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

MONDECA

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

MONDECA

Yes, but … what about time?

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

Sustainability means

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

VocPres report

- 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

Monitoring sustainability : LOV

- ## History
  - ### Started 2011-Q1
    - Initially a deliverable of Datalift project : http://datalift.org
    - First hosted at Mondeca Labs : http://labs.mondeca.com/
  - ### 2012 : Growing interest beyond the initial scope
    - Migration to OKFN since July 2012 : http://lov.okfn.org
    - Open code since 2013-Q1: 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
- LOV-Search : http://lov.okfn.org/dataset/lov/search/
- LOV-Suggest : http://lov.okfn.org/dataset/lov/suggest/
- LOV-Stats : http://lov.okfn.org/dataset/lov/stats/
- LOV-API : http://lov.okfn.org/dataset/lov/apidoc/
- 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
  - Vocab content http://lov.okfn.org/endpoint/lov_aggregator

MONDECA

A vocabulary in LOV

DC-2013 Linking to the Future Lisbon

VocDay Workshop

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

Metadata:

Property	Value
is part of vocabulary space	All > City > Society
Vocabulary URI	http://www.w3.org/ns/org#
Prefix	org
Namespace URI	http://www.w3.org/ns/org#
Description	Vocabulary for describing organizational structures, specializable to a broad variety of types of organization. @en
Date issued	2010-05-28
Last modified	2013-02-15
Language	English, French
Creator	Dave Reynolds, Dominique Guardiola
Publisher	W3C
Class number	9
Property number	35
Homepage	http://www.w3.org/TR/vocab-org/
See also	http://stats.lod2.eu/vocabularies/149
Represented by	format-org
Has review	(2012-10-02) Bernard Vatant: Currently a V 0.4 is a transcription of http://www.epimor editor's draft. Versions 0.1 to 0.3 retrieved

Rich metadata

Vocabulary used in **5** datasets (information coming from LOD Stats project)

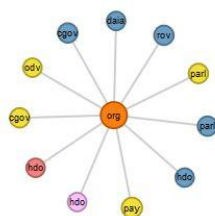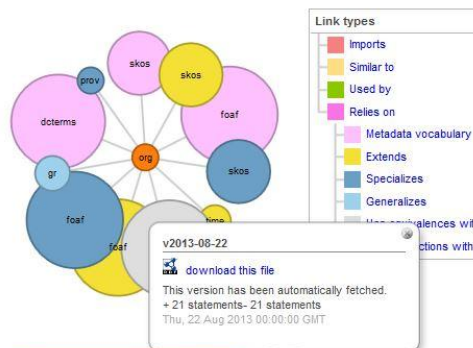| Occurrences	Dataset
177820	EU Who Is Who
2000	Organisation (University of Southampton)
1680	EU Parliament Media Library
24	Buildings and Places (University of Southampton)
2	University of Southampton Profile Document (University of Southampton)
See the full list...	

Usage in linked data

NEW

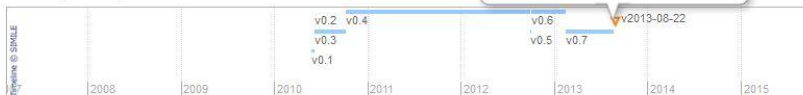Vocabulary links:

Vocabularies referencing "org" (11)

Vocabularies referenced by "org" (11)

Link types	
Imports	
Similar to	
Used by	
Relies on	
	Metadata vocabulary
	Extends
	Specializes
	Generalizes
	Has equivalences with
	...ctions with

v2013-08-22

download this file

This version has been automatically fetched.
+ 21 statements- 21 statements
Thu, 22. Aug 2013 00:00:00 GMT

Relations with other vocabularies

Vocabulary history:

| v0.2 v0.4 | v0.6 | v2013-08-22 |
| v0.3 | v0.5 v0.7 |
| v0.1 |

| 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |

Versions and timeline

MONDECA

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 …**

MONDECA

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 .

MONDECA

Changing URI and namespace

DUL - DOLCE+DnS Ultralite

Metadata:

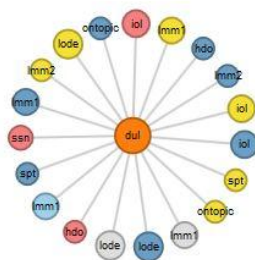| Property	Value
is part of vocabulary space	All > Upper & Meta > General & Upper
Vocab...	
Prefix	
Names	
Description	The DOLCE+DnS Ultralite ontology. It is a simplification of some parts of the DOLCE Lite-Plus library (cf. http://www.ontologydesignpatterns.org/ont/dul/DLP397.owl) @en
Last modified	2012-07-13
Language	English, Italian
Creator	Aldo Gangemi
Class number	75
Property number	109
Has review	(2013-08-05) Bernard Vatant: Refer... http://www.loa-cnr.it/ontologies/DUL...

http://www.ontologydesignpatterns.org/ont/dul/DUL.owl

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.

Vocabulary links:
Vocabularies referencing "dul" (19)   Vocabularies...

Link types
- Imports
- Similar to
- Used by
- Relies on
  - Metadata vocabulary
  - Extends
  - Specializes
  - Generalizes
  - Has equivalences with
  - Has disjunctions with

Vocabulary history:

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?