Linked Open Vocabularies (LOV)

Pierre-Yves Vandenbussche, Lead Researcher, Fujitsu Ireland
Ghislain Atemezing, Mondeca, France
Maria Poveda, Universidad Politécnica de Madrid, Spain
Bernard Vatant, Mondeca, France
What do we mean by vocabulary

- RDF vocabulary defined as a schema (T-box) for the description of Linked Data
  - Limited to Metadata Element Sets for description of Linked Data
  - Does not contain Value Sets such as expressed in SKOS
What do we mean by vocabulary

- **RDF vocabulary defined as a schema (T-box) for the description of Linked Data**
  - Limited to Metadata Element Sets for description of Linked Data
  - Does not contain Value Sets such as expressed in SKOS

- Designed with **reuse** in mind
  - Relaxation of formal constraints (rarely using OWL constructs)
What do we mean by vocabulary

- **RDF vocabulary defined as a schema (T-box) for the description of Linked Data**
  - Limited to Metadata Element Sets for description of Linked Data
  - Does not contain Value Sets such as expressed in SKOS

- **Designed with **reuse** in mind**
  - Relaxation of formal constraints (rarely using OWL constructs)

- **Vocabulary should refer to and **reuse** relevant existing ones**
  - Not reinventing the wheel and maximising the reusability
## Linked Open Vocabularies

Promote and facilitate the reuse of well documented vocabularies in the Linked Data ecosystem

<table>
<thead>
<tr>
<th>In Figures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Since</td>
</tr>
<tr>
<td># vocabs</td>
</tr>
<tr>
<td># vocab terms</td>
</tr>
<tr>
<td>- Median # classes</td>
</tr>
<tr>
<td>- Median # prop</td>
</tr>
<tr>
<td># agents (creators, etc.)</td>
</tr>
</tbody>
</table>

http://datalift.org/

https://lov.okfn.org/
Linked Open Vocabularies

A living ecosystem

Creation date

Last Modification date
Linked Open Vocabularies – DCAT Example

**Metadata**

- **URI**: http://www.w3.org/ns/dcat
- **Namespace**: http://www.w3.org/ns/dcat#
- **Homepage**: http://www.w3.org/TR/vocab-dcat/
- **Description**: DCAT is an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web.

**Language**

- Arabic
- Greek
e1
- English
en
- Spanish
es
- French
fr
- Japanese
ja

**Contributors**

- Richard Cyganiak
  http://google.com/+RichardCyganiak
- Phil Archer
  https://plus.google.com/103670676337547906055
- Fadi Meali

**Relationships**

**Versions**

- W3C Recommendation
  v2014-05-31
- W3C Candidate Recommendation
  v2013-11-28
- W3C Working Draft 30 July 2013
- W3C Working Draft 05 April 2012

© FUJITSU LIMITED 2016
See also:
Schaible, Johann, et al. "TermPicker: Enabling the Reuse of Vocabulary Terms by Exploiting Data from the LOD Cloud" (ISWC 2016)
Act Locally
#1 Provide Basic Metadata (Who, What, Why, When)

- Only 55% of the vocabularies specify at least one creator, contributor or publisher

Agent (creator, contributor, publisher)

- With 55%
- Without 45%

LOV Recommendation: [https://lov.okfn.org/Recommendations_Vocabulary_Design.pdf](https://lov.okfn.org/Recommendations_Vocabulary_Design.pdf)
Best Practices for Publishing Linked Data - W3C: [https://www.w3.org/TR/ld-bp/](https://www.w3.org/TR/ld-bp/)
#1 Provide Basic Metadata (Who, What, Why, When)

- Only 55% of the vocabularies specify at least one creator, contributor or publisher.
- 65% of the vocabularies specify a publication date.

Agent (creator, contributor, publisher)

- With 45%
- Without 55%

Best Practices for Publishing Linked Data - W3C [https://www.w3.org/TR/ld-bp/](https://www.w3.org/TR/ld-bp/)
#1 Provide Basic Metadata (Who, What, Why, When)

- Only 55% of the vocabularies specify at least one creator, contributor or publisher
- 65% of the vocabularies specify a publication date
- Early vocabularies (Usually before 2007) have no metadata at all, e.g. Contact vocabulary
  - Several vocabularies can be used: Dublin Core, ADMS, VANN, and VOAF

Agent (creator, contributor, publisher)

LOV Recommendation  

Best Practices for Publishing Linked Data - W3C  
https://www.w3.org/TR/ld-bp/

ISA (European Comission) cookbook  
#2 Provide Labels with language information

- Labels are the only way for a human to search, understand a term. Must be a strict requirement.
#2 Provide Labels with language information

- Labels are the only way for a human to search, understand a term. Must be a strict requirement

- 27.98% of the vocabularies with no explicit language

- Only 14.68% of the vocabularies are multilingual
  - DCAT is a good example with 6 languages
#2 Provide Labels with language information

- Labels are the only way for a human to search, understand a term. Must be a strict requirement.
- 27.98% of the vocabularies with no explicit language.
- Only 14.68% of the vocabularies are multilingual.
  - DCAT is a good example with 6 languages.
- 45 languages are used by vocabularies in LOV.
#3 License

- 35% of vocabularies explicitly specify a license

<table>
<thead>
<tr>
<th>Kind of License</th>
<th>#licenses (feb2015)</th>
<th>#licenses (sep2013)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not specified</td>
<td>306</td>
<td>313</td>
</tr>
<tr>
<td>Public domain</td>
<td>16</td>
<td>5</td>
</tr>
<tr>
<td>Attribution</td>
<td>97</td>
<td>18</td>
</tr>
<tr>
<td>Share alike</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>With restrictions</td>
<td>6</td>
<td>37</td>
</tr>
<tr>
<td>Closed</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Other</td>
<td>28</td>
<td>4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>469</strong></td>
<td><strong>386</strong></td>
</tr>
</tbody>
</table>

[http://www.cosasbuenas.es/blog/LOVlicenses](http://www.cosasbuenas.es/blog/LOVlicenses)

#3 License

- 35% of vocabularies explicitly specify a license
- How do namespace and License work together?
- Is an extension within an external namespace allowed?

<table>
<thead>
<tr>
<th>Kind of License</th>
<th>#licenses (feb2015)</th>
<th>#licenses (sep2013)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not specified</td>
<td>306</td>
<td>313</td>
</tr>
<tr>
<td>Public domain</td>
<td>16</td>
<td>5</td>
</tr>
<tr>
<td>Attribution</td>
<td>97</td>
<td>18</td>
</tr>
<tr>
<td>Share alike</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>With restrictions</td>
<td>6</td>
<td>37</td>
</tr>
<tr>
<td>Closed</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Other</td>
<td>28</td>
<td>4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>469</strong></td>
<td><strong>386</strong></td>
</tr>
</tbody>
</table>

http://www.cosasbuenas.es/blog/LOVlicenses

#4 Versioning

- A vocabulary is **meant to change**
- Need to keep the history
- Preserve the semantic: URI stability
- Currently no recommendation
  - Proof of concept with VOAF vocabulary
  - Example of schema.org’s version control

#5 Reuse and Link

“If suitable terms can be found in existing vocabularies, these should be reused to describe data wherever possible, rather than reinvented. Reuse of existing terms is highly desirable as it maximises the probability that data can be consumed by applications.” (Heath and Bizer, 2011)

<table>
<thead>
<tr>
<th>Inter-vocabulary relationship</th>
<th># relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>voaf:metadataVoc</td>
<td>2,637</td>
</tr>
<tr>
<td>voaf:specializes</td>
<td>1,269</td>
</tr>
<tr>
<td>voaf:extends</td>
<td>1,031</td>
</tr>
<tr>
<td>owl:imports</td>
<td>373</td>
</tr>
<tr>
<td>voaf:hasEquivalencesWith</td>
<td>201</td>
</tr>
<tr>
<td>voaf:generalizes</td>
<td>57</td>
</tr>
<tr>
<td>voaf:hasDisjunctionsWith</td>
<td>16</td>
</tr>
</tbody>
</table>

- Link to existing vocabularies (e.g. extend, state equivalency, etc.)
  - Well described in methodology and VOAF vocabulary
#6 Leverage the Linked Data and Web architecture

- Machine have their right too!
- Use of RDF and URI namespaces to describe and identify a vocabulary and its terms.
- Dereferencing:
  - Access of the resource (vocabulary or vocabulary term) using its URI
  - Access to documentation for Human and RDF version for machine
- Keep Human documentation and Machine document synchronised
#7 Sustainable governance for long-term preservation

- Preservation will not happen automatically. It has to be planned!
- “Persistent” URIs (e.g. PURL) Institutional Support
  - Publication under W3C namespace
  - Example of FOAF – DCMI agreement: “DCMI could assume maintenance responsibility for the vocabulary if the FOAF Project should cease its normal activity”
- Safety through redundancy (LOCKSS “Lots of copies keep stuff safe”)

Workshop Smart Descriptions & Smarter Vocabularies (SDSVoc)
30 November - 1 December, CWI, Amsterdam Science Park
https://www.w3.org/2016/11/sdsvoc/
#8 Cataloguing and Monitoring

- Provide visibility to vocabulary publisher
  - Empirical data about vocabulary usage
- Mean to search across all vocabularies
  - LOV, Vocab.cc, LodStats, OOSP, FalconS, SWSE, Watson, Swoogle
#8 Cataloguing and Monitoring

- Provide visibility to vocabulary publisher
  - Empirical data about vocabulary usage
- Mean to search across all vocabularies
- Provide a tool set for
  - Syntax and structure checking
  - Pitfalls detection
  - Translation support

| LOV, Vocab.cc, LodStats, OOSP, FalconS, SWSE, Watson, Swoogle |
| RDF triple Checker, OOPS, RDF Unit |
#8 Cataloguing and Monitoring

- Provide visibility to vocabulary publisher
  - Empirical data about vocabulary usage
- Mean to search across all vocabularies
- Provide a tool set for
  - Syntax and structure checking
  - Pitfalls detection
  - Translation support
- Hosting vocabulary

<table>
<thead>
<tr>
<th>LOV, Vocab.cc, LodStats, OOSP, FalconS, SWSE, Watson, Swoogle</th>
</tr>
</thead>
<tbody>
<tr>
<td>RDF triple Checker, OOPS, RDF Unit</td>
</tr>
</tbody>
</table>
#8 Cataloguing and Monitoring

- Provide visibility to vocabulary publisher
  - Empirical data about vocabulary usage
- Mean to search across all vocabularies
- Provide a tool set for
  - Syntax and structure checking
  - Pitfalls detection
  - Translation support
- Hosting vocabulary
- Vocabulary update notification

LOV, Vocab.cc, LodStats, OOSP, FalconS, SWSE, Watson, Swoogle

RDF triple Checker, OOPS, RDF Unit
Toolkit 1/2

Datalift organisation
http://datalift.org/

ISA Team at the University of Oxford’s e-Research Centre
https://github.com/ISA-tools/OntoMaton

Laurens Rietveld
http://yasgui.org/

Agile Knowledge Engineering and Semantic Web (AKSW), Univ. of Leipzig
http://ontowiki.net/

Agile Knowledge Engineering and Semantic Web (AKSW), Univ. of Leipzig
http://aksw.org/Projects/RDFUnit.html

Semantic Web and Ontological Engineering (SWOE) Prague University of Economics
http://owl.vse.cz:8080/OOSP/
Toolkit 2/2

Ontology Engineering Group, Universidad Politécnica de Madrid
http://ontoology.linkeddata.es/

VoCol
Fraunhofer IAIS
https://github.com/vocol/vocol

Oops!
Ontology Engineering Group (OEG), Universidad Politécnica de Madrid
http://oops.linkeddata.es/

LODstats
Agile Knowledge Engineering and Semantic Web (AKSW), Univ. of Leipzig
http://stats.lod2.eu/

RDF Triple Checker
Christopher Gutteridge, Southampton ECS Web Team
http://graphite.ecs.soton.ac.uk/checker/
## Conclusion

<table>
<thead>
<tr>
<th></th>
<th>Provide Basic Metadata (Who, What, Why, When)</th>
<th>Technical</th>
<th>Recommendation</th>
<th>Governance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>X</td>
<td>XXX</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Provide Labels with language information</td>
<td>X</td>
<td>XXX</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>License</td>
<td>X</td>
<td>XXX</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Versioning</td>
<td>XX</td>
<td>XX</td>
<td>X</td>
</tr>
<tr>
<td>5</td>
<td>Reuse and Link</td>
<td>XX</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Leverage the Linked Data and Web architecture</td>
<td>XXX</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Sustainable governance for long-term preservation</td>
<td>XXX</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Cataloguing and Monitoring</td>
<td>XX</td>
<td></td>
<td>XXX</td>
</tr>
</tbody>
</table>

Thank you for your attention

Pierre-Yves Vandenbussche