EURECOM
Sophia Antipolis

MINES TELECOM
INSTITUT
Mines-Télécom

Publishing and Consuming Geo-spatial and Government Data on the Semantic Web

Ghislain Auguste Atemezing

Multimedia Department, Eurecom

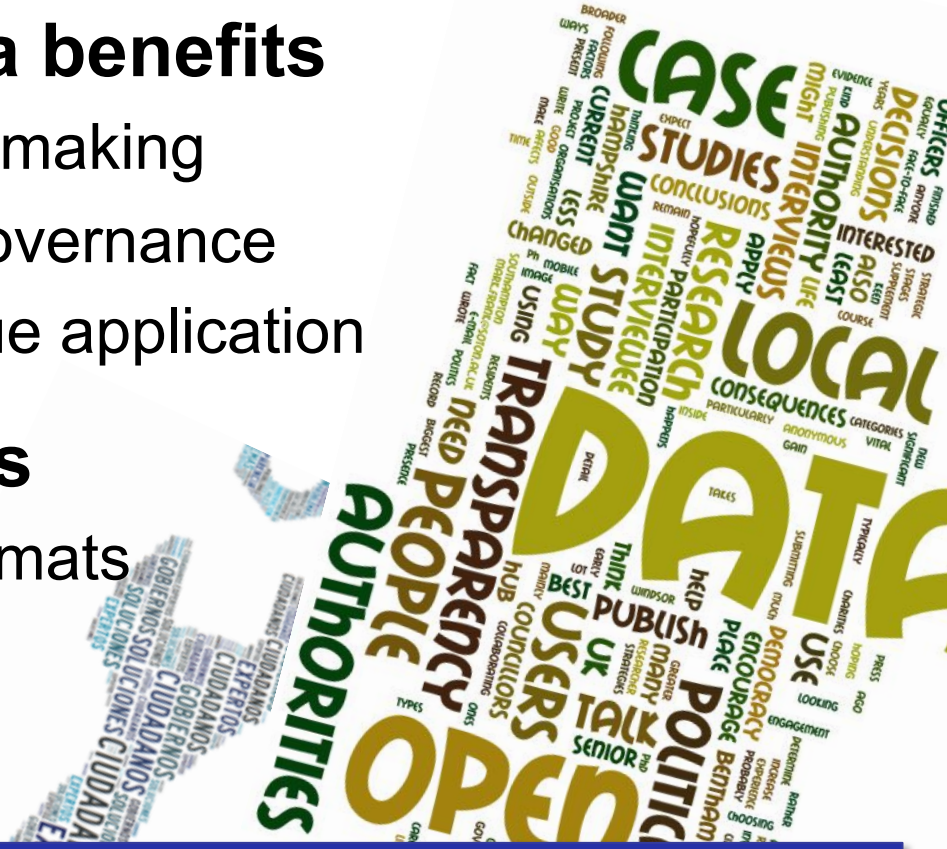
DATALIFT

FINANCÉ PAR ANR

Supervisor:
Dr. Raphaël Troncy, MM Department, Eurecom

Thesis Context

- **Open Government Data benefits**
  - Transparency in decision making
  - Better governance or e-governance
  - Eco-system of added value application
- **Barriers and challenges**
  - Heterogeneity of data formats
  - Variety of access method
  - Lack of nomenclature



In this thesis we explore how Semantic Web technologies can be used for better integration and consumption of geo-spatial data.

Outline

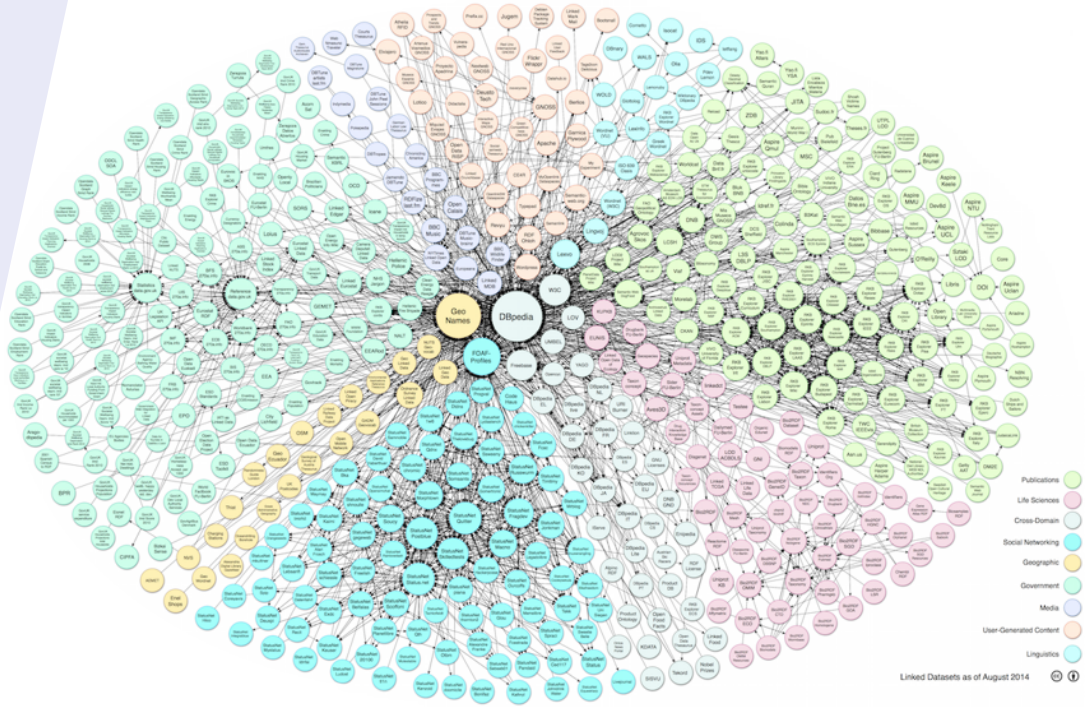
- **Introduction**
- **Research Questions**
- **Contributions**
  - Semantic publishing of geospatial data
  - Visualizations of Government Linked Data
  - Best Practices for metadata in vocabularies
- **Conclusion**
- **Publications**

Geospatial data: why it matters

“80% of needs for decisions from public authorities have a geospatial component”.

(Philippe Grelot, IGN-France)

GeoData on the LOD Cloud

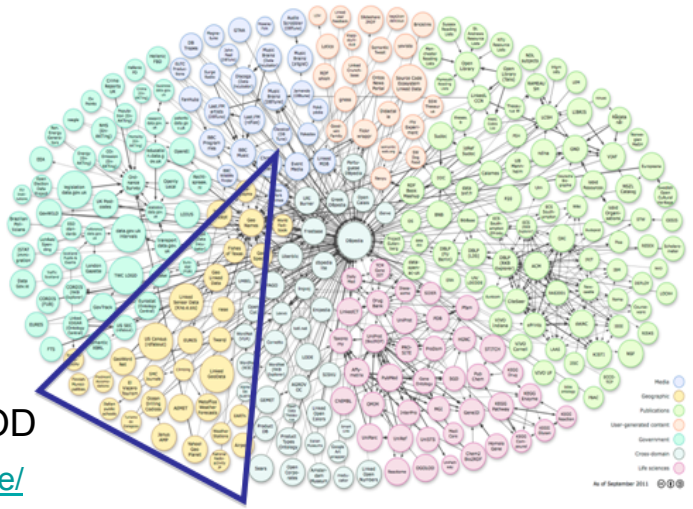


Datasets by topical domain.

Topic	Datasets	%
Government	183	18.05%
Publications	96	9.47%
Life sciences	83	8.19%
User-generated content	48	4.73%
Cross-domain	41	4.04%
Media	22	2.17%
Geographic	21	2.07%
Social web	520	51.28%
Total	1014	

http://lod-cloud.net/versions/2014-08-30/

Linking Open Data cloud diagram 2014, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak.
http://lod-cloud.net/



In 2011 19,43% →31 geo-datasets in LOD

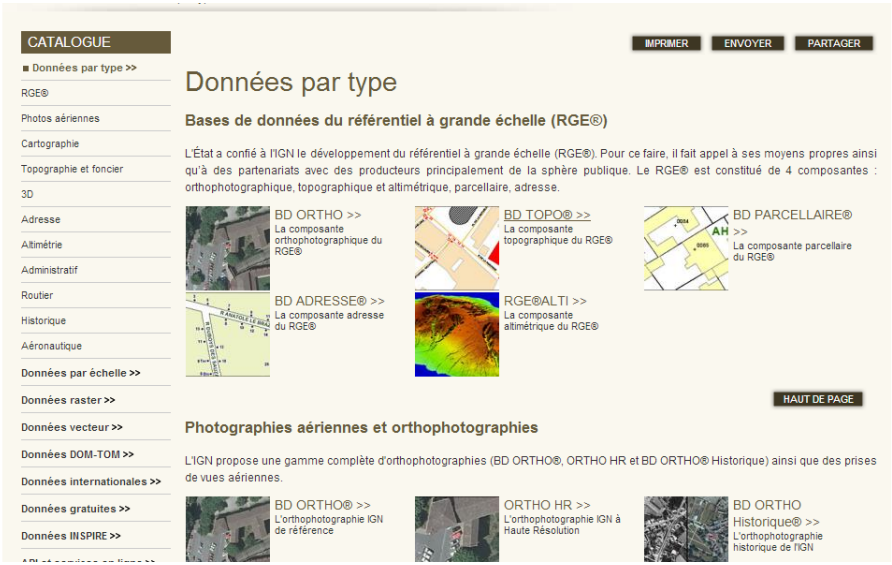
http://lod-cloud.net/state/

French IGN- Reference Geodata

« ..describes the French national territory and the occupation of its land, elaborates and updates perpetual inventory of the forest resources »

- ✓ Different databases with **overlapping** information:
  BD ORTHO, BD PARCELLAIRE, POINT ADRESSE, BD ALTI 25m, BD TOPO; etc.
- ✓ CRS: LAMBERT93 or RGF93

Q: **”Give me all the bridges in a radius of 2km from the "Eiffel Tower“?**
A: Not straightforward



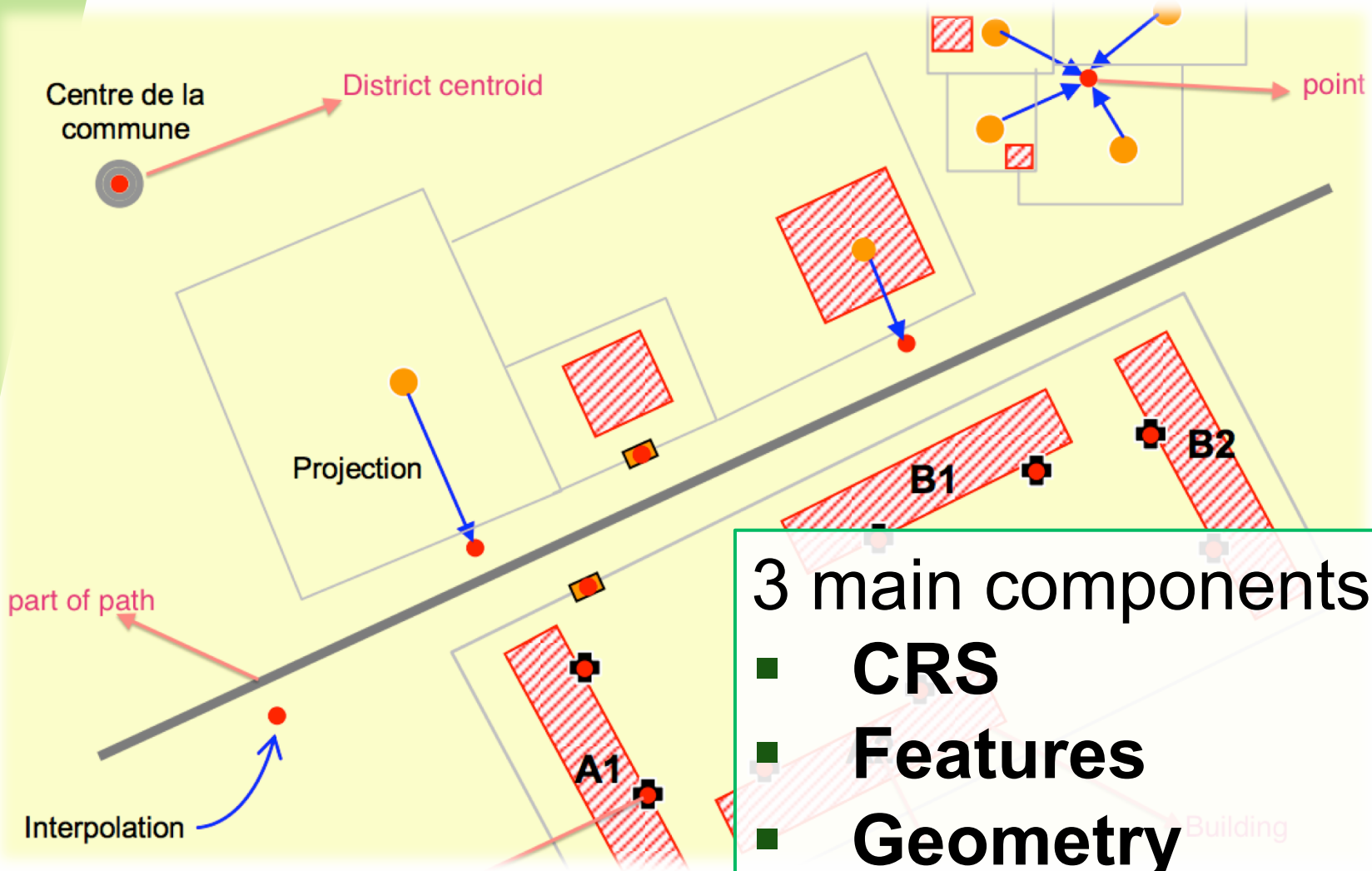
Research Questions

- **How to efficiently represent and store geospatial data on the Web to ensure interoperable applications?**
  - the publication of real datasets
  - the interlinking with other datasets having overlapping coverage
- **What are the best options for a user to discover, browse and interact with semantic content?**
  - Can we propose a generic model for visualizing semantic content?
- **What are the mechanisms to help preserving structured data of a high quality on the Web?**
  - How to improve reusing vocabularies for better interoperability
  - How to detect incompatibility between data and metadata

Part 1

Semantic Publication of Geo-spatial Data

Modeling geospatial objects



3 main components:

- **CRS**
- **Features**
- **Geometry**

Coordinate Reference Systems (CRS)

- **A representation of the locations of geographic features within a common geographic framework**
- **Each CRS is defined by its measurement framework**
  - Geographic: spherical coordinates (unit: decimal degrees)
  - Planimetric: 2D planar surface (unit: meters) + a map projection
  - Additional measurement properties: ellipsoid, datum, standard parallels, central meridians, etc.
- **Situation:**
  - Several hundred Geographical Coordinate Systems (WGS84, ETRS89, etc.)
  - Several thousands Projected Coordinate Systems (UTM, Lambert93, etc.), http://resources.esri.com/help/9.3/arcgisserver/apis/rest/pcs.html

CRS in France (Metropolitan + overseas)

- **10 CRSs coverage (mostly RGF93)**
- **10 projections (mostly used Lambert93)**
- **3 different ellipsoids to define the CRS**

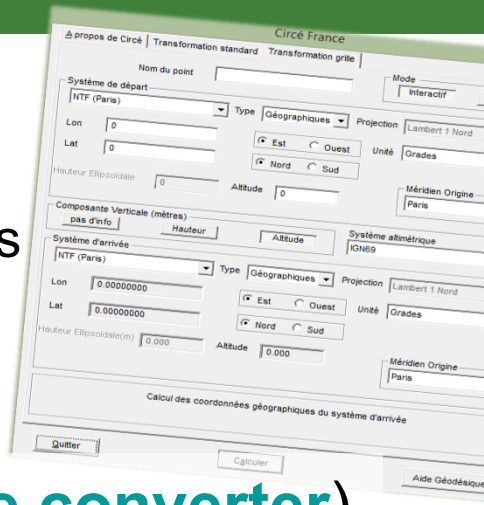
The Web only used WGS84

Publishers need converters for their geodata

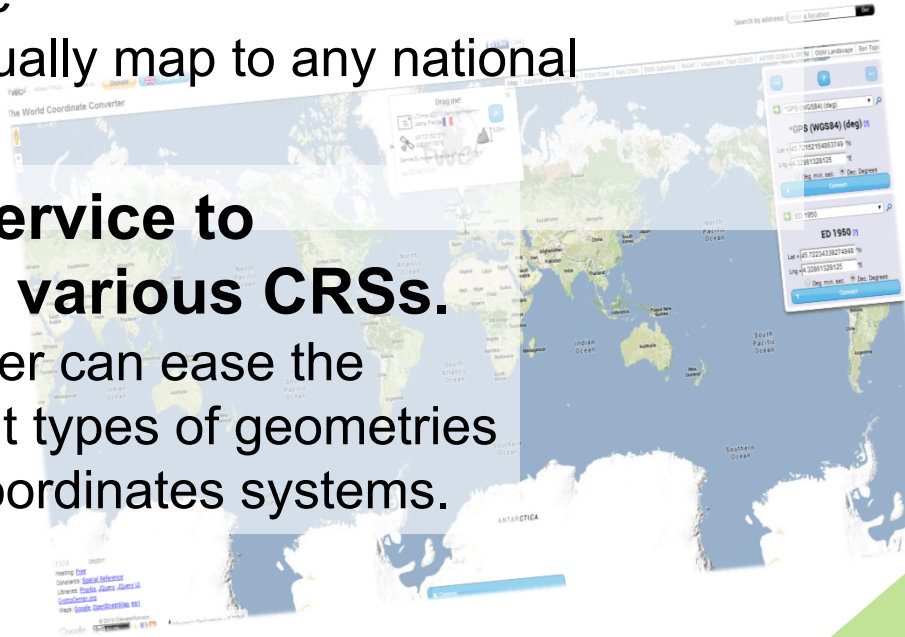
REGION	COORDINATE SYSTEM	ELLIPSOID	PROJECTION SYSTEM	ALTIMETRY SYSTEM
FRANCE METROPOLITAN	RGF93	IAG GRS 1980	Lambert 93 and CC 9 Zones	
MAYOTTE	RGM04 (ITRF2000)	IAG GRS 1980	UTM 38 South	SHOM 1953
GUYANE	RGFG95	IAG-GRS 1980	UTM 21 22 North	
MARTINIQUE	WGS84	IAG-GRS 1980	UTM 20 North	
GUADELOUPE	WGS84	IAG-GRS 1980	UTM 20 North	
LA RÉUNION	RGR92	IAG-GRS 1980	UTM 40 South	GGR 99
NOUVELLE-CALÉDONIE	ITRF90	IAG-GRS 1980	UTM 5, 6, 7 and 8 South	Tahiti IGN 1966
POLYNÉSIE	RGPF	IAG-GRS 1980		
WALLIS ET FUTUNA	MOP87	International 1924		
SAINT-PIERRE ET MIQUELON	RGM01 (ITRF2000)	IAG GRS 1980	UTM 21 North	Danger 1950
ILE CLIPPERTON	Marine 1967	International	UTM 12 South	

CRS Converters and Limitations

- **Circé: A Converter Software from IGN**
  - Allows conversion between some France CRSs to WGS84
  - Closed Source, No open service



- **Existing Web tools** (e.g., world coordinate converter)
  - No open service, closed source
  - Converting algorithms don't usually map to any national mapping authority.



- **Contribution: a Web based service to perform conversion between various CRSs.**
  - The integration of such a converter can ease the process of the publishing different types of geometries on the web regardless of their Coordinates systems.

Developing a REST converter service

- Many regional's' published data involve with local CRS

International Communication

A medium

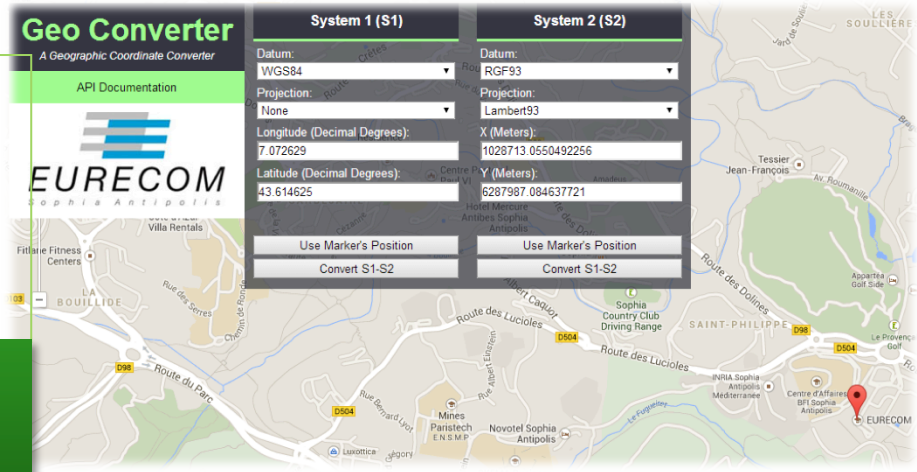
- Interpret the coordinates between these CRSs

- Open services for community
- RESTFul Web Service

A Converter Service

- **WGS 84 <–> Lambert 93**
- **WGS 84 <–> UTM**

GOOD **accuracy** of the results

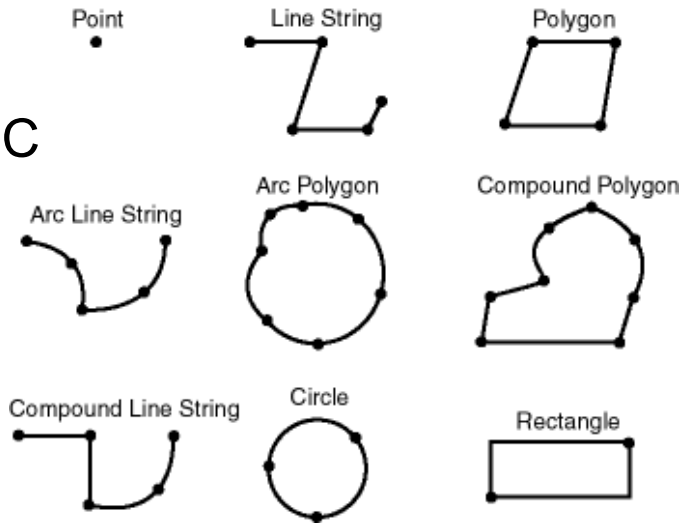


Vocabularies for Modeling Features

- **Authority list of terms (e.g. Foursquare)**
  - No semantics at all!
- **SKOS Categories (e.g. GeoNames)**
  - Classes are `skos:conceptScheme`, codes are `skos:Concept`
- **Domain specific ontologies (OrdS, GeoLD)**
  - Interconnected subdomain ontologies (transport, hydrography, etc.)
- **Data driven ontologies (LGD, GeOnto)**
  - Deeper taxonomy to structure the ontology
  - Many classes

Modeling Geometry: State of the Art

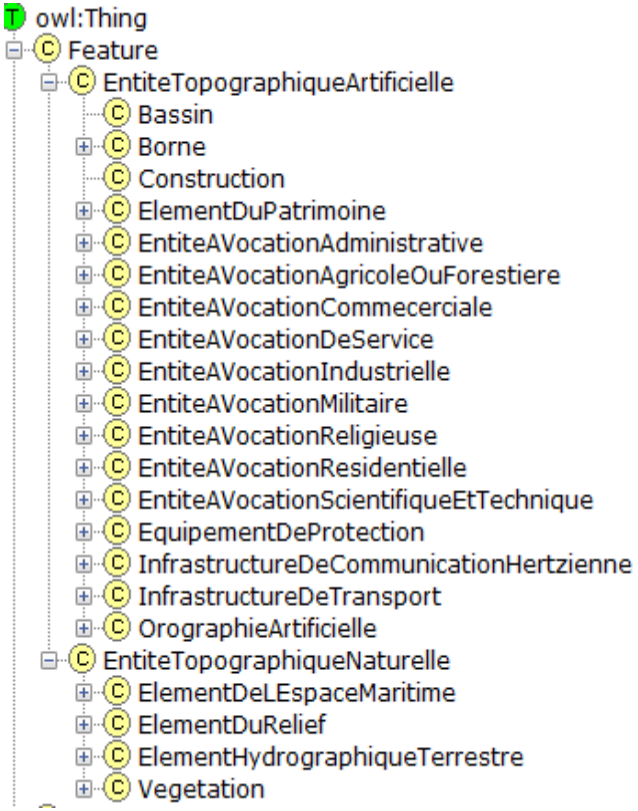
- **Point (lat/long)**
  - WGS 84 vocabulary described by W3C
- **Rectangle ("bounding box")**
  - Geopolitical Vocabulary (FAO)
- **Points in a List**
  - Sequence of points (LinkedGeoData)
  - An object is "*formedBy*" a ListOfPoints (GeoLinkedData.es)



- **Literals WKT datatype in RDF**
  - Ordnance Survey (UK), GeoSPARQL embedding CRS in literals
- **More structured representation of complex geometry**
  - NeoGeo Vocabulary (GeoVocamp), **http://geovocab.org/**

Reusing Existing Ontologies (GeOnto)

- **Ontology for geographic objects (POI)**
  - Output of a French (ANR) research project
  - Obtained from NLP tools
- **Classes in French**
  - `rdfs:labels` in FR & EN
  - No `rdfs:comments`
  - Few `owl:ObjectProperty`
  - 783 classes
- **Overlap with other vocabs**
  - Need for alignment

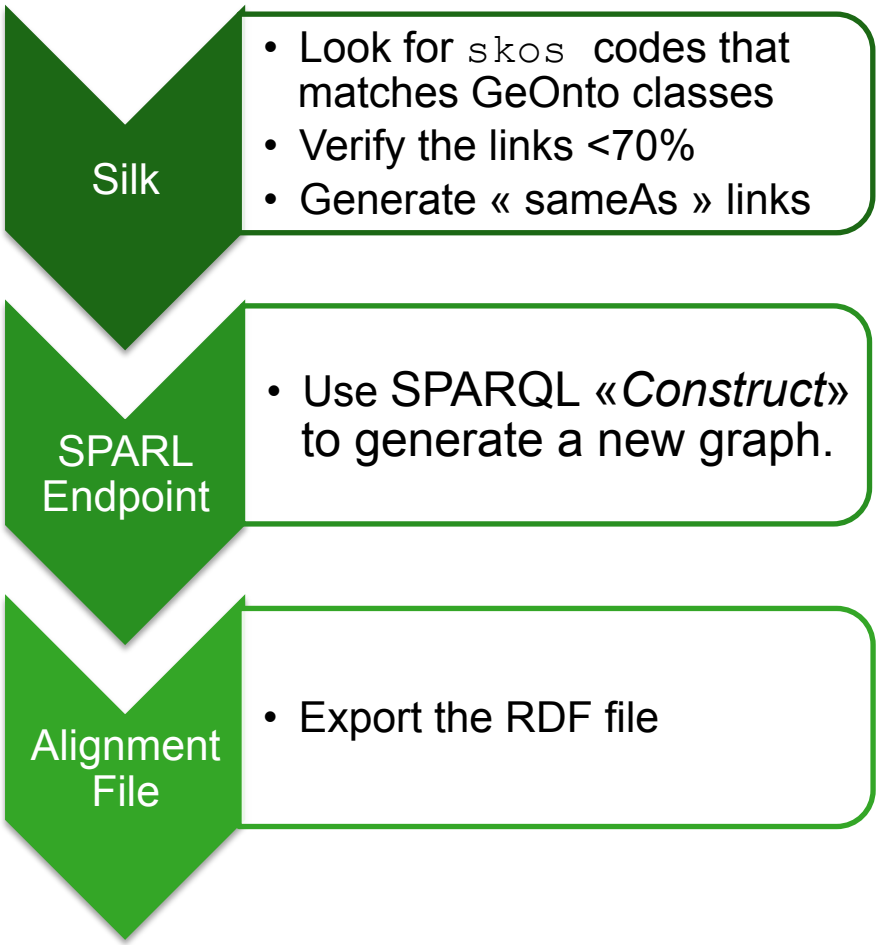


Aligning GeOnto with existing Ontologies

- **Alignment of GeOnto with 5 ontologies and 2 simple taxonomies**
  - LGD, DBpedia, Schema.org, GeoNames, bdtopo
  - Foursquare, Google Places
- **Goal: finding `owl:equivalentClass`**
  - Tool : Silk framework
  - Metrics : `LevenshteinDistance, Jaro`
  - Labels : @en des classes
  - Aggregation Function: Mean
- **Manual validation**
  - For « `rdfs:subClassOf` »
  - Specific alignments with GeoNames codes

Alignment Process (GeoNames)– Results

Vocab/ taxonomies	#Classes	#Classes aligned
Bdtopo	237	153 (64.65%)
GeoNames	699	287 (41.06%)
Google Place (*)	126	41 (32.54%)
Schema.org (*)	296	52 (17.57%)
LGD	1294	178 (13.76%)
Foursquare	359	46 (12.81%)
DBpedia	366	42 (11.48%)



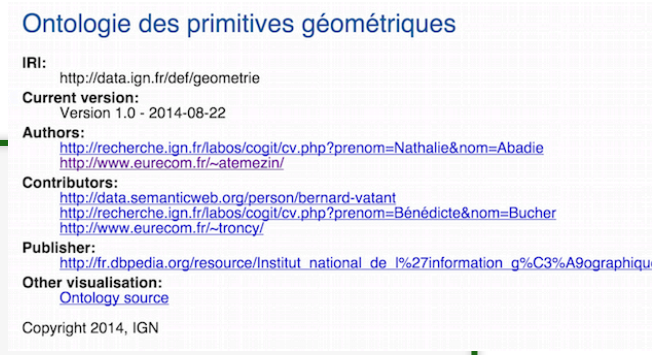
- **High precisions > 80%**
- **BUT P(Schema.org) = 50%** → GeOnto entities are more specific to France

Proposal: CheckList for modeling geodata

- **Complex Geometry Coverage**
  - Need to publish more data with complex geometries
  - Reuse and extend suitable ontologies (NeoGeo, GeoSPARQL)
- **Features MUST be connected to Geometry**
  - Sometimes it may requires two namespaces
- **Serialization in other GIS formats**
  - Provide serialization in other GIS formats (GML, WKT, KML, etc.)
- **Structured Representation**
  - Use of structured representation for complex geometry
  - This covers some of the Use Cases at IGN

Proposal: Vocabulary for complex Geometry

- **Extend** and **reuse** existing vocabularies
- Available at http://data.ignf.fr/def/geometrie



Ontologie des primitives géométriques

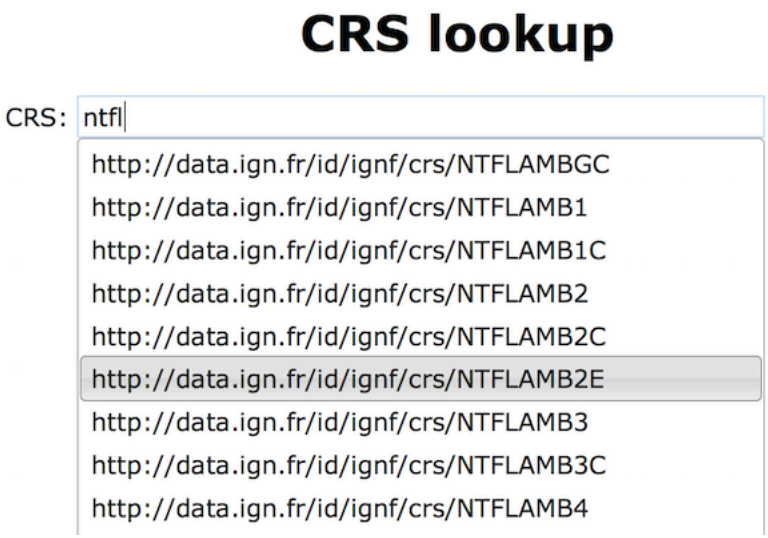
IRI:
http://data.ign.fr/def/geometrie
Current version:
Version 1.0 - 2014-08-22
Authors:
http://recherche.ign.fr/labos/cogit/cv.php?prenom=Nathalie&nom=Abadie
http://www.eurecom.fr/~atemezin/
Contributors:
http://data.semanticweb.org/person/bernard-vatant
http://recherche.ign.fr/labos/cogit/cv.php?prenom=Bénédicte&nom=Bucher
http://www.eurecom.fr/~troncy/
Publisher:
http://fr.dbpedia.org/resource/Institut_national_de_l%27information_g%C3%A9ographiqu
Other visualisation:
Ontology source

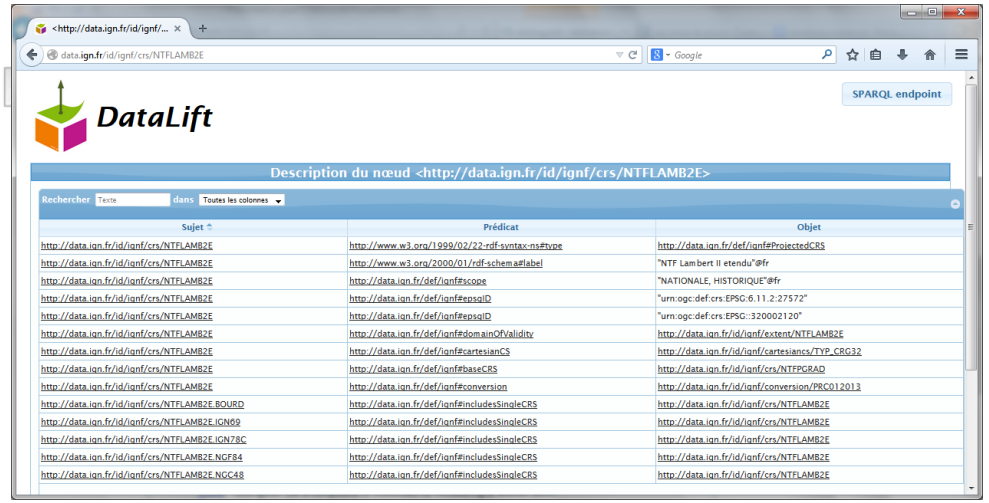
Copyright 2014, IGN

```
@prefix ngeo: <http://geovocab.org/geometry#>.
@prefix sf: <http://www.opengis.net/ont/sf#>.
[…]
geom:Geometry a owl:Class;
        rdfs:comment "Primitive géométrique non instanciable, racine de
                l'ontologie des primitives géométriques. Une géométrie est
        associée à un système de coordonnées et un seul."@fr;
        rdfs:label "Géométrie"@fr, "Geometry"@en;
        owl:equivalentClass [ a owl:Restriction;
                owl:onClass ignf:CoordinatesSystem;
                owl:onProperty geom:crs;
                owl:qualifiedCardinality "1"^^xsd:nonNegativeInteger];
        rdfs:subClassOf ngeo:Geometry;
        rdfs:subClassOf sf:Geometry.
```

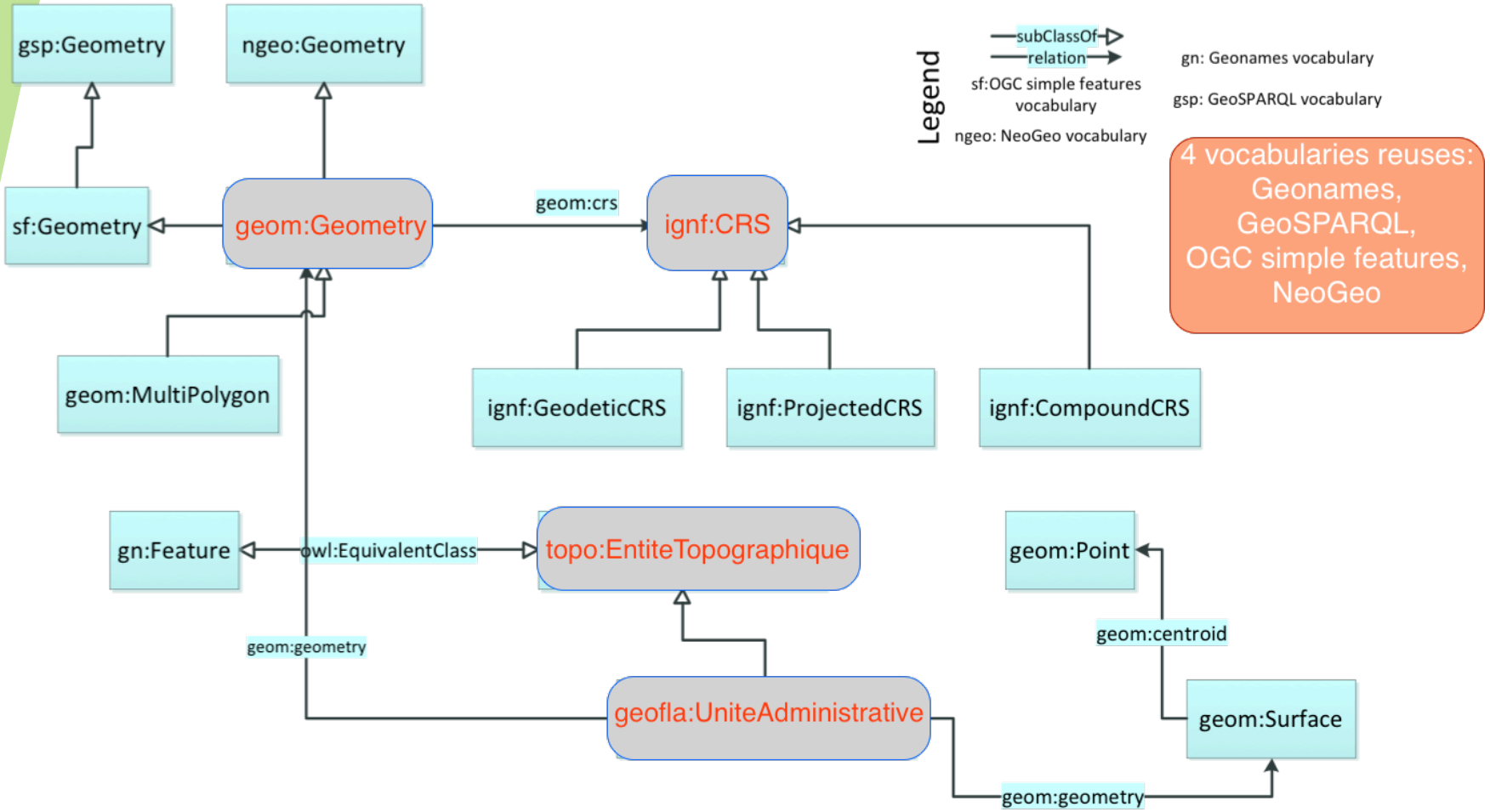
Associating CRS to Geometries

- **A vocabulary for describing CRS**
  - Subset of ISO 19111 model
  - Available at http://data.ignf.fr/def/ingf
- **A dataset for French CRS**
  - Convert from XML data published by IGN France to RDF
  - Eg: "*Lambert 2 étendu*" http://data.ign.fr/id/ignf/crs/NTFLAMB2E



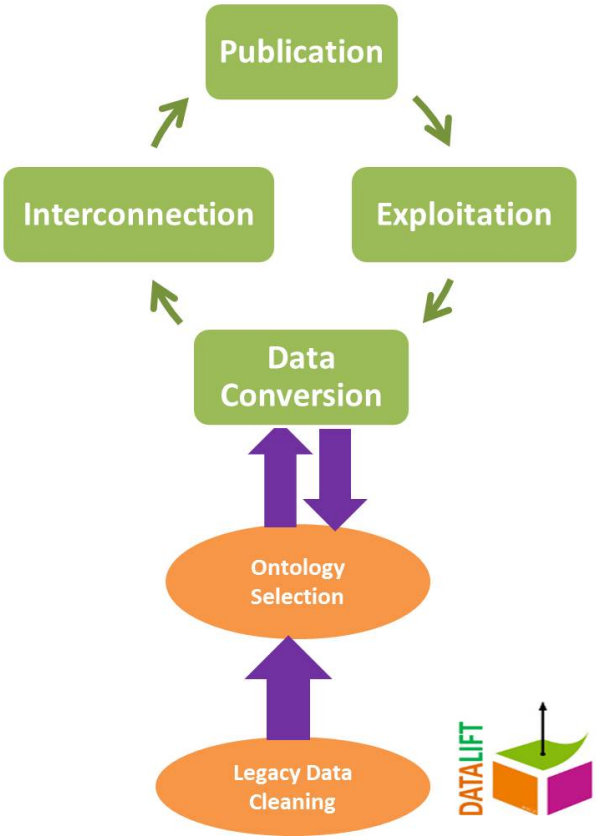


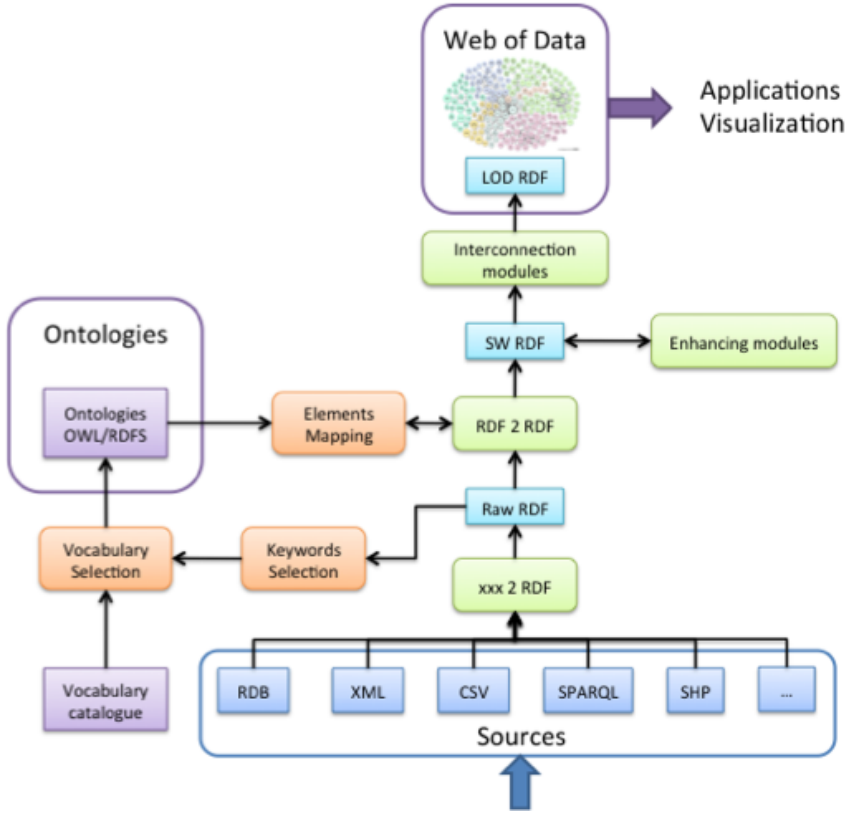
Overview of the vocabularies and relationships



A workflow for publishing Linked(Geo)Data

- **DATALIFT: a set of modular tools for "lifting" raw data in RDF.**





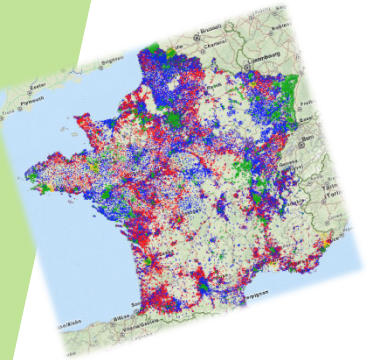
Publishing French Reference Geodata in RDF

- **GEOFLA® DATASET ON FRENCH ADMINISTRATIVE UNITS IN RDF**
  - Features are of type http://data.ign.fr/geofla
- **FRENCH GAZETTEER DATASET**
  - Features are of type http://data.ign.fr/topo

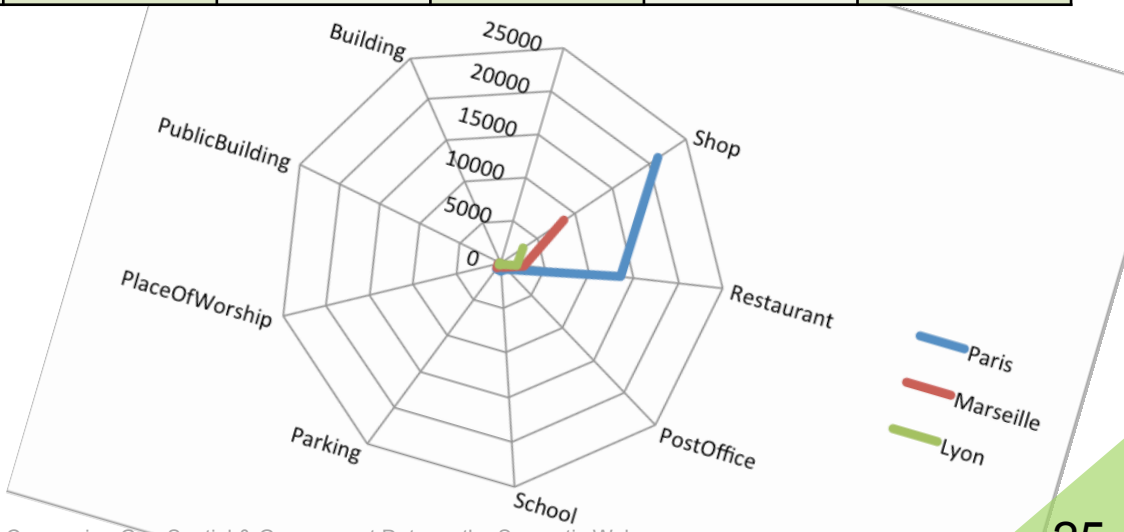
Mapping results with external dataset			
DATASET		#mappings	Tool
GEOFLA DATASET	INSEE	37,020	SILK
	DBPEDIA-FR	23, 252 (communes) and 93 departments	LIMES
	NUTS	105 links (14 comm., 75 depts, 16 regions)	SILK
	GADM	70 links (10 comm., 51 depts, 9 regions)	SILK
FRENCH GAZETTEER	LINKED GEODATA	654 links (lgdo:Amenity)	LIMES

French Open Addresses in RDF Mappings

BANO2RDF

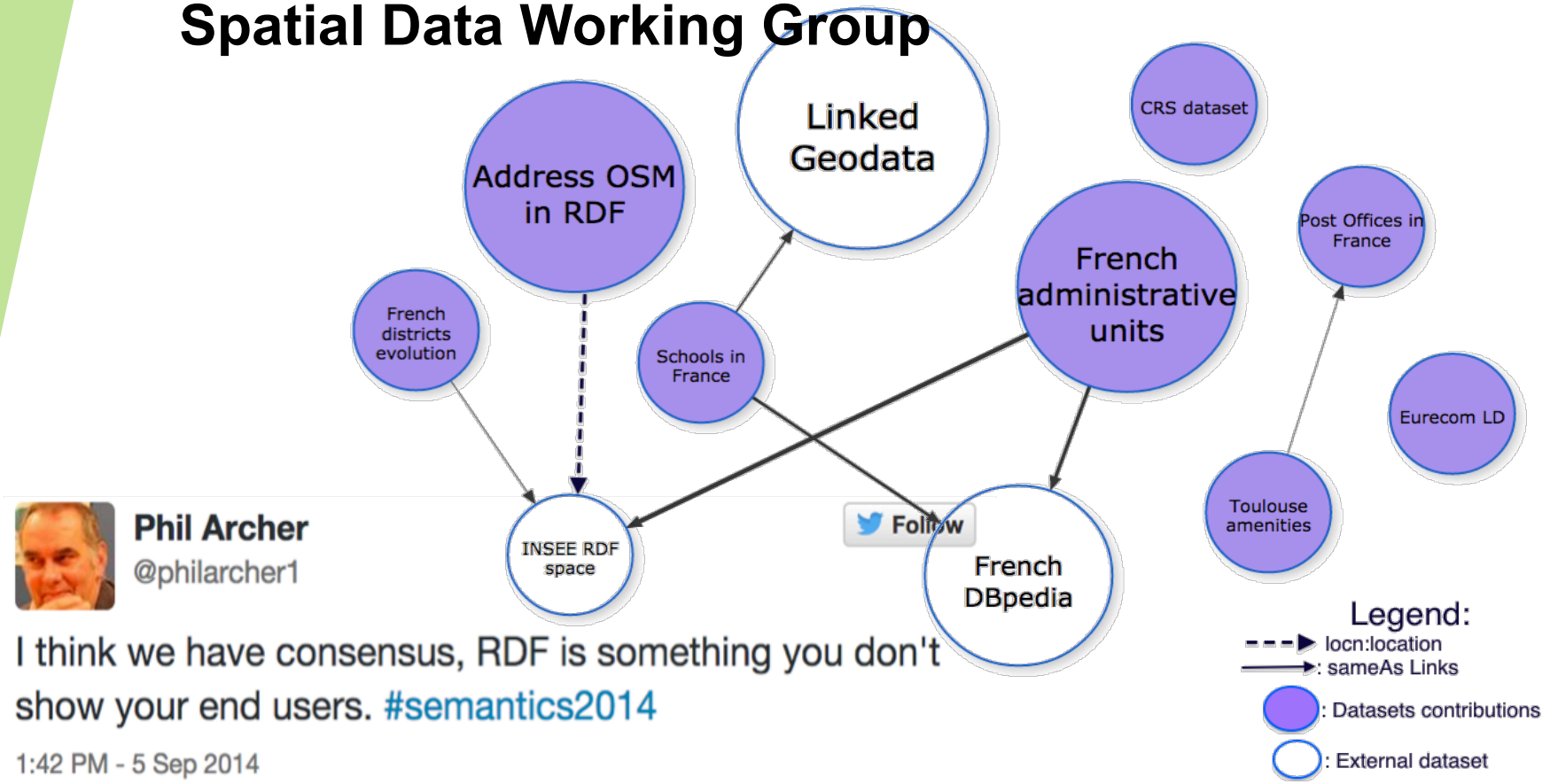


LGData Amenities	Paris (248052)		Marseille (401404)		Lyon (89061)	
	#matched	%matched	#matched	%matched	#matched	%matched
Shop (778680)	21171	**2.71**	8556	**1.098**	3049	**0.391**
Restaurant (260675)	13567	**5.204**	2654	**1.018**	1882	**0.721**
PostOffice (87731)	971	**1.106**	555	**0.632**	173	**0.197**
School (318287)	883	**0.277**	411	**0.129**	197	**0.061**
Parking (250516)	735	**0.293**	625	**0.24**	210	**0.083**
PlaceOfWorship (357445)	272	**0.076**	193	**0.053**	31	**0.008**
PublicBuilding (26735)	97	**0.362**	64	**0.239**	21	**0.078**
Building (22283)	5	**0.022**	12	**0.053**	0	**0**



Contribution to the LOD Cloud (FrLOD)

- 340 millions of triples in RDF (10% of DBPedia 2014)
- Part of our work is under reused by the W3C/OGC Spatial Data Working Group



Phil Archer
@philarcher1

Follow

I think we have consensus, RDF is something you don't show your end users. #semantics2014

1:42 PM - 5 Sep 2014

25 RETWEETS 14 FAVORITES

Why Visualizations Matters

"Don't ask what you can do for the Semantic Web; ask what The Semantic Web can do for you!"
(D. Karger, MIT CSAIL)

"If you use our Linked Data, please let us know, or we might switch off!"
(Ordnance Survey)

1. How to build **bridge** to fill the gap between traditional InfoVis tools and Semantic Web technologies
2. How can Semantic Web help in visualization?

Part 2

Generating Visualizations For Linked Data

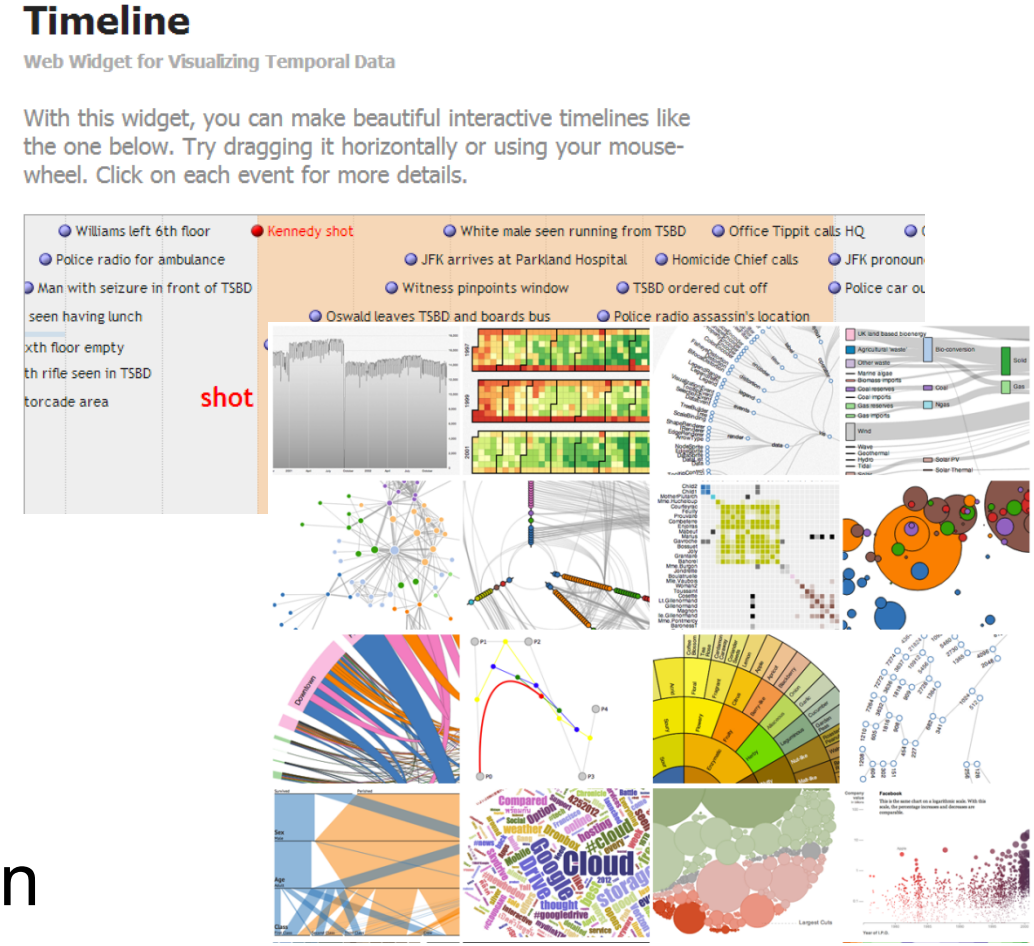
Visualization Categories in Government Portals

- **Study of applications consuming Open Data**
  - 13 applications from UK (7), USA (3) and France (3)
  - Domain: education, health, transport, government, city, housing, criminality, foreign aid

- **Different dimensions**
  - Platform (web, mobile), data sources, which views are available (maps, charts, timeline, etc.)
  - URL policy for identifying data objects
  - Licenses for the application / for the data
  - Commercial / non-commercial

Relevant Features in Visualization Tools

- Data format given as input (csv, xml, shp, rdf, etc.)
- Data access (API, dump, etc.)
- Language code
- Type of view
- External Libraries
- License
- Metadata: author, organization



Describing an Application: Opendatacom

Scope/Domain: Department for Communities and Local Government, datasets access

Description: visualize available datasets (finance, housing, deprivation, geography) by authorities or postcode.
On the dashboard, it provides graphs showing the national distribution of a district and how the values for this local authority compare with others in England.

Supported Platform: Web

URL Policy: http://{domain}/**id**/{...} with redirection to the corresponding document at: http://{domain}/**doc**/{...}.
Hampshire County Council is:
http://opendatacommunities.org/id/county-council/hampshire

Data Sources: 36 datasets from DCLG, Administrative Geography and Postcodes from Ordnance Survey.

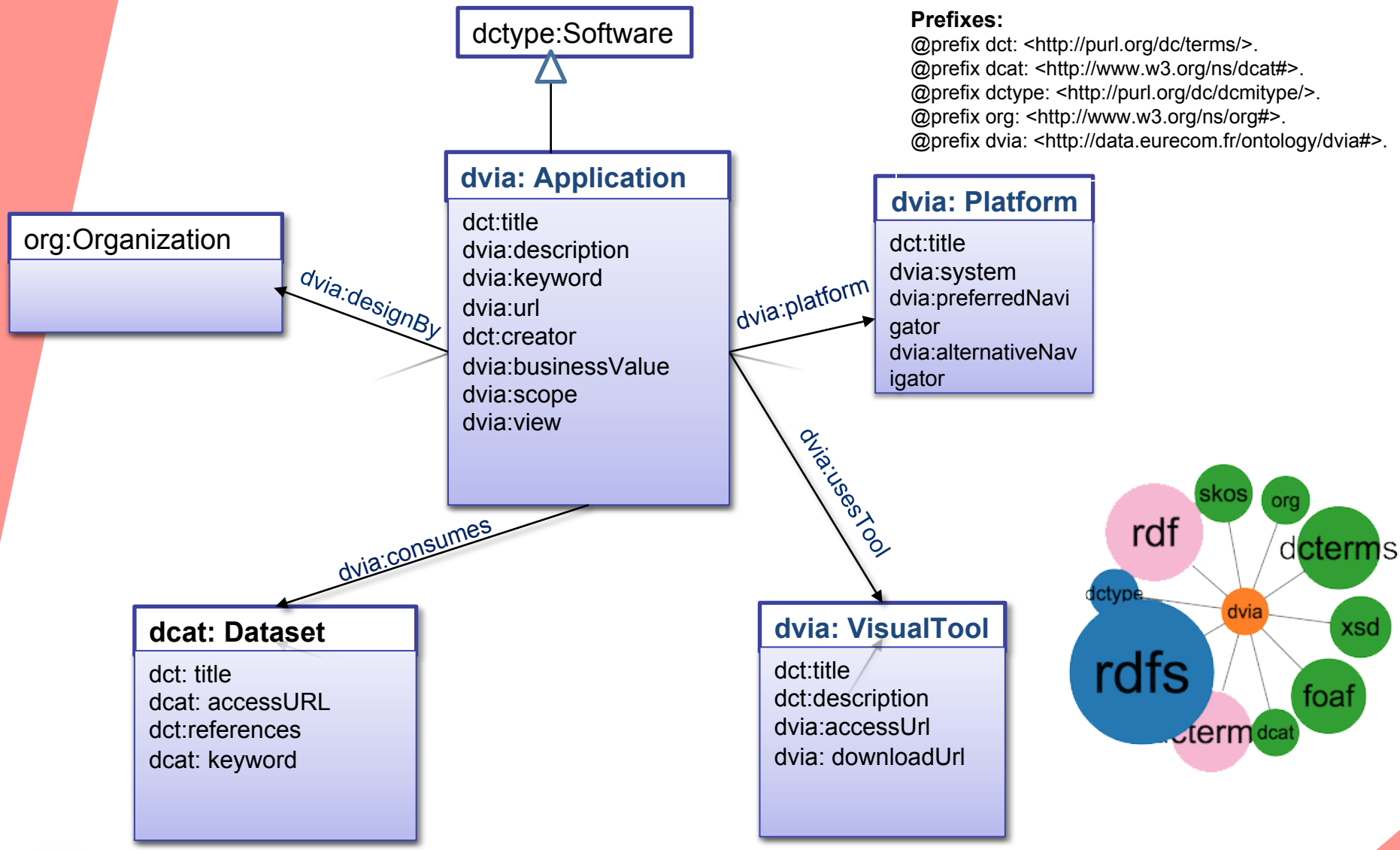
Type of View: Graph, Map views.

Visualization Tools: google visualization API, raphael.js

License: Open Government license [OGL]

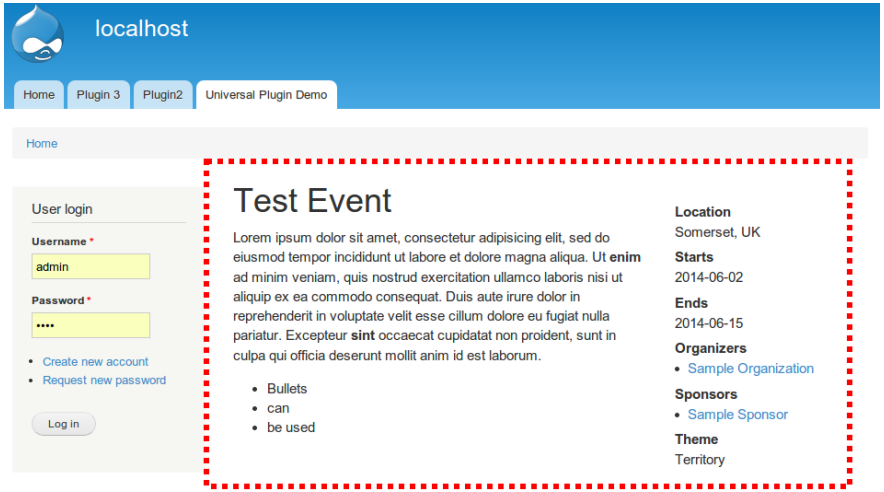
Business Value: Non commercial

DVIA: A vocabulary to describe Applications



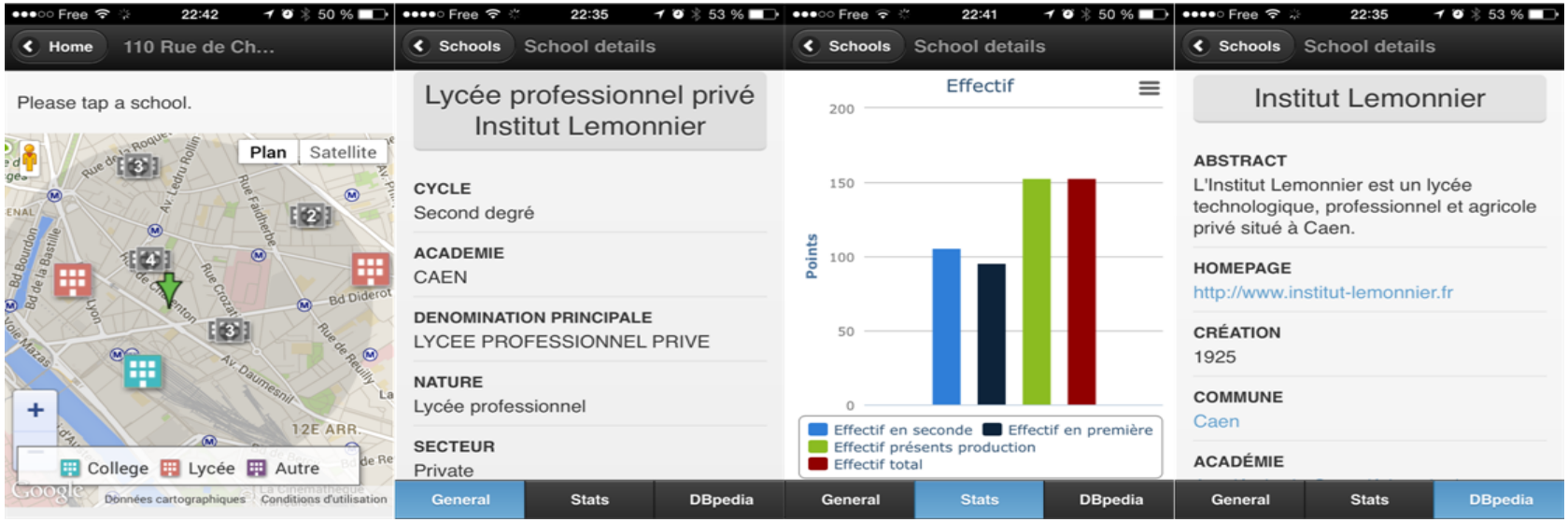
DVIA in Real World Datasets

- **4 applications re-using DVIA**
  - Use to populate 22 past events of hack-athons in Europe, with 889 applications from Apps4Europe.
  - Implementation of a universal JavaScript plugin to embed RDFa in organizers events
  - An extension for Wordpress uses DVIA.



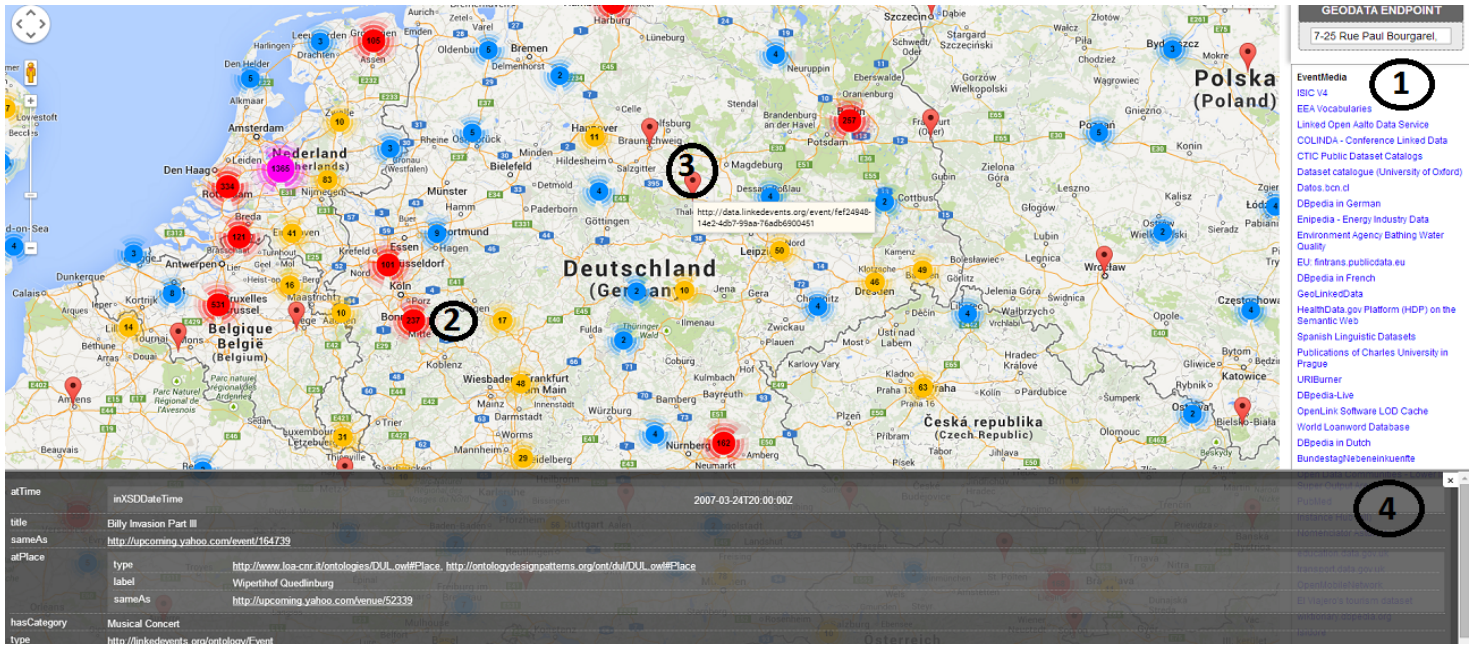
Developing Web Applications: from specific to generic approach

- **Scenario 1**
  - Known Datasets, Known vocabularies → Specific SPARQL queries
  - Visualizations: dataset specific
- **Example**
  - Datasets on schools in France
  - Vocabularies: geo vocab, data cube, geometrie,
  - Application: PerfectSchool



Developing Web Applications: from specific to generic approach

- **Scenario 2**
  - Unknown Datasets, Known domains, so domain-specific SPARQL queries
  - Visualizations: domain specific
- **Example**
  - Endpoints of geodatasets
  - Domain: geospatial
  - Application: GeoRDFviz



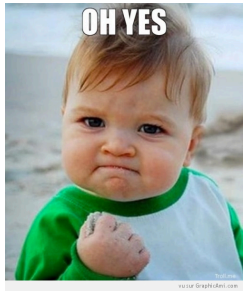
Developing Web Applications: from specific to generic approach

- **Scenario 3**
  - **Unknown** Datasets, **Unknown** domains, so generic SPARQL queries
  - Visualizations: adapted to domains specific
  - Any endpoints
  - Multiple domains: geodata, statistics, persons, cross-domains, etc..
  - Application: ?????

Related work on configuring Semantic Web widgets by data mapping. [1]

Application: Efficient search for Semantic News demonstrator in Cultural Heritage Dataset

Tool: ClioPatria



…but *"method not apply to create interfaces on top of arbitrary SPARLQ endpoints"*

[1] Hildebrand, Michiel, and Jacco Van Ossenbruggen. "Configuring semantic web interfaces by data mapping." *Visual Interfaces to the Social and the Semantic Web (VISSW 2009)* 443 (2009): 96.

Our Proposal

Linked Data Vizualization Wizard (LDVizWiz)

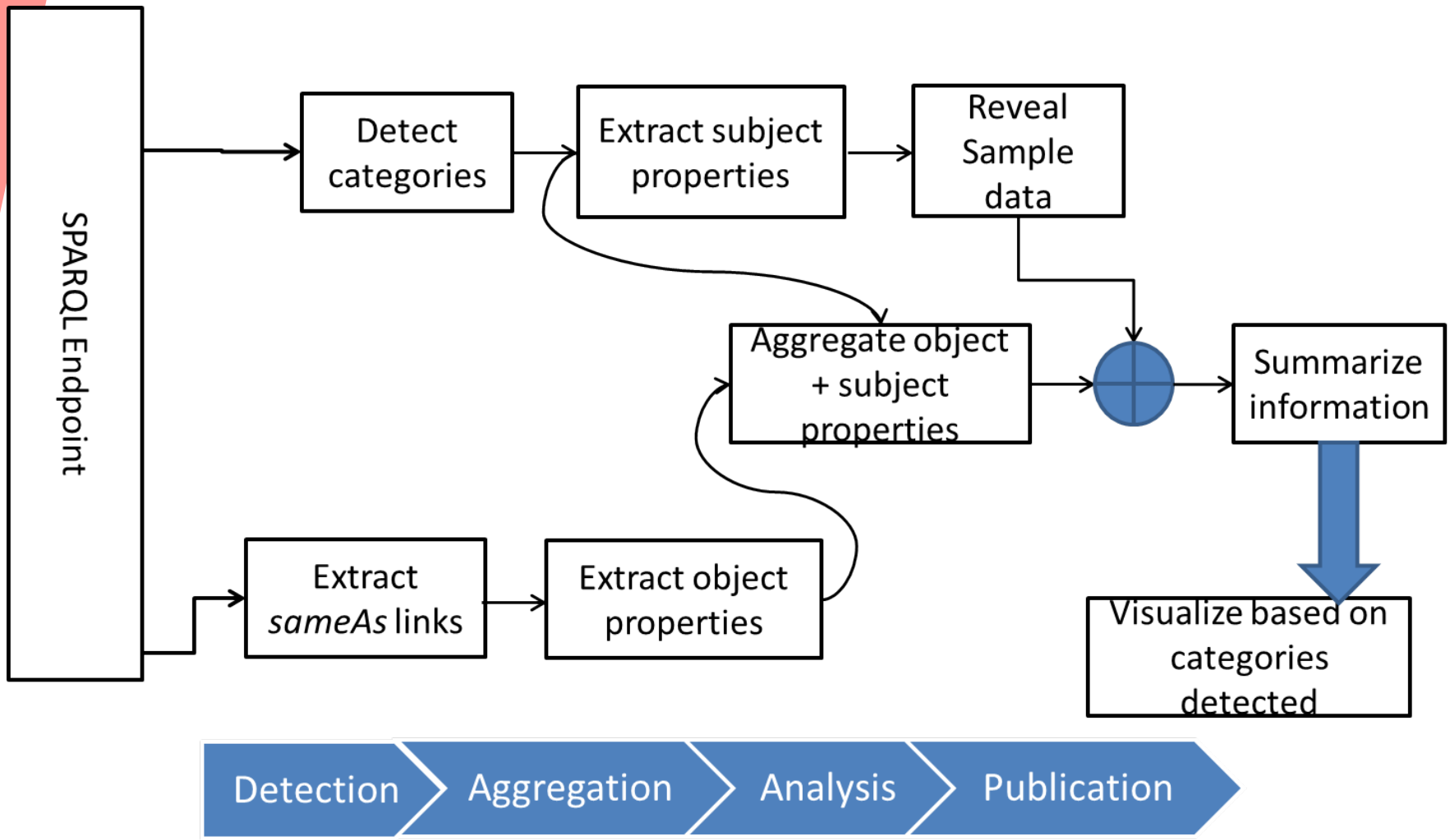
Requirements of LDVizWiz (LDViz-"Wise")

- Predefined **categories** associated to visual elements
- Build on top of RDF standards
  - e.g., SPARQL queries ; Semantic Web technologies
- Reuse existing Visualization libraries
  - e.g., Google Maps, Google Charts, D3.js, etc.
- Reuse On-line Library of Information Visualization (OLIVE)
- Target to non "RDF/SPARQL speakers"
- Input: Datasets published as LOD

Mapping Categories and vocabularies

- **Geographic information**
  - geo: vocab, schema:Place, etc.
- **Temporal information**
  - Time, interval ontologies
- **Event information**
  - lode, event, sport, etc..
- **Agent/Person**
  - foaf:Person/foaf:Agent
- **Organization information**
  - ORG vocabulary, foaf:Organization
- **Statistics information**
  - Data cube, SDMX model
- **Knowledge information**
  - Schemas, classifications using SKOS vocabulary

LDVizWiz Workflow



Step 1: Categories Detection

- **Detection of main categories in datasets**
  - ASK SPARQL queries on predefined categories
  - Uses well-known vocabularies in LOV
  - Condition the type of visual elements

Detection

Dimension	Vocabulary Space	Visual element
Temporal	Time space	TimeLine
one-Dimension	any	Tabular, text
two-Dimension	Geography space	Map view
	Geometry space	Maps view
three-Dimension	Event space	Map + TimeLine
Multi-Dimension	qb, sdmx-model, scovo	Charts, graphs
Tree	skos, Government space	Treemap, Org view
Network	any vocab.	Graph, network map

Experiment: Categories Detection

Category	Number	%
GEO DATA	97	21.84%
EVENT DATA	16	3.60%
TIME DATA	27	6.08%
SKOS DATA	02	0.45%
ORG DATA	48	10.81%
PERSON DATA	59	13.28%
STAT DATA	29	6.6%

- 444 endpoints (*) analyzed, 278 good answers **(62.61%)** using ASK queries.
- Few taxonomies in SKOS, **many GEO DATA**

- **Applications**
  - Automatic detection of endpoints categories
  - More "*trustable*" than human tagging
  - Map categories detected with "suitable" visual elements for the visualizations (e.g., TimeLine + maps for events data)

Detection

(*) All the endpoints retrieved from sparqles.org

Step 2: Properties Aggregation

- **Goal: Exploit the "connectors" between graphs**
- **"connectors" are used to enrich a given graph**
  - e.g., `owl:sameAs ; rdfs:seeAlso, skos:exactMatch`
- **Retrieve properties from external datasets**
  - So called "enriched properties"
- **Build candidate properties for visualization**
  - For pop-up menus
  - For facet browsing
  - For charts display

Step 3: Publication

- **Visualization Generator**
  - Recommend the visual elements based on categories
  - Transform ASK queries to SELECT or CONSTRUCT queries for input to visual library.
- **Visualization Publisher**
  - Export the description of visualization in RDF
  - Add metadata for the visualization (charts) and steps used to create it.
  - e.g., `dcat:Dataset`, `prov:wasDerivedFrom`, `chart` vocabulary, `void:ExampleResource`.

Current Implementation

- Javascript light version as "*proof-of-concept*"
- Url: http://semantics.eurecom.fr/datalift/rdfViz/apps/

Part 3

Best Practices for metadata in vocabularies

W3C Govn't LD Best Practices

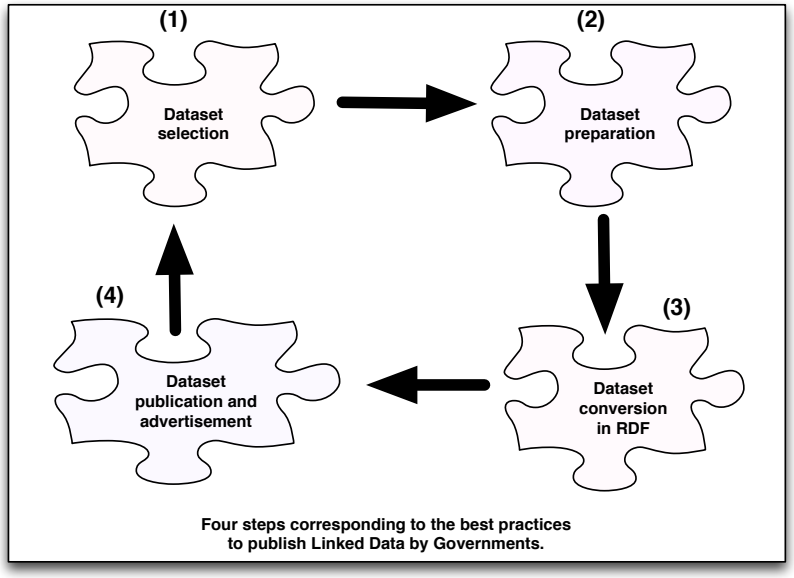
- **10** best practices to help government worldwide to access and reuse their by taking benefit of Linked Data mechanism.
- **4** steps to rich **5★** ratings datasets in TimBL scale.

W3C

Best Practices for Publishing Linked Data

W3C Working Group Note 21 December 2013

This version:
http://www.w3.org/TR/2013/NOTE-ld-bp-20131221/
Latest published version:
http://www.w3.org/TR/ld-bp/
Latest editor's draft:
https://dvcs.w3.org/hg/gld/raw-file/default/bp/index.html
Previous version:
http://www.w3.org/TR/2013/WD-ld-bp-20130625/
Editors:
Bernadette Hyland, 3 Round Stones, Inc.
Ghislain Atemezing, EURECOM
Boris Villazón-Terrazas, iSOCO, Intelligent Software Components S.A.

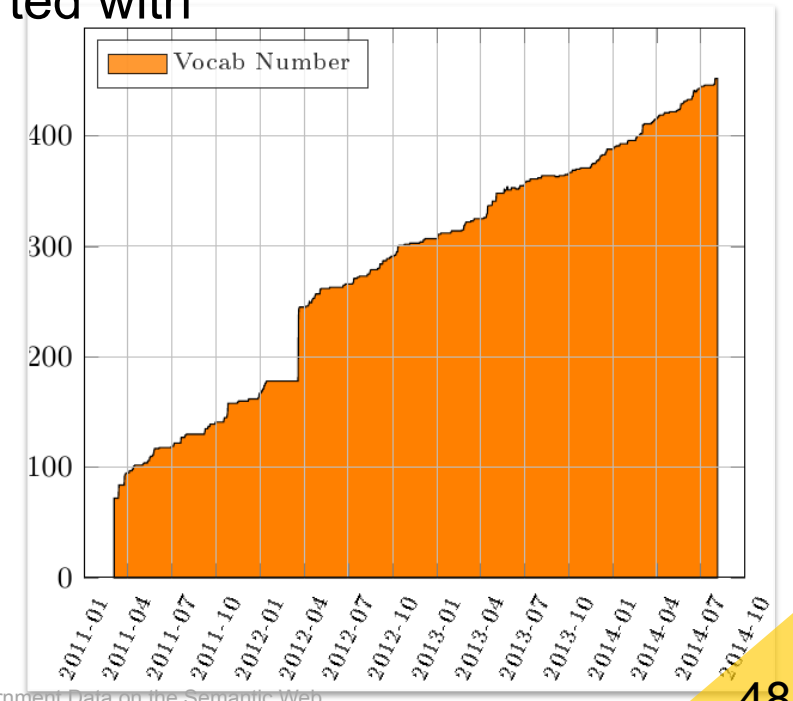


Four steps corresponding to the best practices to publish Linked Data by Governments.

LOV in a Nutshell : http://lov.okfn.org/dataset/lov/

- **A curated list of vocabularies**
  - More than 495 vocabularies
  - Each of them described by vocabulary-of-a-friend **(voaf)**
  - Provide a dump in N3 of the different versions of a vocabulary
  - Quasi linearity of the growth, started with 75 vocabularies

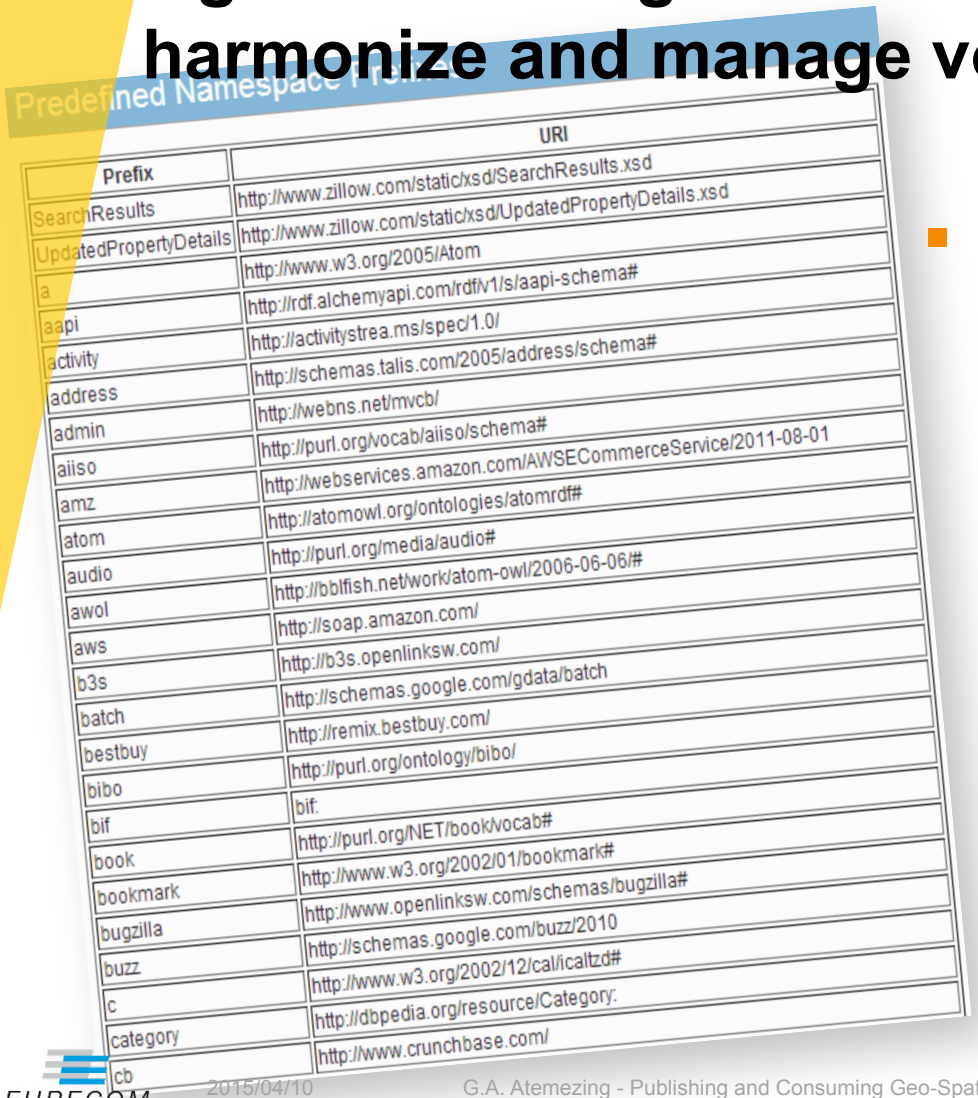
The glitch in 2012 corresponds to the migration to OKFN hosting



Focus on vocabularies: Disambiguating Vocabulary Prefixes

Goal:

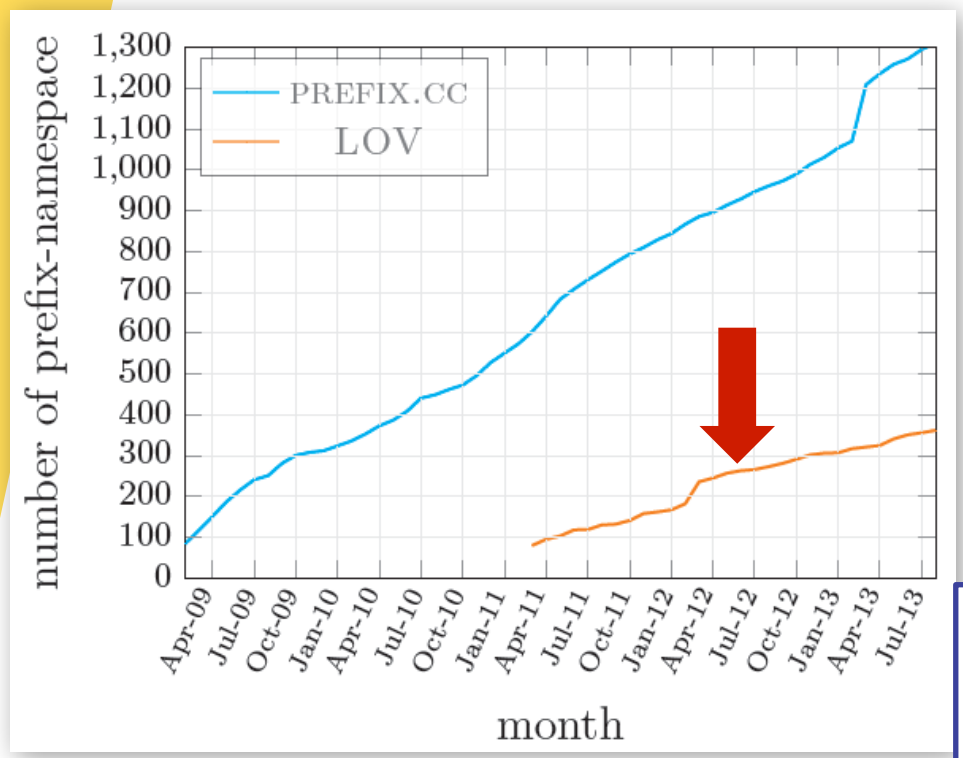
align services against Linked Open Vocabularies to harmonize and manage vocabularies' namespaces



Prefix	URI
SearchResults	http://www.zillow.com/static/xsd/SearchResults.xsd
UpdatedPropertyDetails	http://www.zillow.com/static/xsd/UpdatedPropertyDetails.xsd
a	http://www.w3.org/2005/Atom
aapi	http://rdf.alchemyapi.com/rdf/v1/s/aapi-schema#
activity	http://activitystrea.ms/spec/1.0/
address	http://schemas.talis.com/2005/address/schema#
admin	http://webns.net/mvcb/
aiiso	http://purl.org/vocab/aiiso/schema#
amz	http://webservices.amazon.com/AWSECommerceService/2011-08-01
atom	http://atomowl.org/ontologies/atomrdf#
audio	http://purl.org/media/audio#
awol	http://bblfish.net/work/atom-owl/2006-06-06/#
aws	http://soap.amazon.com/
b3s	http://b3s.openlinksw.com/
batch	http://schemas.google.com/gdata/batch
bestbuy	http://remix.bestbuy.com/
bibo	http://purl.org/ontology/bibo/
bif	bif:
book	http://purl.org/NET/book/vocab#
bookmark	http://www.w3.org/2002/01/bookmark#
bugzilla	http://www.openlinksw.com/schemas/bugzilla#
buzz	http://schemas.google.com/buzz/2010
c	http://www.w3.org/2002/12/cal/icaltzd#
category	http://dbpedia.org/resource/Category:
cb	http://www.crunchbase.com/

- **Global namespaces**
  - With good practices to recommend a prefix
  - Have a more transparent list of built-in prefixes
  - All the services understand each other with prefixes
  - Some de facto prefixes emerging: rdfs:, foaf:, rdf:, owl:, skos:,

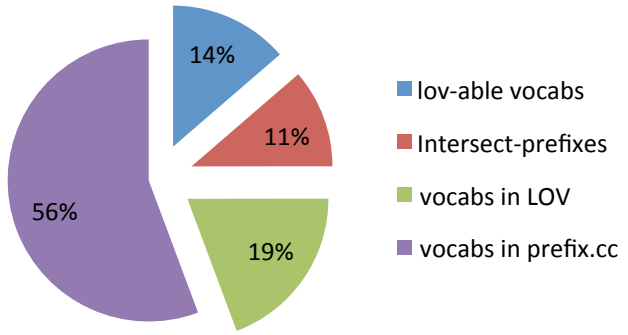
LOV vs PREFIX.CC-Alignment Findings



Category	Number
lov-able vocabs	227
Intersect-prefixes	188
vocabs in LOV	321
vocabs in prefix.cc	925

More than **200** prefixes in prefix.cc are vocabularies

Findings during alignment process



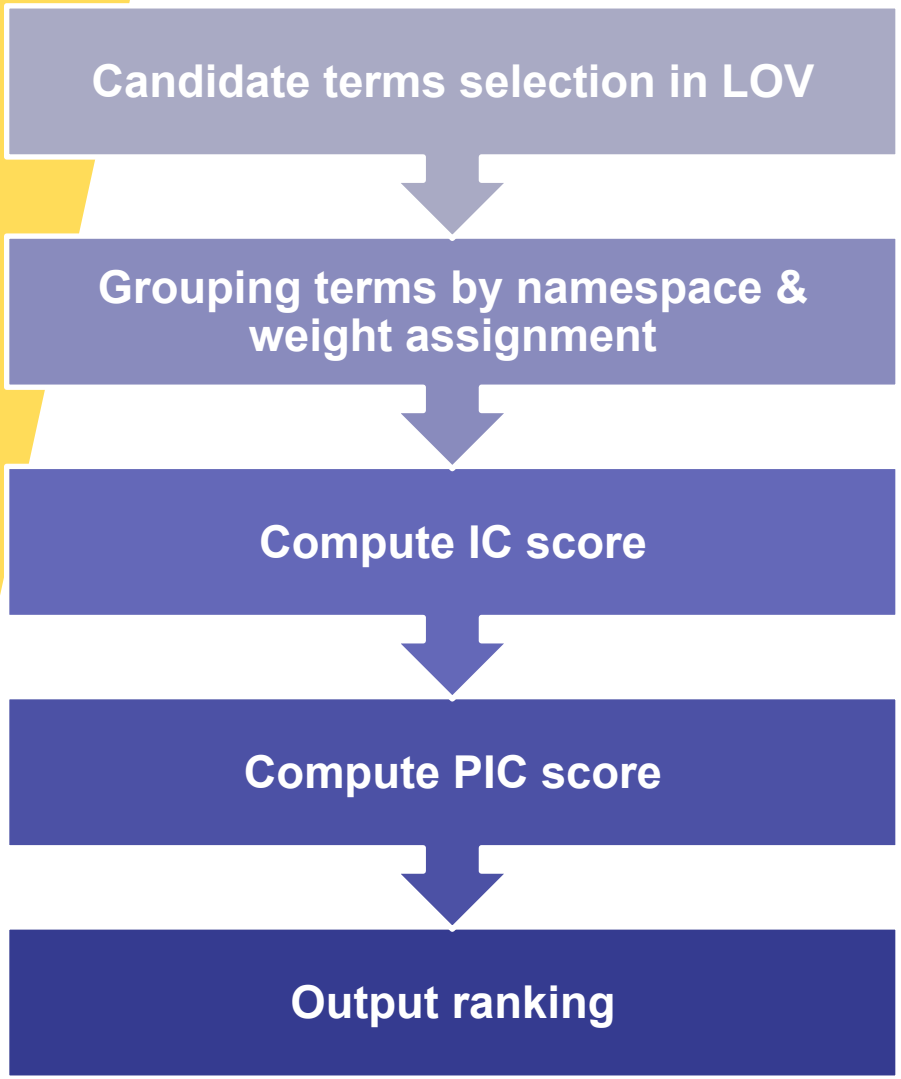
Vocabulary Search and Ranking

Goal

Ranking vocabularies based on Information Content Metrics

- **Metrics**
  - Information Content Metric (IC): **value of information** associated with a given entity
  - Partition Information Content Metric (PIC)
  - Proposed a ranking based on IC and PIC

Ranking Algorithm



Require: Dump of *lovaggregator* file
1: Upload in a triple store for querying
2: Select subset of candidate vocabs *LOVaggregatorendpoint* (1)
3: **for** *term* ∈ *lovaggregator* **do**
4: **if** (*LOVdistribution* ≥ 1)**and** (*LOVPopularity* ≥ 1) **then**
5: *candidateterms* ← append *term*
6: **end if**
7: **end for**
8: **for** each *term* ∈ *candidateterms* **do** (2)
9: GROUP BY vocabulary namespace
10: COMPUTE weight for each vocabulary
11: **end for**
12: INITIALIZE *PICvector* AS a vector
13: **for** each *term* ∈ *candidateterms* **do**
14: **while** *term* ∈ *vocabularySpace* **do**
15: *ICterm* ← function IC(term, vocabPrefix) (3)
16: *ICvocab* ← $\sum$ *ICterm*
17: **end while**
18: *PICvocab* ← *weight*(*vocab*)×ICvocab
19: *PICvector* ← append (PICvocab) (4)
20: ORDER PICvector
21: **end for**
22: **return** PICvector

Ranking Algorithm

- dcterms: http://purl.org/dc/terms/
- Candidate terms: 53 (39 properties + 14 classes)
- wf = 1+ 2+3 = 6
- PIC = 1724.844

Publisher	Dublin Core Metadata Initiative
Class number	22
Property number	55
Homepage	http://dublincore.org/documents/dcmi-terms
See also	http://stats.lod2.eu/vocabularies/9
Represented by	format-dcterms
Has review	(2013-03-07) Bernard Vatant: Prefix restore (2014-03-14) Bernard Vatant: This vocabula purl redirection is sometimes down, like at

- foaf: http://xmlns.com/foaf/0.1/
- Candidate terms: 35 (26 properties + 9 classes)
- wf = 1+ 2+ 3 = 6
- PIC = 1033.197

Property number	62
Homepage	http://www.foaf-project.org/
See also	http://stats.lod2.eu/vocabularies/90
Represented by	format-foaf
Has review	(2013-06-04) Bernard Vatant: From the specificati 2000. There is now a stable core of classes and their documentation to track implementation feed (2014-01-15) Bernard Vatant: Looking forward for

PIC(dcterms) > PIC(foaf)

Ranking Algorithm

- **Top-15 terms (IC value)**

Rank	vocab term	IC value
1	skos:example	7.7806
2	dce:contributor	4.674
3	skos:scopeNote	4.365
4	dcterms:source	4.299
5	mads:code	3.922
6	mads:authoritativeLabel	3.922
7	vs:userdocs	3.847
8	dce:title	3.79
9	skos:hasTopConcept	3.4547
10	dce:description	2.758
11	dcterms:issued	2.553
12	dce:creator	2.518
13	skos:inScheme	2.202
14	skos:notation	1.924
15	dcterms:description	1.646

- **Top-15 vocabs (PIC value)**

Rank	Prefix	PIC score
1	dcterms	1724.844
2	schema	1588.700
3	gr	1261.101
4	foaf	1033.197
5	bibo	876.205
6	time	816.2020
7	skos	805.287
8	dul	797.328
9	ptop	773.167
10	rdafrbr	640.834
11	vaem	630.621
12	ma-ont	508,694
13	prov	497.524
14	swrc	437.394
15	dce	428.618

Applications of the Ranking Metrics

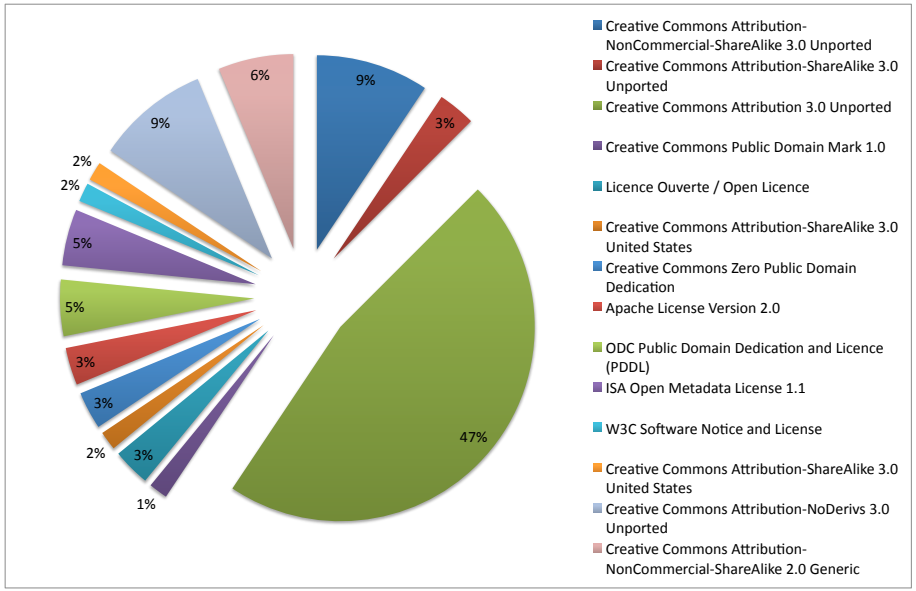
- **Vocabulary life-cycle management**
  - Help assessing the use of terms and vocabulary updates
  - Monitoring the use of *owl:deprecated or* http://www.w3.org/2003/06/sw-vocab-status/ns#:term_status
- **Semantic Web applications**
  - Vocabularies with **higher PIC** might be proposed to a user as much as possible, e.g. for choosing properties to display in a facetted browsing interface
- **Interlinking datasets**
  - Generate *sameAs* links between resources based on vocabularies terms with **lower IC** value

Licenses Compatibility

Goal

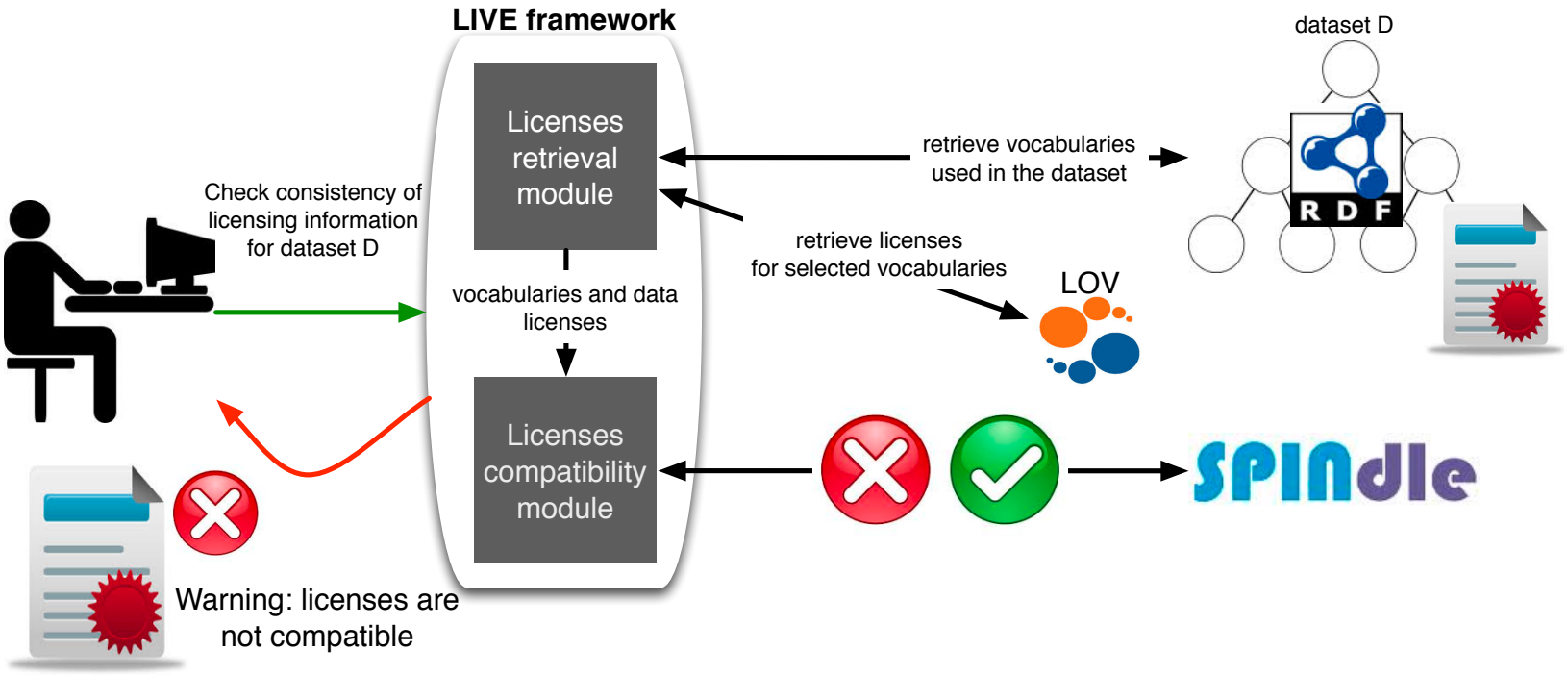
Reasoning on Licenses for checking compatibilities between vocabularies and datasets

- **Approach:**
  - Licenses retrieval both for the dataset and for the vocabularies used in the dataset.
  - Deontic logic [1] to compute compatibility using RDF representation of licenses



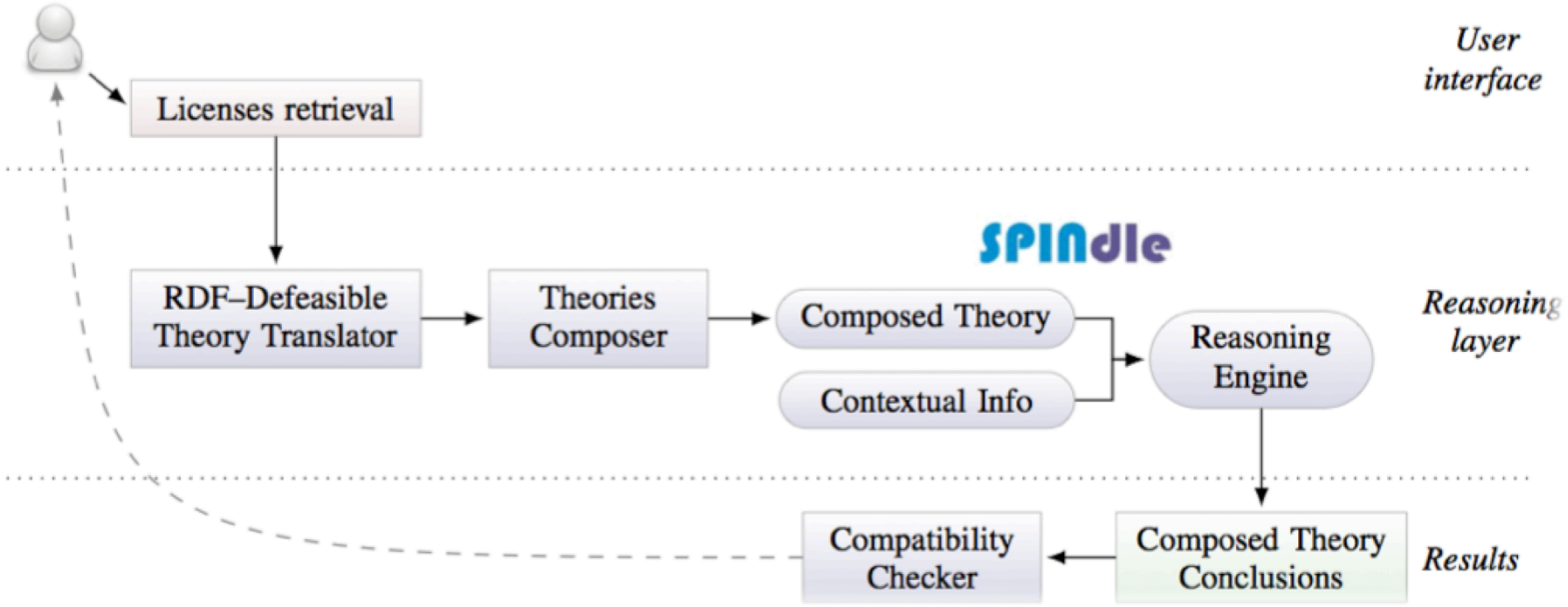
[1] Governatori, Guido, et al. *"One License to Compose Them All."* The Semantic Web–ISWC 2013. Springer Berlin Heidelberg, 2013. 151-166.

Our Proposal: LIVE Framework



http://www.eurecom.fr/~atemezin/licenseChecker/

The Logic



- **RDF licenses: http://purl.org/NET/rdflicense**
- **Logic of deontic rules:**
  - constructive account of basic deontic modalities (obligation, prohibition, permission)
  - compute the set of all conclusions for each license and then check whether incompatible conclusions are obtained.

Live Evaluation

Dataset	LicRe-trieval(ms)	#vocabu-laries	LicCompa-tibility(ms)	LIVE(ms)
rkb-explorer-dblp	4,499	1	0	4,499
rkb-explorer-southampton	14,693	1	0	14,693
rkb-explorer-eprints	3,220	1	0	3,220
rkb-explorer-acm	3,007	1	0	3,007
rkb-explorer-wiki	14,598	1	0	14,598
rkb-explorer-rae2001	3,343	1	0	3,343
rkb-explorer-citeseer	2,760	1	0	2,760
rkb-explorer-newcastle	3,354	1	0	3,354
rkb-explorer-kisti	4,094	5	6	4,100
270a.info	13,202	48	8	13,210

LIVE provides the compatibility in less than **5 seconds** for 7 datasets

LIVE retrieves 48 vocabularies in less than **14 seconds**

Conclusions

- **We have presented models for Geodata**
  - For better handling complex geometries
  - Supporting ***all*** CRS
  - For easy querying in SPARQL
- **We have proposed a generic visualization wizard**
  - Based on predefined categories
  - Targeted to lay-users
- **We have contributed in managing vocabulary metadata**
  - Disambiguating prefixes between vocabularies
  - Improving vocabulary search and ranking
  - Providing a license compatibility framework
  - Contributing to Best Practices standard

Future Work

- **Managing Data updates and versioning**
  - Versioning: integrate Memento protocol?
  - Spatio-temporal evolution
- **Multiple representation: need for metadata?**
  - Level of detail
  - Geometry modeling rules and reasoning
- **Tracking Provenance of geodata**
  - To ensure quality of published dataset
  - To ensure trust from application consumers

Future Work

- **Visualizations**
  - Extend categories and vocabularies for detection
  - Provide templates for generating “mash-ups” to combine domains, an mash-up widget generator
  - Investigate the “***importance***” of a category in dataset
  - Provide a user evaluation

- **Metadata management**
  - Publish a list of common recommended prefixes
  - Foster and support current effort towards a more sustainable governance of vocabularies.
  - Compare (P)IC with other graph-based ranking (e.g. pagerank)
  - Investigate the dependency ranking between vocabularies

Publications

- Pierre-Yves Vandenbussche, **G.A. Atemezing**, Maria Poveda, Bernard Vatant: Linked Open Vocabularies (LOV): a gateway to reusable semantic vocabularies on the Web. Semantic Web Journal, *under review, 2015*.
- **G.A. Atemezing**, Raphael Troncy: Modeling visualization tools and applications on the Web. Semantic Web Journal, *under review, 2015*.
- **G.A. Atemezing** et al.: Transforming meteorological data into linked data. In Semantic Web journal, Special Issue on Linked Dataset descriptions, 2012. IOS Press.

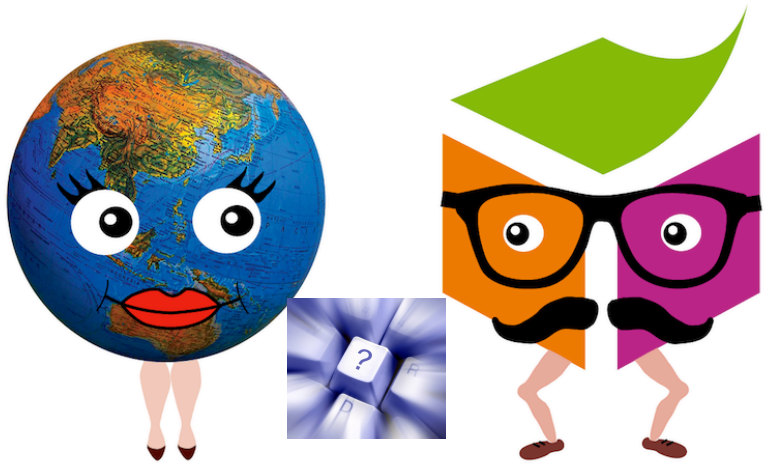
- Guido Governatori, Ho-Pun Lam, Antonino Rotolo, Serena Villata, **G.A. Atemezing** and Fabien Gandon: LIVE: a Tool for Checking Licenses Compatibility between Vocabularies and Data.(ISWC 2014, Demo Track)
- **G.A. Atemezing** and Raphael Troncy: Information content based ranking metric for linked open vocabularies. (SEMANTICS 2014)
- Ahmad Assaf, **G.A. Atemezing**, Raphael Troncy and Elena Cabrio: What are the important properties of an entity? Comparing users and knowledge graph point of view. In 11th Extended Semantic Web Conference (Demo Track, ESWC 2014)

Publications

- Francois Scharffe, **G.A Atemezing**, Raphael Troncy, Fabien Gandon, Serena Villata, Bénédicte Bucher, Faycal Hamdi, Laurent Bihanic, Gabriel Kepeklian, Franck Cotton, Jerome Euzenat, Zhengjie Fan, Pierre-Yves Vandenbussche and Bernard Vatant: Enabling linked-data publication with the datalift platform. (AAAI, W10:Semantic Cities, 2012)
- **G.A. Atemezing** and Raphael Troncy: Vers une meilleure interopérabilité des donneés geographiques francaises sur le Web de donneés. (IC 2012).
- Houda Khrouf, **G.A. Atemezing**, Thomas Steiner, Giuseppe Rizzo and Raphael Troncy: Confomaton: A conference enhancer with social media from the cloud. (ESWC 2012, Demo Track)

- Bernadette Hyland, **G.A. Atemezing** and Boris Villazón-Terrazas (editors): Best Practices for Publishing Linked Data. W3C Working Group Note pub- lished on January 9, 2014. URL: http://www.w3.org/TR/ld-bp/
- Bernadette Hyland, **G.A Atemezing**, Michael Pendleton, Biplav Srivastava (editors): Linked Data Glossary. W3C Working Group Note published on June 27, 2013. URL: www.w3.org/TR/ld-glossary/

Thank you for your attention!



The stars of the show: Miss Globe and Mr Cube, created by Frans Knibbe of Geodan

Credits layout
Mariella Sabatino: mll.sabatino@gmail.com