Digital Representation of Rights for Language Resources

Victor Rodriguez-Doncel

Ontology Engineering Group
Universidad Politécnica de Madrid
`vrodriguez@delicias.dia.upm.es`

Penny Labropoulou

Institute for Language and Speech Processing
Athena RC - Athens
`penny@ilsp.athena-innovation.gr`

Abstract

Language resources are very often valuable assets which are offered to the public under the terms of licenses that determine which uses are allowed and under which circumstances. These licenses have been typically published as natural language texts whose specific contents cannot be easily processed by a computer. This paper proposes a structured representation for the most commonly used licenses for language resources, reusing existing vocabularies and extending the Open Digital Rights Language core model. Examples and guidelines to use the 'Rights Information for Language Resources' vocabulary are given.

1 Introduction

Computational Linguistics started some 50 years ago studying natural language from a computational perspective. The need for Language Resources (LRs), such as lexica, thesauri, terminologies and corpora, was soon appreciated. At first, LRs producers created them mainly for their own use; however it was soon clear that LRs with a minimum size and quality, as those required for the advancement of Computational Linguistics and related disciplines could only live in a sharing paradigm, with LRs being created, distributed, used, re-used, extended and enriched in a shared environment.

LRs were offered to other users, following various distribution models: some LR producers publishing and promoting their resources themselves, either through their institutional sites or through sites dedicated to particular LRs, other producers forming alliances together with other interested parties in order to distribute but also to create new resources (e.g. LDC[1]) or passing on the distribution of their resources to dedicated agencies (e.g. ELRA/ELDA[2], TST-Centrale[3]) etc. The majority of LRs were offered for research and educational purposes, at no cost or for a minimal fee, especially when produced by public funding. The situation, however, changed mainly as the development of Language Technology led to the appearance of profitable business, which also led to the realization that LRs could also be a source of profit. As a consequence, some of the LR publishers have opted to market their LRs (or the rights thereof), thus making licensing an indispensable aspect in the distribution of LRs.

When discussing about licensing of LRs, two are the main dimensions that need to be taken into account: (a) the license itself, either in the form of a proper legal document, or some loosely expressed legal notice, (b) the clear indication of the licensing terms on the LR, in the form of free text or conventional metadata.

One of the priorities set by the FLARENET Strategic Research Agenda (Soria et al., 2014) is the availability of LRs "within an adequate IPR[4] and legal framework". The recommendations include the elaboration of "specific, simple and *harmonised licensing solutions* for data resources", taking into account licensing schemes already in use and simplifying them through broad-based solutions for both R & D and industry, and the adoption of *electronic licensing* and adaptation of current distribution models to new media (web, mobile devices etc.).

The digital formulation of rights and the standardisation of the licensing vocabulary have a number of advantages such as:

- improvement of the understanding of the li-

---

[1] `https://www.ldc.upenn.edu/`
[2] `http://www.elra.info/en/`
[3] `http://tst-centrale.org/nl/home`
[4] IPR: Intellectual Property Rights

censing terms by human users: although licenses are natural language texts, the legal jargon is quite complicated and not easily understood by newcomers. A harmonised vocabulary for licensing terms favours universal understanding of their precise meaning; moreover, the non-flat structure of digital rights information also favours the understanding of the different modalities (e.g. 'free if used for research', but 'non-free if used for commercial purposes')

- processing of the licensing terms by machines; this is extremely important in a re-use scenario of LRs, whereby they can be automatically processed by web services, combined with other LRs, extended and enriched: only LRs that allow such actions should be involved in these activities; and this can only be asserted if rights are expressed in a way understood by machines

- enhancement of the discovery of LRs that allow/forbid particular conditions of use through filtered browsing of LR catalogs based on criteria such as "license", "conditions of use" and "access rights"

- better management of the LRs by publishers, who have a clearer account on which rights have been granted to which resources.

Among the digital structures for representing the rights information, RDF is the one which best favours interoperability. The emergence of the Linked Data paradigm as a manner of publishing LRs on the web urged the publication of licensing information as Linked Data as well. This paper describes a language for expressing rights information for LRs as RDF, starting by the groundings in Section 2 (reviewing the existent practice and the requirements collected), continuing with the ontology in Section 3 and finalizing with examples and conclusions in Section 4 and 5 respectively.

2 Motivation for a common model

2.1 Rights information in LR repositories

LRs are in general considered intellectual property works, and as such they are protected by copyright laws: they should not be used in violation of the terms set by the rights holders. The terms of use declare the actions that are authorized (e.g.

whether they allow derivation, redistribution) and the applicable conditions (e.g. whether they require attribution, payment of a fee). The terms are included in the documentation of most LRs, but their automatic retrieval and processing is difficult because of the many forms they adopt: rights information may appear either as a textual notice or as structured metadata elements, it may consist of a mere reference to a well-known license (like an Open Data Commons or Creative Commons license), or it may point to a license drafted in a non-English language to be used solely for the specific resource. These heterogeneous practices prevent the automated processing of rights information.

Recently, we witness the proliferation of repositories collecting LRs and their metadata descriptions from various communities and sources according to different harvesting methodologies, and publishing them into homogeneous catalogs. The most relevant initiatives for our discussion are: META-SHARE[5] (Piperidis, 2012), CLARIN[6], LRE-Map[7] (Calzolari et al., 2012), OLAC[8] (Simons and Bird, 2003) and Datahub.io[9].

Taking a closer look at the rights metadata present in these catalogs, we see the following tendencies:

- catalogs where the rights information is loosely represented as a free text metadata element: this is mainly the case for portals harvesting from various sources, such as OLAC, the LRE Map and the CLARIN Virtual Language Observatory (VLO[10]); the reason for this is the fact that the sources do not oblige the depositors to document the access rights and/or allow them to use natural language statements for that (e.g. "free for research", "available at resource owner's site", "Public domain resource" etc.); this is also due to the fact that they include resources whose licenses are not available over the internet (e.g. resources from older times, when licenses were not standardised and providers asked legal experts to draft specific contracts for each resource, which were made available only to interested parties upon request); for the LRE Map, this practice has been dictated by the

fact that the metadata are submitted by authors of papers in conferences (e.g. LREC) describing the resources connected to their publication, which may still be under construction and/or not yet be available for distribution with specific licenses, so they simply indicate their intentions;

- catalogs where the rigths information is represented with a controlled vocabulary of values referring to standard licenses; this is the case of META-SHARE and partly Datahub and the CLARIN network repositories; in the case of Datahub, when registering a new dataset, providers can choose a license from a list, but also licensing information can be found in the VoID description of the dataset or even within the dataset itself. In META-SHARE, the provider is also forced to select for the license element among a controlled list of values corresponding to recommended standard licenses[11]; this element (as described in the following section) is part of a more complicated set of metadata elements describing the distribution conditions of the LR. In the case of CLARIN, there is a set of recommended licenses that LR providers are asked to use when depositing their resources in the repositories of the infrastructure, but legacy data can of course come with their own licenses; to help users understand the access rights, licenses are classified to one of three categories: those that can be publicly distributed (PUB), those permitting only academic use, i.e. use for research and educational purposes and which require user authentication, i.e. that users' identity is known (ACA) and those which impose additional restrictions or whose use requires additional consent from the rightsholder (RES); the use of easy-to-understand icons and symbols (e.g. a money icon for resources distributed with-a-fee) is recommended (Oksanen and Lindn, 2011).

- faceted browsing with the criterion of access rights/ license is a feature integrated in most of these catalogs but it is actually useful mostly when the set of values is limited to a manageable number of values that users can browse through; in addition, META-SHARE allows faceted browsing with a filter for conditions of use (e.g. whether the license allows commercial use, derivatives etc.)

The most recent initiative in this line is the Linghub portal[12], supported by the European LIDER project, which collects metadata from some of the repositories mentioned before (META-SHARE, CLARIN, Datahub.io and LRE Map) and publishes the records as Linked Data. All licensing information present in the original metadata records is harvested and collected together in the element "rights", bringing together license names, urls, free text statements etc. The work presented in this paper is related to this effort and the need for a common licensing metadata framework (McCrae et al., 2015).

2.2 Rights information in the META-SHARE model

The META-SHARE (MS) metadata schema constitutes an essential ingredient of the META-SHARE infrastructure, which is an open, integrated, secure and interoperable exchange infrastructure where LRs are documented, uploaded, stored, catalogued, announced, downloaded, exchanged and discussed, aiming to support reuse of LRs (Piperidis, 2012). The MS schema is a complex but rich model and, most important for our work, provides extensive support for the detailed representation of licensing information, making a remarkable effort that in some regards goes beyond of what has been described by license-specialized models. In consequence, the MS model has been taken as a basis for the rest of this work.

The original META-SHARE metadata model (Gavrilidou et al., 2012) [13] has been implemented as an XML Schema[14]. The META-SHARE schema encodes information about the whole life-cycle of the LR from production to usage. The central entity of the schema is the LR per se, which encompasses both datasets and technologies used for their processing. In addition to the

---

[11]http://www.meta-net.eu/meta-share/licenses

[12]http://linghub.lider-project.eu/

[13]Documentation and User Manual of the META-SHARE Metadata Model, found at http://www.meta-net.eu/meta-share/META-SHARE\%20\%20documentationUserManual.pdf

[14]Schemas can be found at github https://github.com/metashare/META-SHARE/tree/master/misc/schema/v3.0

central entity, other entities are also documented in the schema; these are reference *documents* related to the LR (papers, reports, manuals etc.), *persons/organizations* involved in its creation and use (creators, distributors etc.), related *projects* and activities (funding projects, activities of usage etc.) and accompanying *licenses*, all described with metadata taken as far as possible from relevant schemas and guidelines (e.g. BibTex for bibliographical references). The five root entities are represented as boxes in Figure 1. The META-SHARE schema proposes a set of elements to encode specific descriptive features of each of these entities and relations holding between them, taking as a starting point the LR. Following the CMDI approach (Broeder et al., 2012), these elements are grouped together into "components". The core of the schema is the *resourceInfo* component, which subsumes

- administrative components relevant to all LRs, e.g. *identificationInfo* (name, description and identifiers), *usageInfo* (information about the intended and actual use of the LR);

- components specific to the relevant resource and media type combinations, e.g. text or audio parts of corpora, lexical/conceptual resources etc., such as language, formats, classification etc.

The META-SHARE schema recognises obligatory elements (minimal version) and recommended and optional elements (maximal version).
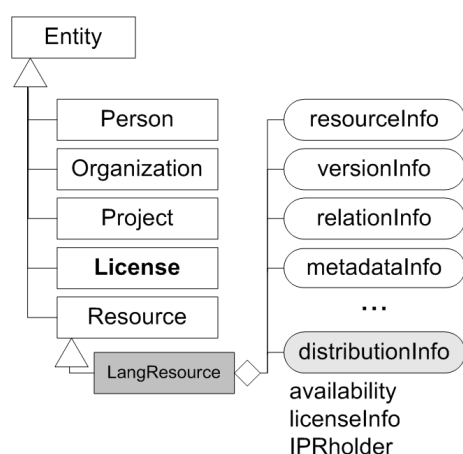

Figure 1: Main entities in the MS model

For our discussion, the most relevant component is the *distributionInfo* which brings together all information related to licensing and IPR issues, e.g. the IPR holder(s), the distribution rights holder(s), availability status (i.e. whether the LR is available for access, with or without restrictions); the embedded *licenseInfo* component encodes all information related to the licensing terms, e.g. the license short name and specific terms and conditions, the medium with which the LR can be accessed (i.e. whether it cam be downloaded or used via an i/f etc.). Each resource may be linked to one or more *licenseInfo* components, in case the same resource is made available under different formats and/or licensing conditions (e.g. for free for non-commercial purposes vs. at a price for commercial purposes, downloadable for commercial users vs. accessible through interface for academic users).

In the framework of the LD4LT group, the META-SHARE model has been the base for the development of an ontology in OWL; the MS/OWL ontology has been based on the on the ontology developed by Villegas et al. (Villegas et al., 2014) (covering part of the original schema) and extended to the complete schema (in order to cover all relevant LRs) (McCrae et al., 2015). The transformation from the XSD schema to the OWL ontology involved the transformation of components to classes and that of elements to properties[15].

3 The Rights Information for Language Resources Ontology

In the course of this activity, the original module of licensing and rights information has been re-structured (in order to better accommodate RDF modelling considerations) and enhanced with RELs, capable of describing rights information in a generally understood manner. RELs also provide a hierarchical organization for the rights information whose structure more naturally depicts dual licenses, nested permissions and the relationship between conditions and rights. In addition, some other vocabularies like CreativeCommons'[16] or the price specification with GoodRelations have been considered.

The licensing and rights module as perceived in the model has also been released as a separate ontology ("Rights Information for Language Resources" ontology) at:

---

[15]This is an simplified description of the actual transformation process; for more on this, see (McCrae et al., 2015)

[16]http://creativecommons.org/ns#

The rights ontology builds upon the META-SHARE schema for the LanguageResource and the Distribution classes and for the License class integrates elements of the ODRL model. In fact, the ontology revolves around three entities/classes:

- the *Language Resource*, perceived in the same way as in the original MS model;

- the *Distribution*, which comes from the original *distributionInfo* component but is remodeled and adapted to the concept of the *dcat:Distribution*[17] class; thus, it now represents an accessible form of an LR, which for instance can be available through different delivery channels (e.g. as a downloadable file, on a CD-ROM or accessible via an interface), in different forms (e.g. as a csv or txt file), through different distributors and with different licensing terms;

- the License, coming from the *licenseInfo* component.

The elements included in the *distributionInfo* and *licenseInfo* components have been transformed to OWL object and datatype properties, while a careful study has been made in order to attach them to the appropriate classes. For instance, the *iprHolder* which was included in the *distributionInfo* component has been attached to the *Language Resource* class, given that this is a property that remains the same irrespective of the different forms of access an LR may take; the *distributionRightsHolder*, however, may differ for different forms and is thus attached to the *Distribution* class. Similarly, there has been a careful separation of the elements included in the *licenseInfo* between properties attached to the license and those moved to the *Distribution* class. Here, the main consideration was to detach the License class from Language resources, in an effort to generalize over them and standardize their representation as far as possible. By attaching, for instance, the exact sum to be paid for the acquisition of an LR to the Distribution class while the information that a payment is due on the license class, we can re-use the same

license representation for all LRs distributed under this condition.

We have also introduced additional properties (e.g. *licenseCategory*, *licenseName* and *licenseURL*) and individuals (*languageEngineeringResearch* for the *ConditionsOfUse*).

Licenses represented with the Rights Information for Language Resources ontology permit a dual representation of the information: preserving the META-SHARE elements and structure and/or adhering to the ODRL schema. Both are compatible and satisfy different requirements. Redundancy is the preferred option, but expressing rights information in either manner is acceptable. This section describes both alternatives, introducing first the ODRL-style and then the schema inherited from META-SHARE.

3.1 Rights Expressions in ODRL

ODRL 2.1[18] is a policy and rights expression language suitable to represent the licensing terms of the language resources. ODRL specifies both an abstract core model and a common vocabulary, which can be extended for the particular domains ODRL is applied to. There have been ODRL profiles for representing contents' rights in mobile devices (OMA DRM), for the news industry (RightsML by IPTC), for the eBook (ONIX) and for general Creative Commons licenses, but no specific terms exist for the language resources domain. ODRL 2.0 can be serialized in XML, JSON and RDF. The latter serialization is based on the ODRL 2.1 Ontology (McRoberts and Rodriguez-Doncel (eds.), 2015).

The main entities in the ODRL Core Model[19] are presented in Figure 2. An ODRL *policy* is a set of *rules*, which can be *permissions*, *prohibitions* or *duties*. Permissions allow executing certain *actions* over an *asset*, provided that certain *constraints* are respected. An *assignee* can be specified for the action to be executed by.

The example of ODRL expression in Figure 3, serialized as RDF describes a language resource as being reproducible (downloaded, copied) but not derivable nor commercializable[20]. The absence of assignee is understood as 'applicable to anybody'.

---

[17]The prefix *dcat* stands for Data Catalog Vocabulary. DCAT is a W3C Recommendation `http://www.w3.org/TR/vocab-dcat/`

[18]`https://www.w3.org/community/odrl/`

[19]`http://www.w3.org/community/odrl/model/2/`

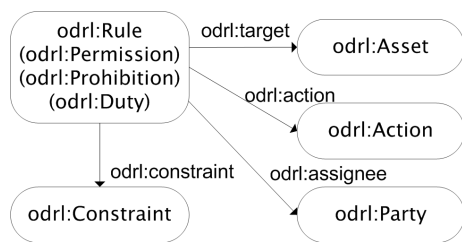[20]The prefix *odrl* points to `http://www.w3.org/ns/odrl/2/`

Figure 2: Main entities in the ODRL model

```
:example0
 a odrl:Set;
 odrl:permission [
   odrl:target :langResource ;
   odrl:action odrl:reproduce
 ] ;
 odrl:prohibition [
   odrl:target :langResource ;
   odrl:action odrl:derive,
     odrl:commercialize
 ] .
```

Figure 3: Simple example of ODRL policy

3.2 Rights expression within the META-SHARE structure

The ODRL model satisfies most of the concepts that are required for the licensing of LRs. Some adjustments have been required mainly to separate general conditions from the specifics that can instantiate them: for instance, payment is a general term of use but the exact amount to be paid for each LR may differ and vary depending on a number of other parameters (e.g. no fee for non-commercial use, X euros for commercial use, X euros but with a discount for a specific group of users etc.); by keeping the payment as a general condition in the RDF representation of the license and putting the amount to be paid on the LR, we can have the same standard license used for a large number of LRs. Consequently, the semantics of the ODRL model have been slightly altered for the Rights Information for Language Resources: *missing attributes in the policy can be found as attributes of the licensed asset.* Besides the vocabulary additions over the ODRL Common Vocabulary, which are foreseen by the specification, this is the only divergence that was made from the ODRL language.

The primary META-SHARE metadata schema presents conditions and rights in a flat structure. While this information is expressed in ODRL within the rules, having it directly accessible improves readability by simple processors. Hence,

```
:langResource a ms:languageResource .
:langResource ms:distribution :distrib1 .
:distrib1 dct:license :lic1 .
:lic1 ms:conditionsOfUse ms:noRedistri-
  bution, ms:nonCommercialUse .
```

Figure 5: Example equivalent to 3 using the MS structure

as a second design decision, *rights and conditions can be redundantly given as attributes of the policy or within the rule structure.*

The licensing information of a language resource can be entirely described with the MS/OWL ontology. In Figure 4, key classes are represented with orange ovals and minor classes with gray ovals. Class individuals are rectangles next to a class they are instances of. Properties are represented with arrows. For our regards, the four key elements in the META-SHARE structure are: a 'language resource' is published as a 'distribution', which may have attached a 'license'. 'Licenses' can have 'conditions of use'. The language resource can have different levels of availability (restricted, unrestricted, upon negotiation etc.). The distribution has a specific access medium and it can be granted to users of different nature (academic users, commercial users, etc. or combinations thereof). Licensors and distribution rights holders can also be expressed at the distribution level.

The License can belong to a License Category (ACA, RES, PUB) and it may contain different conditions of use –the fine grain but flat description of the license.

4 Examples of license

In the most simple setting, the metadata records describing a language resource may point to an RDF document with the license description. The RDF License dataset (Rodriguez-Doncel et al., 2014) contains a set of well-known licenses and licenses recommended by META-SHARE[21] which have been already written using the elements of the ontology.

To facilitate end users, we identified commonly used licenses in the LR domain from the values used for LRs distributed through META-SHARE. For our conversation we can identify the following categories that impose different treatment as

---
[21]The list of RDF licenses can be checked at http://rdflicense.appspot.com/
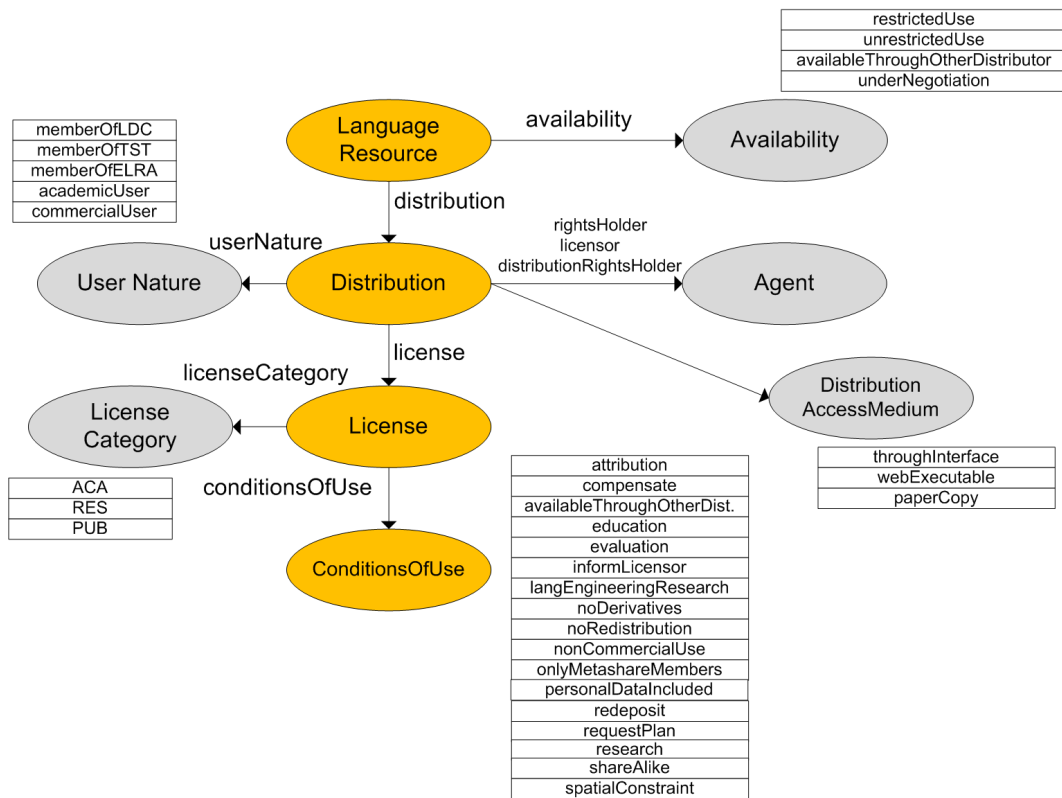
Figure 4: Rights Information for Language Resources

regards their RDF representation:

- Licenses, such as CC and FOSS that do not require any signatures; they are legal documents with a general text specifying the offering terms for end-users, they apply to all and do not ask for formal commitments from them. The text is published on a web site and can be accessed by anyone. They can have a direct representation in RDF.

- Standard licenses that include instantiation elements (e.g. ELRA, META-SHARE): legal documents that need to be signed by both contracting parties; they consist of a general text but include also specific terms that must be instantiated for each LR: the LR identification data as well as those of the signatories, but also specific fields such as the amount to be paid, the place where the LR will be used etc.; the licenses are available over the internet and can be accessed by anyone. In this case, the general text can be represented in RDF but we separated what is particular to the resource (e.g. the amount of money) and what is general and can be included in the RDF

of the license (e.g. the obligation to pay). For example, in order to declare that a resource is distributed under a META-SHARE Commercial-NoRedistribution-ForAFee license, the RDF fragment in Figure 6 can be used in its metadata record. The first line declares that *:resource* is a *dcat:Distribution*. The Dublin Core *license*[22] property links the resource with the license, and the price -whose precise number is not specific in the generic license online- is given. The price is specified using the GoodRelations[23] vocabulary.

- License templates with potential extra terms (e.g. CLARIN [24]): legal documents that include a general text and extra potential terms (e.g. attribution, request for a research plan, usage of the resource only at a specific location etc.); i.e. the use or not of specific terms leads to a new combination and the creation of a new license. The texts are also avail-

---

[22]dct is the prefix of http://purl.org/dc/terms/
[23]www.heppnetz.de/projects/goodrelations/
[24]http://clarin.eu/content/licenses-agreements-legal-terms

```
:resource a dcat:Distribution ;
 dct:license <http://purl.org/NET/
 rdflicense/ms-c-nored-ff> ;
 gr:hasPriceSpecification [
  gr:hasCurrencyValue "400"^^xsd:float;
  gr:hasCurrency "USD"^^xsd:string
 ].
] .
```

Figure 6: Example showing the use of a license template

able over the internet, but the combinations of the terms are free. The basic text itself can be represented in RDF, and so can the terms but the full RDF representation of all combinations must be dynamically constructed, with a combination of the RDF representation of the general text and the RDF representations of each additional term, once this is selected.[25]

- Non-standard licenses, such as proprietary ones, legal notices, terms of use etc.: there's a large variety of them, not all of the texts are available over the internet. There cannot be a ready-made RDF representation available for all of them. In this case, the conditionsOfUse element can help the end users get a quick grasp of what they are allowed to do with the LR.

The next example, in Figure 8, shows unabridged the "META-SHARE Commercial No Redistribution" license. The main resource in the license is an *odrl:Policy* (line 02) which has attributed some metadata elements: version (03), label (04), alternative name (05) or location of the legal code[26] (10). The policy additionally has information regarding the language and a flat list with the conditions (*ms:NoRedistribution*, *cc:Attribution*, etc. in lines 07-09).

The main permission (lines 12-25), which explicitly authorizes for making derivative works, making commercial use has the duty of attribution (15-17) and the constraints of being used only for language engineering purposes (lines 18-21) and on the users' site (lines 21-24). Distribution is forbidden in lines 26-28.

---

[25]see, for instance, https://www.clarin.eu/content/clarin-license-category-calculator with possible combination of license categories and terms of use.

[26]cc is prefix of http://creativecommons.org/ns#

5 Conclusions and future work

This paper has presented the Rights Information for Language Resources Ontology, the outcome of a cooperation between the META-SHARE project and the LIDER project, in the framework of the W3C Linked Data for Language Technology Group, which is expected to enhance the accessibility of language resources, following the Linked Data model, and facilitate their automatic processing by web services.

In the future, we expect to improve on the model, especially as regards the user modelling, as well as implement a mechanism for the dynamic generation of RDF representations of non-standard licences. Finally, the use of SPARQL queries to fill in predefined data structures will be investigated, so that the original ODRL structure is preserved while keeping the concept of license template.

Acknowledgments

References

Daan Broeder, Menzo Windhouwer, Dieter Van Uytvanck, Twan Goosen, and Thorsten Trippel. 2012. CMDI: a component metadata infrastructure. In *Describing LRs with metadata: towards flexibility and interoperability in the documentation of LR workshop programme*, page 1.

Nicoletta Calzolari, Riccardo Del Gratta, Gil Francopoulo, Joseph Mariani, Francesco Rubino, Irene Russo, and Claudia Soria. 2012. The LRE Map. Harmonising community descriptions of resources. In *Proceedings of the Eighth Conference on International Language Resources and Evaluation*, pages 1084–1089.

Maria Gavrilidou, Penny Labropoulou, Elina Desipri, Stelios Piperidis, Haris Papageorgiou, Mon-
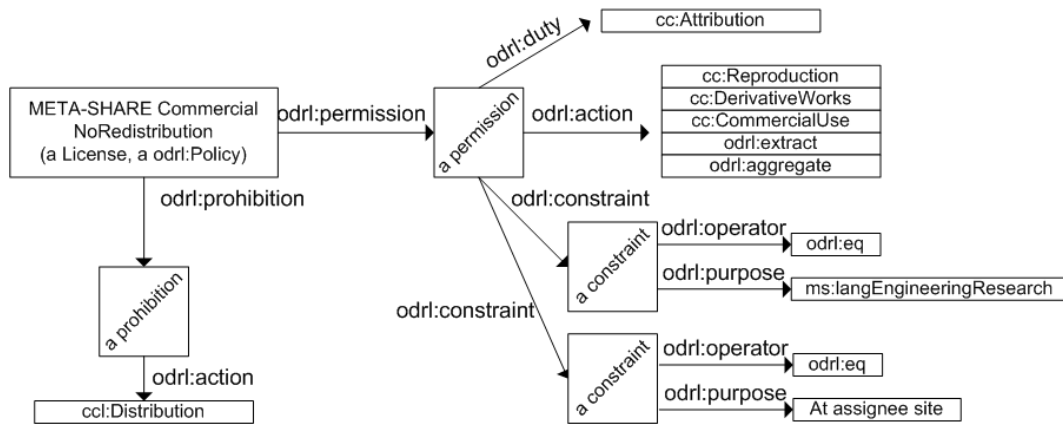
Figure 7: Graphical representation of the license in Figure 8

```
01<http://purl.org/NET/rdflicense/ms-c-nored>
02  a odrl:Policy ;
03  dct:hasVersion "1.0" ;
04  rdfs:label "META-SHARE Commercial NoRedistribution" ;
05  dct:alternative "MS C-NoReD" ;
06  dct:language <http://www.lexvo.org/page/iso639-3/eng> ;
07  ms:conditionsOfUse ms:noRedistribution, cc:Attribution,
08    cc:CommercialUse, ms:conditionsOfUse,
09    ms:languageEngineeringResearch ;
10  cc:legalcode <http://www.meta-net.eu/meta-s[...etc...].pdf> .
11  ms:licenseCategory ms:PUB ;
12  odrl:permission [
13     odrl:action cc:Reproduction, cc:DerivativeWorks , odrl:extract,
14        odrl:aggregate, cc:CommercialUse ;
15     odrl:duty [
16  odrl:action cc:Attribution ;
17     ] ;
18     odrl:constraint [
19       odrl:operator odrl:eq ;
20       odrl:purpose ms:languageEngineeringResearch
21     ] , [
22       odrl:operator odrl:eq ;
23       odrl:spatial "only at assignee's site"
24     ]
25 ];
26  odrl:prohibition [
27     odrl:action cc:Distribution ;
28  ] .
```

Figure 8: The META-SHARE Commercial No-redistribution license

ica Monachini, Francesca Frontini, Thierry Declerck, Gil Francopoulo, Victoria Arranz, and Valrie Mapelli. 2012. The META-SHARE metadata schema for the description of language resources. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, pages 1090–1097.

John McCrae, Penny Labropoulou, Jorge Gracia, Marta Villegas, Victor Rodriguez-Doncel, and Philipp Cimiano. 2015. One ontology to bind them all: The META-SHARE OWL ontology for the interoperability of linguistic datasets on the Web. In *Proceedings of the 4th Workshop on the Multilingual Semantic Web (to appear)*.

Mo McRoberts and Victor Rodriguez-Doncel (eds.). 2015. ODRL Version 2.1 Ontology. Final specifiction, ODRL W3C Community Group, March. http://www.w3.org/ns/odrl/2/.

Ville Oksanen and Crister Lindn. 2011. Open content licenses - how to choose the right one. In *NEALT Proceedings Series Vol. 13*.

Stelios Piperidis. 2012. The META-SHARE language resources sharing infrastructure: Principles, challenges, solutions. In *Proceedings of the Eighth Conference on International Language Resources and Evaluation*, pages 36–42.

Victor Rodriguez-Doncel, Serena Villata, and Asuncion Gomez-Perez. 2014. A dataset of RDF licenses. In Rinke Hoekstra, editor, *Proceedings of the 27th International Conference on Legal Knowledge and Information System*, pages 187–189.

Gary Simons and Steven Bird. 2003. Building an open language archives community on the oai foundation. *Library Hi Tech*, 21(2):210–218.

Claudia Soria, Nicoletta Calzolari, Monica Monachini, Valeria Quochi, Nria Bel, Khalid Choukri, Joseph Mariani, Jan Odijk, and Stelios Piperidis. 2014. The language resource strategic agenda: the flarenet synthesis of community recommendations. *Language Resources and Evaluation*, 48(4):753–775.

Marta Villegas, Maite Melero, and Nuria Bel. 2014. Metadata as Linked Open Data: mapping disparate XML metadata registries into one RDF/OWL registry. In *Proceedings of LREC 2014*.