WORK DONE AT MONDECA

25 JULY 2013

María Poveda Villalón

Ontology Engineering Group. Departamento de Inteligencia Artificial.
Facultad de Informática, Universidad Politécnica de Madrid.
Campus de Montegancedo s/n.
28660 Boadilla del Monte. Madrid. Spain
mpoveda@fi.upm.es

Acknowledgments: Consejo Social UPM

Index

- Why Mondeca?
- Publication Good Practices and Pitfalls: identification and description
- Automatic detection, debugging and visualization
- Integration proposals with LOV
- OOPS! integration

Why Mondeca?

- **Datalift** aims at providing tools to facilitate each step of the publication process*:
  - ***selecting ontologies for publishing data***
  - converting data to the appropriate format (RDF using the selected ontology)
  - publishing the linked data
  - interlinking data with other data sources
- **LOV** is main **input** for my thesis workflow
  - Extended with validation features
    - Dereferenceability
    - Content negotiation
    - Metadata
    - Bad/Best practices
  - Possible OOPS! integration (web service) in LOV?
- Evaluation features would be included in OOPS! (win win)

*http://datalift.org/

Publication Good Practices and Pitfalls: identification and description

Good practices

GP1. Provide RDF description (1)
GP2. Provide HTML documentation (2)
GP3. Content negotiation for RDF (2)
GP4. Content negotiation for HTML (2)
GP5. Provide *vann* metadata (2)
GP6. Well established/consistent/known Prefix (3)

Pitfalls

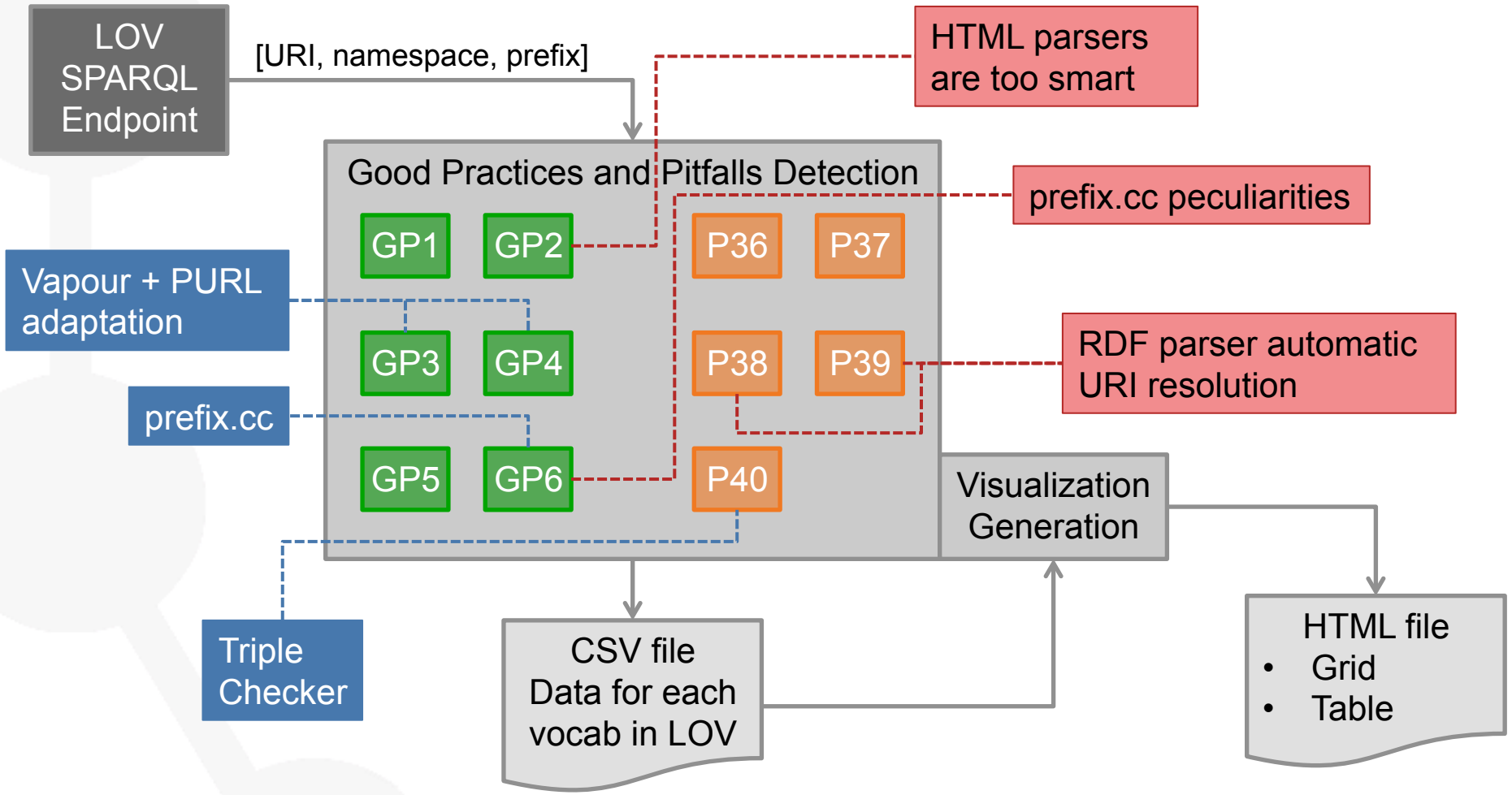
P36. URI contains file extension (3)
P37. Not available (1)
P38. No OWL ontology declaration (2)
P39. Ambiguous namespace (1)
P40. Namespace hijacking (1)

- **Critical (1):** this characteristic is crucial, it could affect the ontology consistency, reasoning, applicability, etc.
- **Important (2):** not critical for ontology function but important for reusability, processing, etc.
- **Minor (3):** It is not really necessary or a problem but meeting this good practice or avoiding this pitfall will make the ontology nicer.

Some sources and previous work:

- *Linked Open Data 5 Star rating system* (Tim Bernes-Lee) http://www.w3.org/DesignIssues/LinkedData.html
- *Is your linked data vocabulary 5-star?* (Bernard Vatant) http://bvatant.blogspot.fr/2012/02/is-your-linked-data-vocabulary-5-star_9588.html
- Archer, P., Goedertier, S., and Loutas, N. *D7.1.3 – Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC*. Deliverable. December 17, 2012.

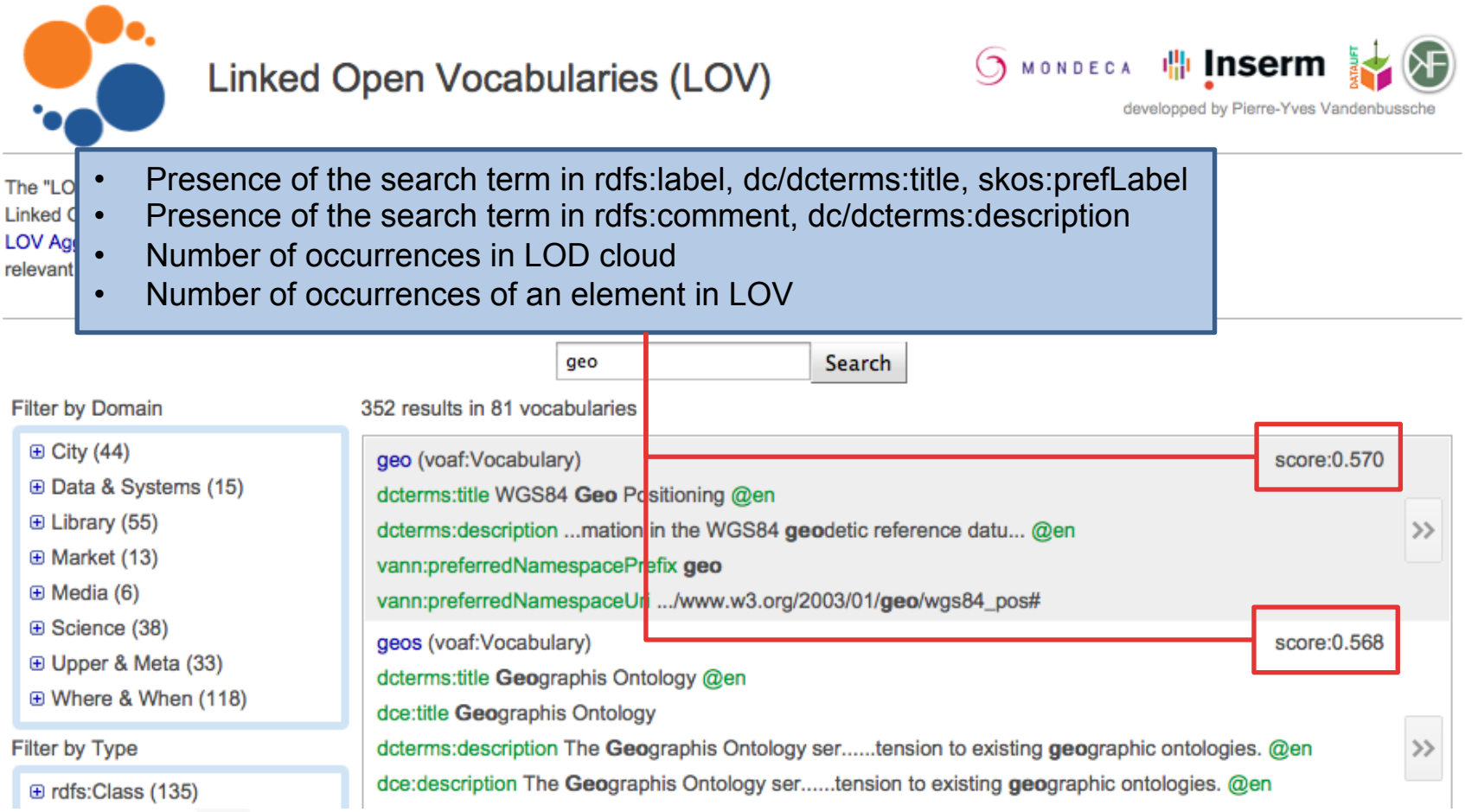
Automatic detection, debugging and visualization



Jena API: http://jena.sourceforge.net/
Java EE: http://www.oracle.com/technetwork/java/javaee/overview/index.html
HTML: http://www.w3.org/html/wg/
JSP: http://www.oracle.com/technetwork/java/javaee/jsp/index.html

CSS: http://www.w3.org/Style/CSS/
Vapour: http://validator.linkeddata.org/vapour
prefix.cc: http://prefix.cc/
Triple Checker: http://graphite.ecs.soton.ac.uk/checker/
Google Charts: https://developers.google.com/chart/?hl=es

Integration proposals (I)



- Presence of the search term in rdfs:label, dc/dcterms:title, skos:prefLabel
- Presence of the search term in rdfs:comment, dc/dcterms:description
- Number of occurrences in LOD cloud
- Number of occurrences of an element in LOV

- Tune the score according to the presence or absence of good practices and pitfalls.
- Weighted by importance: critical, important, minor
- Warning if the vocab is not available in RDF

Integration proposals (II)

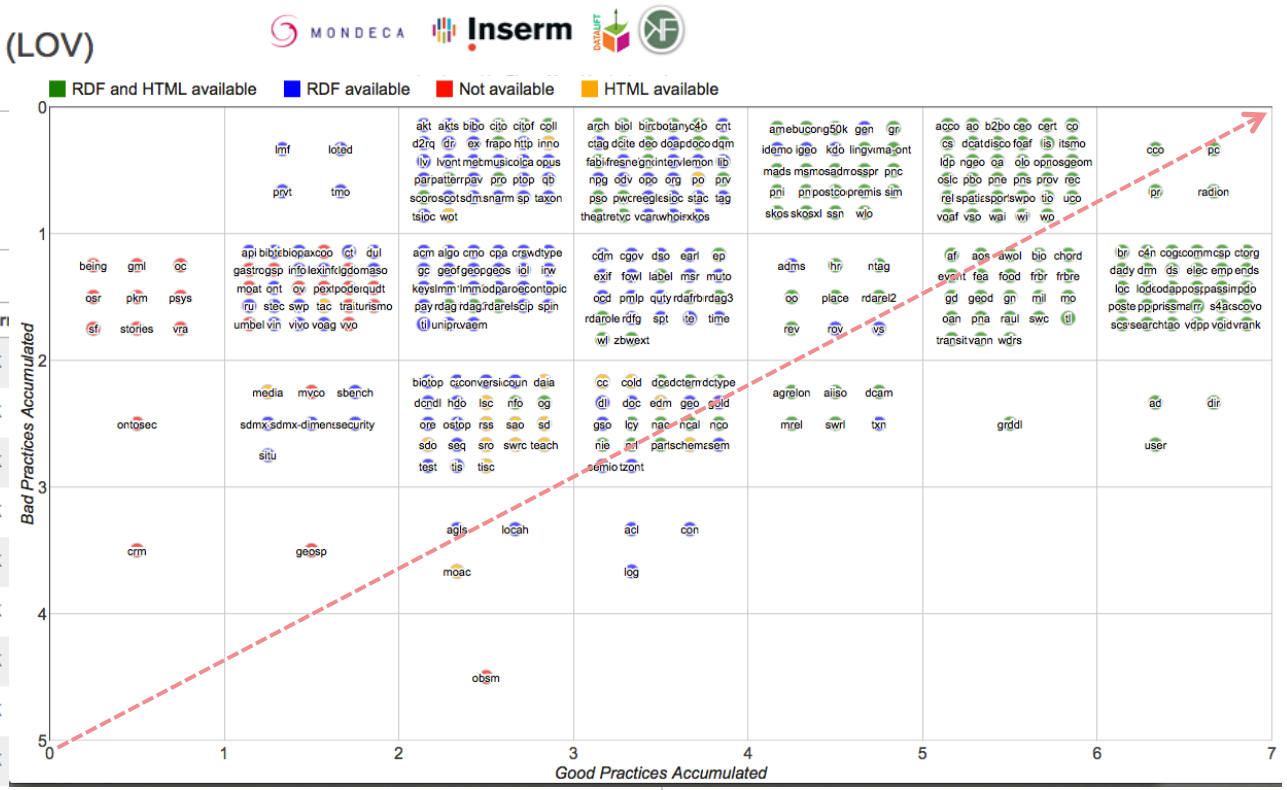
Linked Open Vocabularies (LOV)

LOV aggregator try to import all vocabularies in a single endpoint.

Vocabulary aggregation (300/355):

Prefix	Title	Curi
acco	Accomodation Ontology	OK
acl	Basic Access Control ontology	OK
acm	ACM Classification Ontology	OK
ad	Address Schema	OK
adms	Asset Description Metadata Schema	OK
af	Audio Features Ontology	OK
agls	AGLS Metadata Terms	OK
agrelon	Agent Relationship Ontology	OK
aiiso	Academic Institution Internal Structure Ontology	OK
akt	AKT Reference Ontology	OK
akts	AKT Support Ontology	OK
algo	Algorithms Ontology	OK
am	OSLC Asset Management Vocabulary	OK
ao	Association Ontology	OK
aos	Appearances Ontology Specification	Invalid URL or Content Negotiation problem
api	Linked Data API Vocabulary	OK
arch	Archival collections ontology	OK
awol	Atom Syndication Ontology	OK

date: 2013-06-06	
date: 2013-06-06	
date: 2013-06-06	
date: 2013-06-22	
date: 2013-06-06	
No file ever fetched !	
date: 2013-06-06	
date: 2013-06-06	



- Provide more detailed information
- Provide an overview of availability and quality (grid)

Integration proposals (III)

- Provide evaluation tips or reports for each vocabulary.
- Natural language text?

OOPS! integration (I)

- Create Good Practice Catalogue

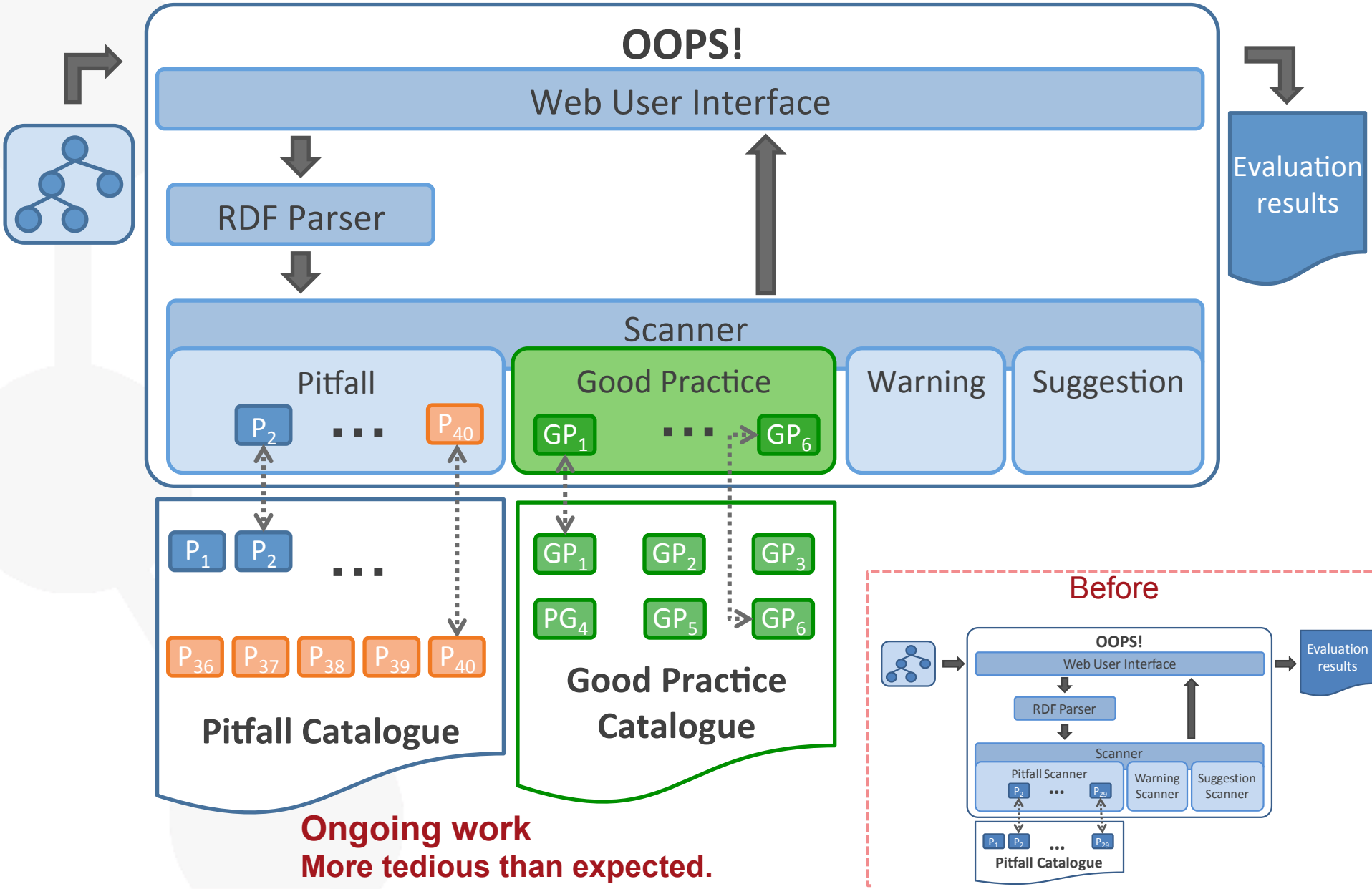
Catalogue of common pitfalls

Here you can find a catalogue of pitfalls that usually appear when developing ontologies. Some of them are very common and have been identified by several works about ontology evaluation (see References).

. . .

- **P36. URI contains file extension:** Guidelines in [ISA] suggest avoiding file extension in persistent URIs, particularly those related to the technology used, as for example ".php" or ".py". In our case we have adapted it to the ontology web languages used to formalized ontologies and their serializations. In this regard, we consider as pitfall including file extensions as ".owl", ".rdf", ".ttl", ".n3" and ".rdfxml" in an ontology URI. An example of this pitfall is defining an ontology uri as "http://www.biopax.org/release /biopax-level3.owl" containg the extension ".owl" related to the technology used.
- ★ *(Linked Data Feature)* **P37. Ontology not available:** This bad practice is about not meeting LOD1 from Linked Data star system that stars "On the web" and LDV1 that says "Publish your vocabulary on the Web at a stable URI". An example of this pitfall could be the following case: "Ontology Security (ontosec)" which URI is http://www.semanticweb.org/ontologies/2008/11 /OntologySecurity.owl and it is not available online as RDF nor as HTML (at the moment of carrying out this work).
- **P38. No OWL ontology declaration:** The owl:Ontology tag aims at gathering metadata about a given ontology as version information, creation date, etc. It is also used to declare the inclusion of other ontologies. Not declaring this tag is consider as a bad practice for owl ontologies as it is a symptom of not providing useful metadata as proposed in "LDV2". Example: "Creative Commons Rights Expression Language (cc)" ontology with URI http://creativecommons.org/ns does not have any owl:Ontology declaration in its RDF file even though there are other OWL elements used as, for example, owl:equivalentProperty.
- ★ *(Linked Data Feature)* **P39. Ambiguous namespace:** In the case of not having defined the ontology URI nor the xml:base namespace the ontology namespaces is matched to the file location. This situation is not desirable as the location of a file might change while the ontology should remain stable as proposed in "LDV1". Example: "Basic Access Control ontology (acl)" with URI http://www.w3.org/ns/auth/acl has no owl:Ontology tag nor xml:base definition.
- ★ *(Linked Data Feature)* **P40. Namespace hijacking:** this bad practice refers to the situation when an ontology is reusing or referring to terms from other namespaces that are not defined in such namespace. This is an undesirable situation as no information could be retrieve when looking up those undefined terms, in addition, there would be no meaning or semantic behind them. In addition this practice is against Linked Data publishing guidelines provided in [LDbook] "Only define new terms in a namespace that you control." Example: the "WSMO-Lite Ontology (wl)" which URI is http://www.wsmo.org/ns/wsmo-lite#, uses http://www.w3.org /2000/01/rdf-schema#Property" that is not defined in the rdf namespace (http://www.w3.org/2000/01/rdf-schema#) instead of using http://www.w3.org/1999/02/22-rdf-syntax-ns#Property, that is ac-tually defined in the rdfs namespace (http://www.w3.org/1999/02 /22-rdf-syntax-ns#).

OOPS! integration (II)



Last but not least

Linked Open Vocabularies (LOV)

MONDECA

develop

Welcome to LOV, your entry point to the growing ecosystem of linked open vocabularies (RDFS or OWL ontologies) used in the Linked Data Cloud. Here you will find vocabularies listed and individually described by metadata, classified by vocabulary spaces, interlinked using the dedicated vocabulary VOAF.
You will enjoy querying the LOV dataset either at vocabulary level or at element level, exploring the vocabulary content using full-text faceted search, and finding metrics about the use of vocabularies in the Semantic Web.
Not finding your favourite one? Suggest a new vocabulary to add to LOV!
Read more ...

RDF

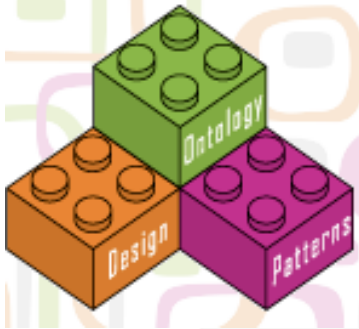
Endpoint

Search

Metadata:

Property	Value
Creator	Bernard Vatant, Pierre-Yves Vandenbussche
Contributor	Ghislain Atemezing, María Poveda-Villalón, Thomas Francart

- LOV edition and curating tasks
- Became a contributor ☺



wop | discussion | view source | history

WOP:Main

Welcome to the workshop series on Ontology Patterns.

Contents [hide]

- Documentation
- Submitted as research paper to WOP2013

Questions?

Thanks!

WORK DONE AT MONDECA

25 JULY 2013

María Poveda Villalón

Ontology Engineering Group. Departamento de Inteligencia Artificial.
Facultad de Informática, Universidad Politécnica de Madrid.
Campus de Montegancedo s/n.
28660 Boadilla del Monte. Madrid. Spain
mpoveda@fi.upm.es

Acknowledgments: Consejo Social UPM