Towards high quality, industry-ready Linguistic Linked Licensed Data

Víctor Rodríguez-Doncel\textsuperscript{1}, Daniel Vila-Suero\textsuperscript{1}, Asunción Gómez-Pérez\textsuperscript{1}, Philipp Cimiano\textsuperscript{2}, John P. McCrae\textsuperscript{2}, and Guadalupe Aguado-de-Cea\textsuperscript{1}

\textsuperscript{1}Ontology Engineering Group, Facultad de Informática, UPM. Madrid, Spain
\{dvila, vrodriguez, asun, lupe\}@fi.upm.es
\textsuperscript{2}Forschungsbau Intelligente Systeme (FBIIS). Universität Bielefeld. Bielefeld, Germany
\{cimiano, jmccrae\}@cit-ec.uni-bielefeld.de

FEISGILTT 2014, Localization World Conference, Dublin 04 June 2014
Towards high quality, **industry-ready** Linguistic Linked Licensed Data

**Community building networking**
- LD4LT
- BP-MLOD W3C-CG
- OntoLex W3C-CG

**Technological activities**
1. Roadmap on 3LD for Content Analytics
2. Guidelines for 3LD
3. 3LD Reference Architecture

**Surveys Requirements**

**Industry use cases**

**Technical activities**

[http://www.w3.org/community/Ld4Lt/](http://www.w3.org/community/Ld4Lt/)
[http://www.w3.org/community/bpmlod/](http://www.w3.org/community/bpmlod/)
Linguistic Linked Licensed Data

3LD

Linguistic Linked Licensed Data

Language resources such as:
- Lexica
- Corpora
- Dictionaries
- Grammars ..

Using RDF and standard data models (vocabularies):
- Lexica
- Corpora
- ......

Published along with a machine-readable license.

ODRL
Open Digital Rights Language

NIF
NLP Interchange Format
RDF(S) models

Unique identifiers: URI
identify or name a resource

Equivalence links to other datasets
Owl:same As

http://iflastandards.info/ns/fr/frbr/frbrer/C1005
Is creator of
Work
Is a
Cervantes

http://datos.bne.es/resource/XX1718747
same As
Cervantes

http://datos.bne.es/resource/XX3383563
Is creator of
El Quijote
Is a

http://dbpedia.org/resource/Miguel_de_Cervantes
same As
Cervantes

http://viaf.org/viaf/17220427
Linguistic **Linked** Licensed Data: Ontology and Data

- **Ontology**
  - Work
  - Person
  - Library
  - Language

- **Data**
  - Don Quijote de la Mancha
  - Vida de Miguel de Cervantes Saavedra
  - BNE
  - Cervantes Saavedra, Miguel de

- **URLs**
  - http://datos.bne.es/resource/XX1924295
  - http://datos.bne.es/resource/XX1718747
  - http://datos.bne.es/resource/XX1924295
  - http://datos.bne.es/resource/XX3383563
  - http://datos.bne.es/resource/bimo0002045496
  - http://datos.bne.es/resource/XX1718747
  - http://datos.bne.es/#

- **Dates**
  - http://.../date
  - http://.../publicationDate
  - http://.../translation
  - http://.../hasTopic
  - http://.../isAuthor
  - http://.../locatedAt
Language resources as Linked Data:

- Lexica
- Language descriptions
- Corpora

Linguistic Linked Open Data (LLOD) cloud

Linguistic LOD (LLOD) cloud

1 "Open Data and Linguistics" working group, Open Knowledge Foundation, see more http://linguistics.okfn.org/
Linguistic **Linked** Licensed Data: accessed by machines
Linguistic Linked Licensed Data

http://creativecommons...

dc:license

Ontologies

Datasets

http://creativecommons...
Linguistic Linked Licensed Data: current practice

Resource

Contains a link to a linguistic resource (a RDF dump file)
Contains a link to a license
Linguistic Linked Licensed Data: Guidelines

1. Add "rights" metadata in the dataset description (e.g., VoID, DCAT)

2. Use standard predicates to declare "rights" statements (e.g., Dublin Core terms: dc:rights, dct:license)

3a. Use URI of standard license e.g., CC0

3b. Use rights declaration language, e.g., ODRL

Standard license available?

Yes

No
META-SHARE NonCommercial NoRedistribution NoDerivatives For-a-Fee Licence

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix odrl: <http://www.w3.org/ns/odrl/2/> .

<http://example.com/nc-nored-nd-ff> a odrl:Policy ;
  rdfs:label "NC-NoReD-ND-FF" ;
  rdfs:comment "MetaShare NonCommercial, No Redistribution, No Derivatives, for a fee. Perpetual, worldwide, allowing no redistribution of the original." @en ;
  rdfs:seeAlso <http://www.meta-net.eu/meta-share....pdf> ;
  odrl:permission [ a odrl:Permission ;
    odrl:action odrl:reproduce; 
    odrl:duty [ a odrl:Duty ;
      odrl:action odrl:pay ;
      odrl:target "XXX EUR"
    ]
  ] ;
  odrl:prohibition [ a odrl:Prohibition ;
    odrl:action odrl:commercialize, odrl:distribute, odrl:derive
  ] .

See demo at http://conditional.linkeddata.es
Linguistic Linked Licensed Data
Demo: Conditional access to Linked Data

- **Prototype** developed at OEG-UPM
- A licenses-aware Linked Data server and a data policies and licenses manager
- Using Web standards (DCAT descriptions, SPARQL constructs, ODRL RDF policies, etc.)

Excerpt for a Demo taken from Princeton Wordnet³.⁰ in RDF
**Browse data: the synset “life”**

---

**Conditional access to Linked Data**

**Resource info**

---

### What is this?
- Introduction
- Technical details

### User's view (demo)
- Browse dataset
- Offers
- SPARQL
- Account

### Admin's view (demo)
- Control panel
- Accountability

---

**life (Synset)**

**Open triples**

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>sample</td>
<td>the oceans are teeming with life</td>
</tr>
<tr>
<td>part_of_speech</td>
<td>noun</td>
</tr>
<tr>
<td>translation</td>
<td>'יְדֹ   (life)</td>
</tr>
<tr>
<td>translation</td>
<td>vie</td>
</tr>
<tr>
<td>lexical_domain</td>
<td>noun tops</td>
</tr>
<tr>
<td>translation</td>
<td>takdir</td>
</tr>
</tbody>
</table>

**Limited access triples**

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>sameAs</td>
<td>Limited access (Offer --- Per RDF Statement)</td>
</tr>
</tbody>
</table>
Set some policies (admin)

Conditional access to Linked Data
Control panel

Assignment of policies to resources
Choose licenses/policies for each pre-defined dataset

Datasets

Dataset management

Mappings

Description: Mappings to terms in other resources
Number of triples: 519
Dataset license: [License1Cent]

Add  Remove  View
Conditional access to Linked Data
GeoLinkedData dataset

**Dataset RDF dump offers**

Different subsets of data can be offered under different conditions. These subsets can be portions of the dataset, mappings, specific resources or even individual statements.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>RDF dump</th>
<th>Licenses</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mappings</td>
<td></td>
<td>1- License1Cent</td>
<td>Mappings to terms in other resources</td>
</tr>
<tr>
<td>Thing</td>
<td></td>
<td>1- License15euros</td>
<td>Synset for Thing</td>
</tr>
<tr>
<td>General</td>
<td>Download</td>
<td>1- Creative Commons CC-BY</td>
<td>General dataset</td>
</tr>
</tbody>
</table>
Gain access to restricted data (user)

License: License1Cent

Rule (Permission)

Rights
reproduce,

Constraints
Pay ( Pay 0.01 EUR) - per unit,
Target

To be acted on http://localhost:99/ldr/resource/Provincia/Barcelona - http://www.w3.org/2004/02/skos/core#exactMatch

View license in RDF

<http://localhost:99/ldr/policy/ee32f675-ccae-4ca9-a544-3c07abf0b16e>  
a  <http://www.w3.org/ns/odrl/2/Policy> , <http://www.w3.org/ns/odrl/2/Set> ;  
<http://www.w3.org/2000/01/rdf-schema#comment>

"Individual triples are available upon payment of 1 euro cent" ;  
<http://www.w3.org/ns/odrl/2/permission> ....
Conditional access to Linked Data Account

Logged in as vroddon@gmail.com
User: Victor Rodriguez Doncel logout

You have purchased the following permissions:

<table>
<thead>
<tr>
<th>Price</th>
<th>Asset</th>
<th>License</th>
</tr>
</thead>
<tbody>
<tr>
<td>15,00 EUR</td>
<td><a href="http://salonica.dia.fi.upm.es/ldr/data">http://salonica.dia.fi.upm.es/ldr/data</a></td>
<td>License15euros (view)</td>
</tr>
</tbody>
</table>

Account
Conditional access to Linked Data

Introduction

**SPARQL endpoint**

This SPARQL endpoint provides access to the limited subset of RDF triples you are allowed to see.

Output:

- [ ] Text

Superuser: [ ]

Get Results
THANK YOU FOR YOUR ATTENTION

QUESTIONS?

Demo available at:

http://salonica.dia.fi.upm.es/ldr/en/
http://conditional.linkeddata.es
• META-SHARE recommends a set of 21 licenses, classified in:
  – META-SHARE Non Redistribution
  – META-SHARE Commons («distribution towards META-SHARE members»)
  – Creative Commons

• ALL of them can be represented as RDF using extendedly used vocabularies such as ODRL (Open Digital Rights Licenses)

• Advantages of expressing licenses as RDF:
  – Unambiguous identification of well known licenses by their URIs
  – Enables conditional access to resources
  – Automatic license compatibility analysis when integrating resources
We offer a machine-readable RDF file with information about languages and other entities provided on Lexvo.org. This dump file can be downloaded to your local computer and distributed in accordance with the Creative Commons Attribution-ShareAlike (CC-BY-SA) 3.0 Unported license:

- Lexvo dump (2013-02-09)
- Lexvo dump (direct link to latest version 2013-02-09)

We are grateful to a number of people for providing the original data sources that were used to create Lexvo.org.

<!--
This data file is a part of
Lexvo
http://www.lexvo.org/
Gerard de Melo, 2008-2013

License: Creative Commons Attribution-Share Alike 3.0 Unported
http://creativecommons.org/licenses/by-sa/3.0/

Portions copyright various contributors (e.g. article authors on Wikipedia). For information about the data sources and the respective copyrights, please see:
http://www.lexvo.org/linkeddata/sources.html
-->

<!-Description rdf:about="http://id.loc/vocabulary/iso639-2/abk">
</Description>

<!-Description rdf:about="http://id.loc/vocabulary/iso639-2/ace">
</Description>