SenseCtn id="SE00025242" num="">
  <EnCtn>
    <EnHeadword>join</EnHeadword>
    <EnPOS>verb</EnPOS>
    <DefinitionCtn>
      <Definition>(often with up, onetc) to put together or connect</Definition>
    </DefinitionCtn>
    <ExampleGrp>
      <Example>The electrician joined the wires (up) wrongly</Example>
      <Example>You must join this piece (on) to that piece</Example>
      <Example>He joined the two stories together to make a play</Example>
      <Example>The island is joined to the mainland by a sandbank at low tide.</Example>
    </ExampleGrp>
  </EnCtn>
  <TranslationCtn index="TR00025242">
    <Translation lang="de">verbinden</Translation>
    <Translation lang="nl">verbinden</Translation>
  </TranslationCtn>
</SenseCtn>
<SenseCtn id="SE00025243" num="">
  <EnCtn>
    <EnHeadword>rally</EnHeadword>
    <EnPOS>verb</EnPOS>
    <DefinitionCtn>
      <Definition>to come or bring together for a joint action or effort</Definition>
    </DefinitionCtn>
    <ExampleGrp>
      <Example>The supporters rallied to save the club from collapse</Example>
      <Example>The politician asked his supporters to rally to the cause.</Example>
    </ExampleGrp>
  </EnCtn>
  <TranslationCtn index="TR00025243">
    <Translation lang="de">sich zusammentun</Translation>
    <Translation lang="nl">zich verzamelen</Translation>
  </TranslationCtn>
</SenseCtn>
<SenseCtn id="SE00025244" num="">
  <EnCtn>
    <EnHeadword>lump</EnHeadword>
    <EnPOS>verb</EnPOS>
    <DefinitionCtn>
      <Definition>(usually with together) to treat or think of as (all) alike</Definition>
    </DefinitionCtn>
    <ExampleGrp>
      <Example>She lumped all men together, and didn't trust any of them.</Example>
    </ExampleGrp>
  </EnCtn>
  <TranslationCtn index="TR00025244">
    <Translation lang="de">in einen Topf werfen</Translation>
    <Translation lang="nl">op een hoop gooien</Translation>
  </TranslationCtn>
</SenseCtn>
<SenseCtn id="SE00025245" num="">
  <EnCtn>
    <EnHeadword>pool</EnHeadword>
    <EnPOS>verb</EnPOS>
    <DefinitionCtn>
      <Definition>to put together for general use</Definition>
    </DefinitionCtn>

```

```

<ExampleGrp>
  <Example>We pooled our money and bought a caravan that we could all use.</Example>
</ExampleGrp>
</EnCtn>
<TranslationCtn index="TR00025245">
  <Translation lang="de">zusammenlegen</Translation>
  <Translation lang="nl">bij elkaar leggen</Translation>
</TranslationCtn>
</SenseCtn>
</SenseBlock>
</DictionaryEntry>

```

RANDOM HOUSE Webster's College Dictionary

HTML

ma•chine /mə'jɪn/ , -chined, -chin•ing.

1. an apparatus consisting of interrelated parts with separate functions, used in the performance of some kind of work: *a sewing machine*.
2. **a.** a device that transmits or modifies force or motion. **b.** Also called **simple machine**, any of several elementary mechanisms, as the lever, wheel and axle, pulley, wedge, or inclined plane. **c.** a combination of simple machines.
3. an automobile or airplane.
4. any of various apparatus, devices, etc., that dispense things, esp. a vending machine.
5. any complex agency or operating system: *the machine of government*.
6. a group of persons that conducts or controls a political party or organization.
7. a person or thing that acts in a mechanical or automatic manner.
8. a mechanical contrivance formerly used for producing stage effects.
9. a literary contrivance introduced for special effect.

10. to make, prepare, or finish with a machine or machine tool.

[1540–50; < F < L *māchina* < Attic Gk *mēchanē*; see MECHANICmagh-1]

XML

```

<DictionaryEntry OriginalCode="cdr2.03:machine" identifier="RDE00040551">
  <HeadwordCtn>
    <Headword>machine</Headword>
    <SearchForm>machine</SearchForm>
    <EntryBody hi="12">
      <HeadwordDisplay>ma•chine</HeadwordDisplay>
      <PronunciationCtn Id="PR00035489">
        <Pronunciation>mə'jɪn</Pronunciation>
      </PronunciationCtn>
      <PreDefinition>
        <PartOfSpeechCtn>
          <PartOfSpeechBody value="noun" />
          <PartOfSpeechBody value="verb" />
        </PartOfSpeechCtn>
        <InflectionCtn>
          <Inflection>-chined, -chin•ing.</Inflection>
        </InflectionCtn>
      </PreDefinition>
      <DefinitionCtn Id="DC00041999">
        <EmPartOfSpeech>
          <PartOfSpeechBody value="noun" />
        </EmPartOfSpeech>
        <Definition id="1" identifier="RSE00087917" SenseNumber="1">
          <CategoryCtn>
            <Category>MAC</Category>
          </CategoryCtn>
          <DefBody>an apparatus consisting of interrelated parts with separate functions, used in the performance
of some kind of work:</DefBody>
          <Example>a sewing machine.</Example>
        </Definition>
        <Definition id="3" hi="3" identifier="RSE00087918" SenseNumber="2">
          <SubDefinition id="3a">
            <CategoryCtn>
              <Category>MEC</Category>
            </CategoryCtn>
            <DefBody>a device that transmits or modifies force or motion.</DefBody>
          </SubDefinition>
          <SubDefinition id="3b">

```

```

<CategoryCtn>
  <Category>MEC</Category>
</CategoryCtn>
<CrossReferenceCtn>
  <ReferencePreface>Also called</ReferencePreface>
  <Target>
    <TargetDisplay>
      <b>simple machine.</b>
    </TargetDisplay>
    <ReferenceTo>
      <ReferenceHW>simplemachine</ReferenceHW>
    </ReferenceTo>
  </Target>
</CrossReferenceCtn>
<DefBody>any of several elementary mechanisms, as the lever, wheel and axle, pulley, wedge, or
inclined plane.</DefBody>
</SubDefinition>
<SubDefinition id="3c">
  <CategoryCtn>
    <Category>MEC</Category>
  </CategoryCtn>
  <DefBody>a combination of simple machines.</DefBody>
</SubDefinition>
</Definition>
<Definition id="4" hi="1" identifier="RSE00087919" SenseNumber="3">
  <SubDefinition id="4a">
    <DefBody>an automobile or airplane.</DefBody>
  </SubDefinition>
</Definition>
<Definition id="6" identifier="RSE00087920" SenseNumber="4">
  <CategoryCtn>
    <Category>CVB</Category>
  </CategoryCtn>
  <DefBody>any of various apparatus, devices, etc., that dispense things, esp. a vending
machine.</DefBody>
</Definition>
<Definition id="7" identifier="RSE00087921" SenseNumber="5">
  <DefBody>any complex agency or operating system:</DefBody>
  <Example>the machine of government.</Example>
</Definition>
<Definition id="8" identifier="RSE00087922" SenseNumber="6">
  <CategoryCtn>
    <Category>GOV</Category>
  </CategoryCtn>
  <DefBody>a group of persons that conducts or controls a political party or organization.</DefBody>
</Definition>
<Definition id="9" identifier="RSE00087923" SenseNumber="7">
  <DefBody>a person or thing that acts in a mechanical or automatic manner.</DefBody>
</Definition>
<Definition id="10" identifier="RSE00087924" SenseNumber="8">
  <DefBody>a mechanical contrivance formerly used for producing stage effects.</DefBody>
</Definition>
<Definition id="11" identifier="RSE00087925" SenseNumber="9">
  <CategoryCtn>
    <Category>LIT</Category>
  </CategoryCtn>
  <DefBody>a literary contrivance introduced for special effect.</DefBody>
</Definition>
<EmPartOfSpeech>
  <PartOfSpeechBody value="verb-transitive" />
</EmPartOfSpeech>
<Definition id="12" identifier="RSE00087926" SenseNumber="10">
  <CategoryCtn>
    <Category>MAC</Category>
  </CategoryCtn>
  <DefBody>to make, prepare, or finish with a machine or machine tool.</DefBody>
</Definition>
</DefinitionCtn>
<Etymology>[1540-50; &lt; F &lt; L <i>māchina</i> &lt; Attic Gk <i>mēchanē</i>; see
<Suppercase>mechanic</Suppercase><root>magh-1</sup></root>]</Etymology>
</EntryBody>
</HeadwordCtn>
</DictionaryEntry>

```

Appendix 4. JSON-LD conversion of DIV

Data is often stored and distributed in internal formats, designed to cover the needs of the specific application and the technology stack available to developers². Even when the data is available in a parseable format (CSV, XML, JSON, etc), it is still far from being machine-readable. In some cases there is descriptive metadata provided, however it's usually only targeted to developers to read when implementing yet-another-parser for the data.

JSON is the data format of modern REST Web-APIs³. However usually each API uses a custom defined JSON schema to describe its data. This issue can be addressed by normalization using JSON-LD. JSON-LD is a JSON based syntax to express Linked Data⁴. The main design concept in JSON-LD is the “context”, which provides a mapping of JSON properties to an RDF vocabulary, giving meaning to data. In simple cases this can be achieved even without changing the original JSON structure.

Consider the following example:

```
{  
  "first_name": "Joe",  
  "last_name": "Doe",  
  "alias": "UknonowMan",  
  "email": "joe.doe@example.com"  
}
```

This could be the API output of a social network platform describing a person. This data can be described using the Schema.org ontology, simply by providing a “@context” that maps the key names of the JSON structure to specific Schema.org predicates. The “@context” object can be added directly into the JSON data as a discrete object, or it can refer to a URL where the context is provided as a remote context.

```
{  
  "@context": {  
    "@vocab": "http://schema.org/",  
    "first_name": "givenName",  
    "last_name": "familyName",  
    "alias": "alternateName",  
    "email": "email"  
  },  
  "first_name": "Joe",
```

²<https://developer.ibm.com/dwblog/2014/webizing-database-json/>

³<https://blog.codeship.com/json-ld-building-meaningful-data-apis/>

⁴<https://json-ld.org/spec/latest/json-ld/>


```
"last_name": "Doe",  
"alias": "UknonowMan",  
"email": "joe.doe@example.com"  
}
```

Using a JSON-LD parser, this JSON-LD can be parsed and stored as RDF data according to the JSON-LD 1.0 Processing Algorithms and API⁵.

```
@prefix schema: <http://schema.org/> .
```

```
[]
```

```
  schema:alternateName "UknonowMan" ;  
  schema:email "joe.doe@example.com" ;  
  schema:familyName "Doe" ;  
  schema:givenName "Joe" .
```

The complete context object defined for mapping the Lexicala API vocabularies is included below and is accessible through the web to be used as remote context⁶.

```
"@context":{  
  "kd":"http://lexicala.com/def/ontolexKD#",  
  "kd-base":"http://lexicala.com/id/global/",  
  "lexinfo":"http://www.lexinfo.net/ontology/2.0/lexinfo#",  
  "lime":"http://www.w3.org/ns/lemon/lime#",  
  "ontolex":"http://www.w3.org/ns/lemon/ontolex#",  
  "owl":"http://www.w3.org/2002/07/owl#",  
  "rdf":"http://www.w3.org/1999/02/22-rdf-syntax-ns#",  
  "rdfs":"http://www.w3.org/2000/01/rdf-schema#",  
  "skos":"http://www.w3.org/2004/02/skos#",  
  "synsem":"http://www.w3.org/ns/ontolex-synsem#",  
  "vartrans":"http://www.w3.org/ns/lemon/vartrans#",  
  "xsd":"http://www.w3.org/2001/XMLSchema#",  
  "dc":"http://purl.org/dc/elements/1.1/",  
  "lexicog":"http://www.w3.org/ns/lemon/lexicog#",  
  "void":"http://rdfs.org/ns/void#",  
  "lexicographicEntryIn":{  
    "@reverse":"lexicog:entry"  
  },  
  "describes":{  
    "@id":"lexicog:describes"  
  },  
  "dictionaryEntryId":{  
    "@id":"kd:dictionaryEntryId"  
  },  
  "language":{  
    "@id":"dc:language"  
  },  
  "limeLanguage":{  
    "@id":"lime:language"  
  },  
  "pos":{  
    "@id":"lexinfo:partOfSpeech"  
  },  
  "headword":{  
    "@id":"ontolex:lexicalForm"  
  }  
}
```

⁵ <https://www.w3.org/TR/json-ld-api/>

⁶ https://api.lexicala.com/contexts/entry_context.json

```
},
"text":{
  "@id":"ontolex:writtenRep",
  "@container":"@language"
},
"pronunciation":{
  "@id":"ontolex:phoneticRep",
  "@container":"@language"
},
"senses":{
  "@id":"ontolex:sense"
},
"lexicalizedSense":{
  "@id":"ontolex:isLexicalizedSenseOf"
},
"definition":{
  "@id":"skos:definition",
  "@container":"@language"
},
"reversedRelates":{
  "@reverse":"vartrans:relates"
},
"category":{
  "@id":"vartrans:category"
},
"relates":{
  "@id":"vartrans:relates"
},
"lexiconSenseToEntry":{
  "@reverse":"ontolex:sense"
},
"source":{
  "@id":"dc:source"
},
"entryIn":{
  "@reverse":"lime:entry"
},
"examples":{
  "@id":"lexicog:usageExample"
},
"value":{
  "@id":"rdf:value",
  "@container":"@language"
},
"translations":{
  "@reverse":"vartrans:source"
},
"target":{
  "@id":"vartrans:target"
},
"tranSet":{
  "@reverse":"vatrans:trans"
},
"compositionalPhrases":{
  "@id":"vatrans:senseRel"
}
}
```