

Vocabulary Selection

Michael Hausenblas (DERI)

Ghislain Atemezing (EURECOM)

Rapaël Troncy (EURECOM)

Bernadette Hyland (3 Round Stones)

Boris Villazón-Terrazas (OEG-UPM)



ToC

- Introduction
- Modelling within GLD
- The Big picture
- Search for available vocabularies
- Summary of vocabularies
- Summary of search engines/ontology repositories
- Criteria for selecting a particular vocabulary/ontology
- LOV features
- Build the vocabulary by reusing the vocabularies we have found



Introduction

- From the GLD WG Charter
 - The group will provide advice on how governments should select RDF vocabulary terms (URIs), including advice as to when they should mint their own. This advice will take into account issues of stability, security, and long-term maintenance commitment, as well as other factors that may arise during the group's work.

Why?

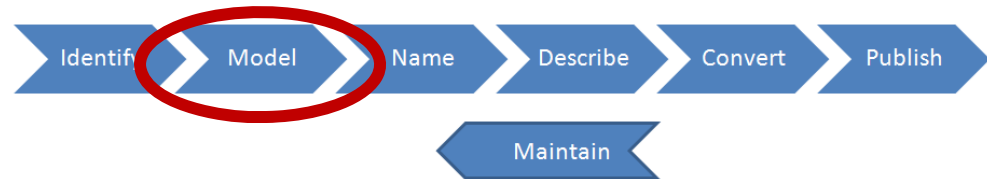
- RDF is a triple data model that aims to represent knowledge. However, it does not provide any domain-specific terms for describing classes of things in the world and how they relate to each other ¹.
- We need vocabularies/ontologies expressed in RDFS/OWL.



¹ Linked Data Book
<http://linkedatabook.com/editions/1.0/#htoc49>

Modelling is an important phase in any GLD Life cycle, for example

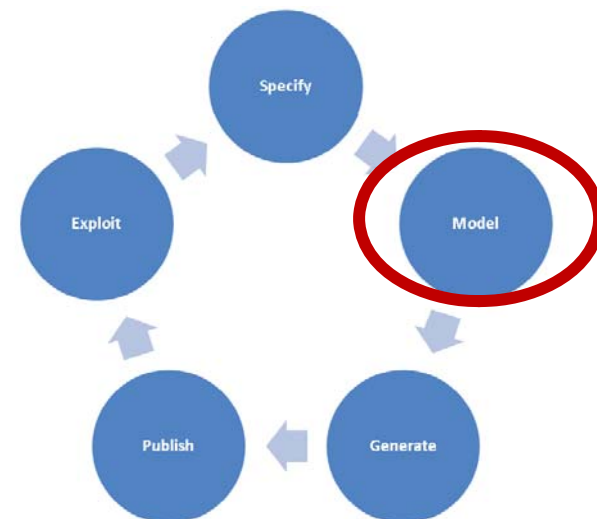
- Hyland et al.



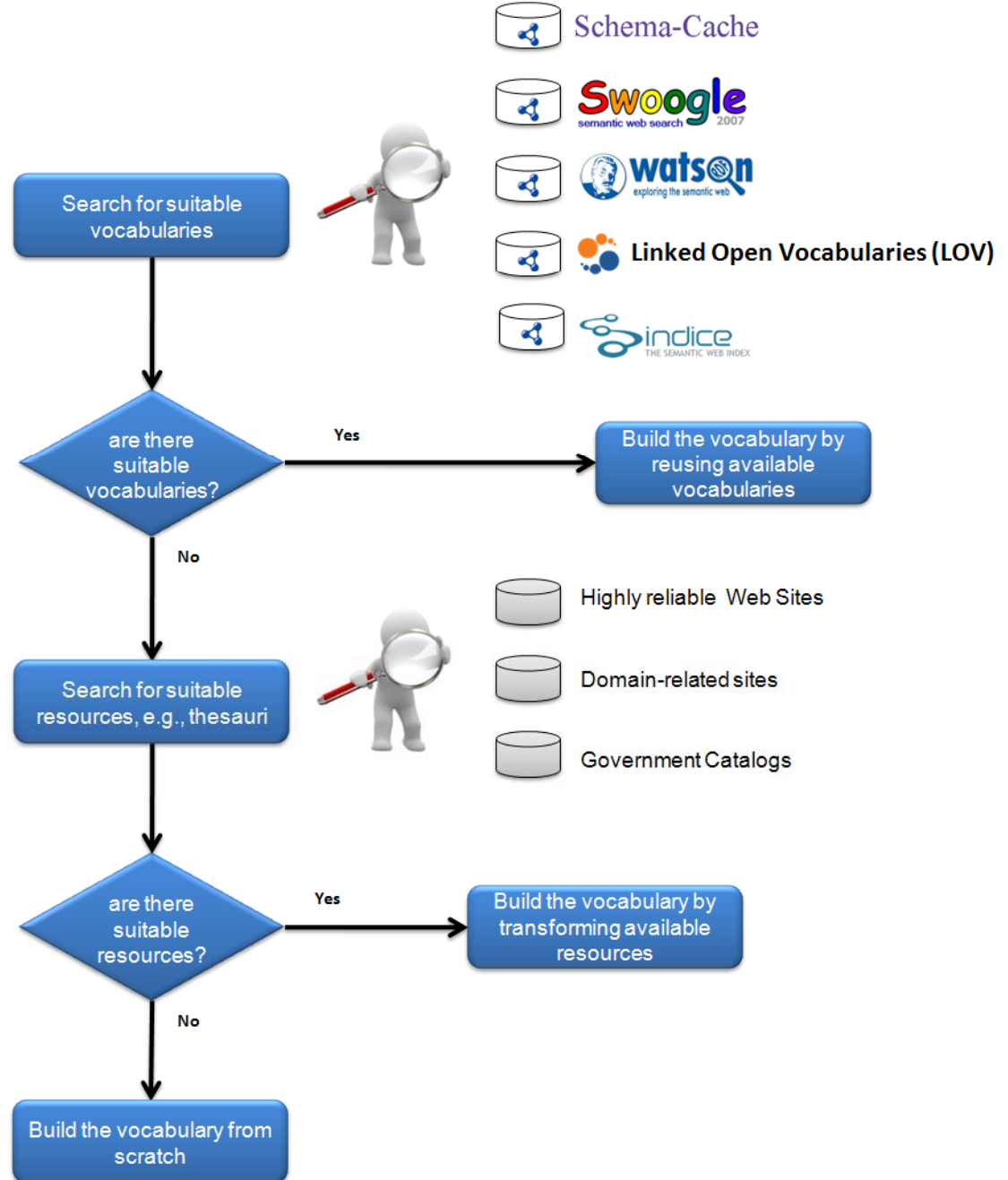
- Hausenblas et al.



- Villazon-Terrazas et al.



Big picture





Search for suitable vocabularies

- There is no definitive directory that can be consulted to find suitable vocabularies and ontologies.

Summary of available directories/search engines

	Ontology Directory/ Search engine/ Repository of Ontologies/ etc	Manually curated	Maintainer	Number of vocabs/ ontologies (approx)	Status	URL
vocab.org	ontology directory	No	Ian Davis	30	Up and running	http://vocab.org/
ontologi.es	ontology directory	No	Linked Railway Data Project	30	Up and running	http://ontologi.es
open vocab	community maintained vocabulary	No	Ian Davis	65 classes 188 properties	Up and running	http://open.vocab.org/
schemaweb	ontology directory	No			Not working	http://www.schemaweb.info/
schemapedia	community built compendium of RDF schemas	No	Ian Davis	283	Up and running	http://schemapedia.com/
vocab.cc	RDF vocabulary search and lookup	No	Andreas Hart (AIFB)	Pending	Up and running	http://www.vocab.cc/
swoogle	semantic web search	No	ebiquity group at UMBC	10000	Up and running	http://swoogle.umbc.edu/
watson	semantic web search	No	KMI at Open University	Pending	Up and running	http://kmi-web05.open.ac.uk/WatsonWUI/
Sindice	semantic web index	No	DERI	Pending	Up and running	http://www.sindice.com/
Falcons	semantic web search	No	Websoft Research Group, Nanjing University	Pending	Up and running	http://ws.nju.edu.cn/falcons/
schema-cache	vocabulary directory	No	Talis	Pending	Up and running	http://schemacache.com/
LOV	vocabulary directory	Yes	Mondeca Labs	185	Up and running	http://labs.mondeca.com/dataset/lov/

- There are no guidelines helping ontology developers to decide between one semantic web engine.

Summary of vocabularies relevant for government agencies

Name	Domain	URL
WGS84	for geo positioning defines terms for lat(itude), long(itude) and other information about spatially-located things	http://www.w3.org/2003/01/geo/wgs84_pos#
Bibliographic Ontology (BIBO)	describing citations and bibliographic references	http://bibliontology.com/specification
Creative Commons Rights Expression Language	for copyright licenses	http://creativecommons.org/ns#
Description of a Project	vocabulary to describe software projects, with particular emphasis on Open Source projects.	http://trac.usefulinc.com/doap
The Dublin Core Metadata Initiative (DCMI)	defines general metadata attributes for published works including title, creator, date, subject and publisher	http://dublincore.org/documents/dcmi-terms/
Friend-of-a-Friend (FOAF) vocabulary	for describing people. their activities (collaboration) and their relations to other people and objects.	http://xmlns.com/foaf/spec/
GeoNames Ontology	for describing geographical features	http://www.geonames.org/ontology/ontology_v2.2.1.rdf
Good Relations	for E-commerce that defines terms for describing products, price, and company data.	http://www.heppnetz.de/projects/goodrelations/
Object Reuse and Exchange	defines standards for the description and exchange of aggregations of Web resources.	http://www.openarchives.org/ore/
Semantically-Interlinked Online Communities (SIOC)	to describe information about an online community sites, such as users, posts and forums	http://rdfs.org/sioc/spec/
vCard	for address books	http://w3.org/TR/vcard-rdf/
vocabulary of Interlinked Datasets (VOID)	defines key metadata about RDF datasets	http://www.w3.org/2001/sw/Interest/void/
RDF Data Cube vocabulary	for statistical data	http://linked-statistics.org/datacube/
VOAF	for description of RDF(S)/OWL vocabularies	http://labs.mondeca.com/vocab/voaf#
dcat	for describing data catalogs	http://www.w3.org/ns/dcat#



Widely used Vocabulary prefixes

	Prefix	Vocabulary URI	Description
1	dc	http://purl.org/dc/terms/	Dublin Core Metadata Terms
2	og	http://ogp.me/ns#	Facebook's Open Graph protocol
3	cc	http://creativecommons.org/ns#	Creative Commons Rights Expression Language
4	foaf	http://xmlns.com/foaf/0.1/	Friend of Friend Vocabulary
5	v	http://rdf.data-vocabulary.org/#	Google Rich Snippets' Vocabularies
6	sioc	http://rdfs.org/sioc/ns#	Semantically-Interlinked Online Communities
7	vcard	http://www.w3.org/2006/vcard/ns#	vCard
8	gr	http://purl.org/goodrelations/v1#	GoodRelations Ontology
9	rev	http://purl.org/stuff/rev#	RDF Review Vocabulary
10	ctag	http://comontag.org/ns#	Common Tag Ontology
11	ical	http://www.w3.org/2002/12/cal/icaltzd#	iCalendar terms in RDF



LOV



- LOV suggest - <http://labs.mondeca.com/dataset/lov/suggest/>

New Vocabulary URI:

validate vocabulary



Vocabulary <http://www.w3.org/ns/r2rml>

Namespace URI	http://www.w3.org/ns/r2rml#
Namespace Prefix	r2rml
Classes nb	14
Properties nb	26
Language	English

cancel submit vocabulary



Please enter your email for further correspondence

email:

Send Cancel

LOV













- LOV aggregator - <http://labs.mondeca.com/dataset/lov/agg/>

LOV aggregator gathers all vocabularies in a single [endpoint](#).



Vocabulary aggregation (178/185):

Prefix	Title	Download Status
acco	Accomodation Ontology	 last date: 2011-12-18
acl	Basic Access Control ontology	 last date: 2011-12-18
ad	Address Schema	 last date: 2011-12-18
af	Audio Features Ontology	 last date: 2012-01-13
agls	AGLS Metadata Terms	 last date: 2011-12-18
aiiso	Academic Institution Internal Structure Ontology	 last date: 2011-12-18
akt	AKT Reference Ontology	 last date: 2011-12-18
akts	AKT Support Ontology	 last date: 2011-12-18
ao	Association Ontology	 last date: 2011-12-18
api	Linked Data API Vocabulary	 last date: 2011-12-18

LOV



- LOV search - <http://labs.mondeca.com/dataset/lov/search/>

latitude Search

22 results in 8 vocabularies

geo:lat (rdf:Property)	score:1.000	>>
rdfs:label latitude		
rdfs:comment The WGS84 latitude of a SpatialThing (...)		
ogp:latitude (rdf:Property)	score:0.882	>>
rdfs:label latitude @en-us		
rdfs:comment [DEPRECATED] The latitude of the resource e.g.,..., the latitude of a company. @en-us		
schema:latitude (rdf:Property)	score:0.882	>>
rdfs:label Latitude @en		
rdfs:comment The latitude of a location. For ... @en		



Information about geo:lat

Metrics:

Overall score with current parameters	1.000
Best ratio search words in labels	1.000
Nb occurrences in main labels	1
Nb occurrences in secondary labels	1
Nb occurrences in LOD datasets	22486967
Nb occurrences in LOV datasets	0
Nb of vocabularies which reference this element	0


Element information:

URI	http://www.w3.org/2003/01/geo/wgs84_pos#lat
rdfs:type	owl:Thing > rdf:Property
vocabulary	geo
is dc:hasPart of	All > Space-Time > Space
rdfs:label	latitude
rdfs:comment	The WGS84 latitude of a SpatialThing (decimal degrees).

Close

Criteria for selecting a particular vocabulary/ontology

- Using specific score from the SW engine, e.g., Watson, LOV.
- Usage
- Maintenance
- Coverage
- Expressivity
-



Build the vocabulary by reusing the vocabularies we have found

- Available tools
 - Neologism
 - Protege
 - Topbraid composer
 - NeOn toolkit
 - SWOOP



Vocabulary Selection

Michael Hausenblas (DERI)

Ghislain Atemezing (EURECOM)

Rapaël Troncy (EURECOM)

Bernadette Hyland (3 Round Stones)

Boris Villazón-Terrazas (OEG-UPM)