Summary

• What is a vocabulary anyway?
  • And what does sustainability mean?

• Main points from VocPres session
  • Who should care of what?

• Lessons learned in LOV project
  • Various examples of non sustainable practices

• Discussion
  • What can (should) we do as vocabulary managers?

bernard.vatant@mondeca.com
What is a vocabulary?

A vocabulary is a set of terms

- Put together (for some functional purpose)
- Published on the Web (by some agent)
- Documented (what, why, who)

Each term provides

- A human readable definition
- A machine readable description
- A stable identifier (URI or shared key)
**What is a vocabulary term?**

Each vocabulary term is an **interface**

**Human readable**
- label, description

**Machine actionable**
- identifier, description

*Event* - An arbitrary classification of a space/time region, by a cognitive agent. An event may have actively participating agents, passive factors, products, and a location in space/time.

http://purl.org/NET/c4dm/event.owl#Event
A vocabulary is a **changing thing** ...

Every bit and piece is likely to be modified: Terms, URI, namespace, publisher, creator ...

http://en.wikipedia.org/wiki/Theseus%27s.Ship_paradox
The change is under control

- Vocabulary history is available
- Term meaning is preserved as far as possible
- If the meaning changes, it’s documented

Users can cope with change

- Changes are not too frequent
- Users are made aware of change in due time
• Various technical approaches to sustainability
  • PURL, LOCKSS, LOV

• But not only a technical problem!
  • Needs governance with each stakeholder playing its role

• Libraries
  • Should (will) engage in long-term preservation
  • Creation of vocabularies collections

• The community (W3C, DCMI) will
  • Spread best practices
  • Help to publication of quality vocabularies
  • Help users to discover the quality vocabularies they need
History

- **Started 2011-Q1**
  - Initially a deliverable of Datalift project: [http://datalift.org](http://datalift.org)

- **2012**: Growing interest beyond the initial scope
  - Migration to OKFN since July 2012: [http://lov.okfn.org](http://lov.okfn.org)
  - Open code since 2013-Q1: [https://github.com/pyvandenbussche/](https://github.com/pyvandenbussche/)

Status as of September 2013

- 360+ vocabularies, and counting
- Continuous feedback from the community
- Small management team: 5 curators, 1 technical manager
LOV interfaces

- LOV Home: [http://lov.okfn.org](http://lov.okfn.org)
- LOV-Search: [http://lov.okfn.org/dataset/lov/search/](http://lov.okfn.org/dataset/lov/search/)
- LOV-Suggest: [http://lov.okfn.org/dataset/lov/suggest/](http://lov.okfn.org/dataset/lov/suggest/)
- LOV-Bot: Daily or on-demand scan of vocabularies
- LOV-Edit: Administration and curation back-office
- SPARQL endpoint(s)
  - Vocab metadata [http://lov.okfn.org/endpoint/lov](http://lov.okfn.org/endpoint/lov)
  - Vocab content [http://lov.okfn.org/endpoint/lov_aggregator](http://lov.okfn.org/endpoint/lov_aggregator)
A vocabulary in LOV

http://lov.okfn.org/dataset/lov/details/vocabulary_org.html

Rich metadata

Usage in linked data

NEW

Relations with other vocabularies

Versions and timeline
Vocabulary Sustainability

Lessons learned

• Vocabularies in LOV are less than 13 years old
  • Oldest were published in 2000, but ...

• Tracing history is a pain
  • Finding versions
  • Finding documentation
  • Finding responsible contacts

• Publication practices are a pain
  • Variety of RDF formats
  • Variety of URI, namespace and content negotiation policies
  • URIs change without notice ...
Absent or inconsistent metadata

- DC basics often missing in vocabularies
  - Title
  - Creation, publication, modification date
  - Creator, contributor, publisher

- Easy to add in owl:Ontology header!

- Metadata inconsistent across formats
  - Different in HTML and in RDF
URI and namespace

• The vocabulary URI is the one from which you can access the vocabulary.

• It should be the same as the vocabulary namespace, but they are often different.

• The URI where you find the vocabulary is often different from the one declared inside the vocabulary (if present)

• Namespace declared and effectively used can be different.

• Several vocabularies can use the same namespace

• Content negotiation can be broken

• Etc.
Changing URI and namespace

Referenced by some vocabularies using its former URI http://www.loa-cnr.it/ontologies/DUL.owl#. This URI is still online but stuck at version 3.22.
Versioning

- No versioning at all in many vocabularies
  - Exemple: LoC authorities

- Often non-standard versioning information

- Although OWL and DC provides for basic stuff
  - owl:versionInfo, owl:priorVersion, owl:incompatibleWith
  - dcterms:isReplacedBy
Conclusion

• **Do your homework!**
  • Quality of data depends on your vocabulary quality
  • Other vocabularies will rely on yours

• **A healthy ecosystem starts by healthy individuals**

• **Where do we start?**