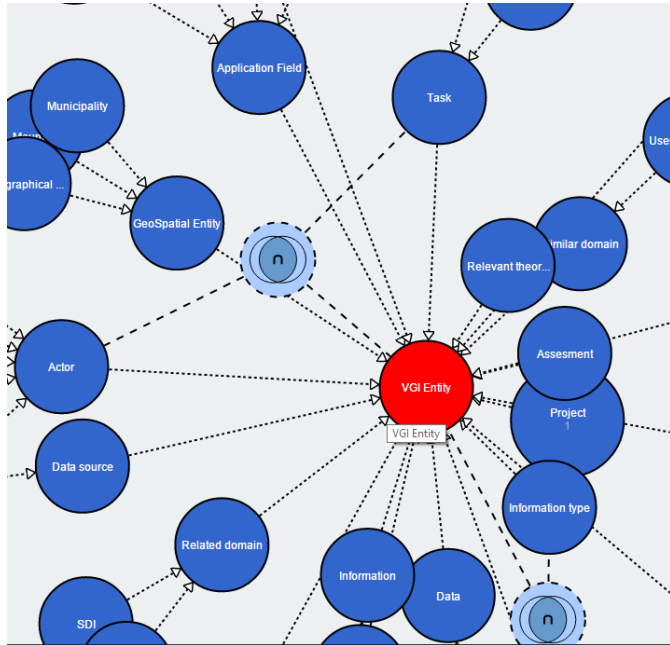


# AGILE Link-VGI workshop, Helsinki 14 June 2016



## Towards Linked Data and ontology development for the semantic enrichment of volunteered geo-information

Rob Lemmens

Gilles Falquet,  
Claudine Métral

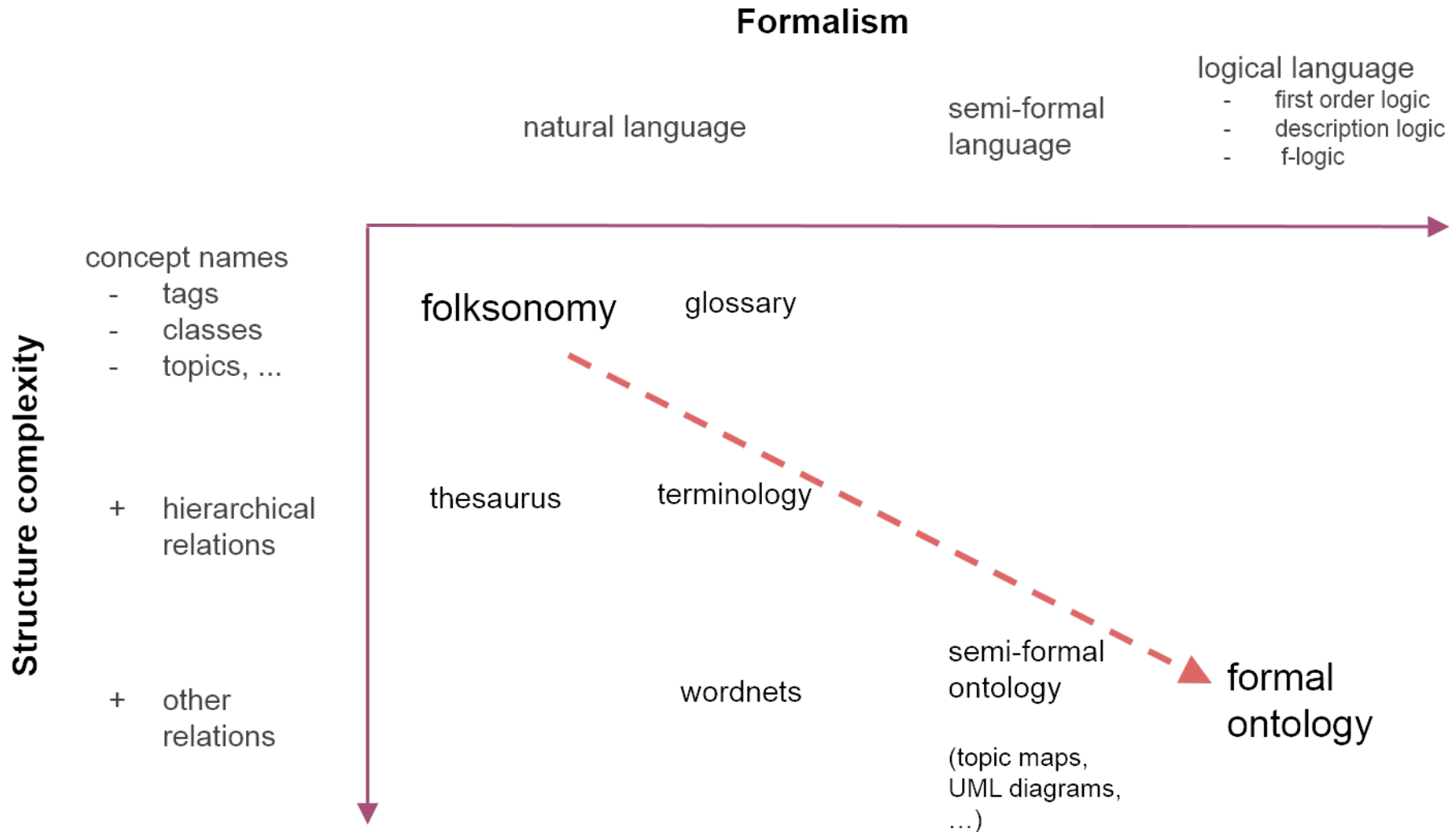
University of Twente,  
Faculty of Geo-Information  
Science and Earth  
Observation (ITC)  
Enschede, The Netherlands  
r.l.g.lemmens@utwente.nl

University of Geneva,  
Faculty of economics and  
social sciences,  
Geneva, Switzerland  
{Gilles.Falquet;  
claudine.metral}@unige.ch

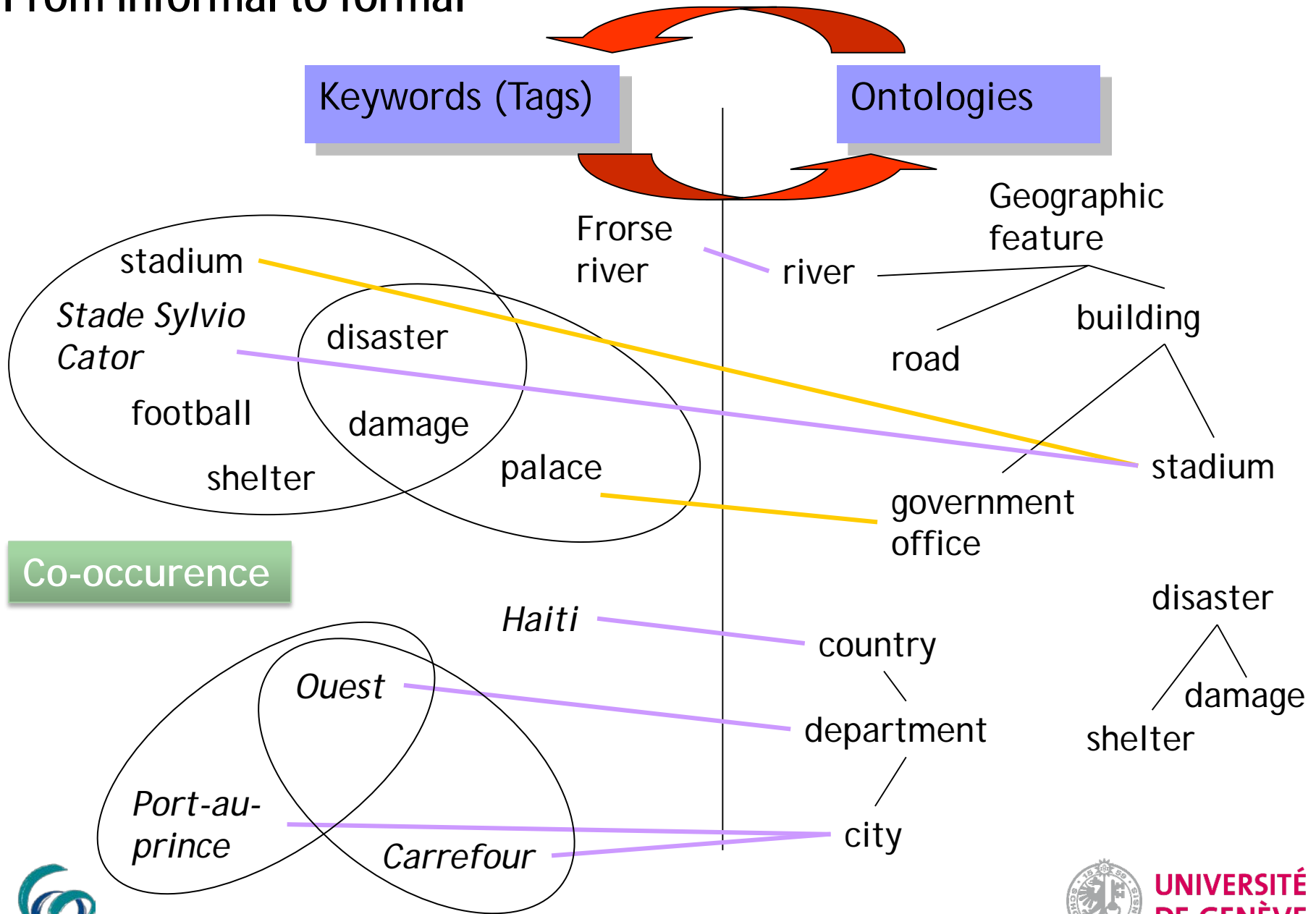
# ENERGIC Project – Objectives - <http://vgibox.eu/>

- The ENERGIC project (European Network Exploring Research into Geospatial Information Crowdsourcing: software and methodologies for harnessing geographic information from the crowd; COST Action IC1203)
- Finding new solutions for the exploitation, integration and application of user generated geo-information.
- Working groups:
  1. Societal and human aspects of VGI,
  2. Spatial data quality and infrastructures and
  3. Data mining, semantics and VGI.
- Links to COST Action TD1202 Mapping and the citizen sensor
- Deliverable: Open and updatable repository of VGI analysis and integration tools and methods, literature and case studies.  
**Ontology serves as semantic backbone.**

# Informal and formal semantic reference spaces



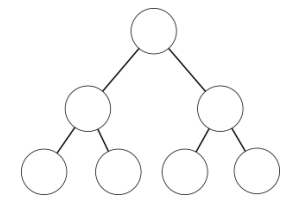
# From informal to formal



Co-occurrence



# Semantic enrichment with ENERGIc VGI ontology



## Classes

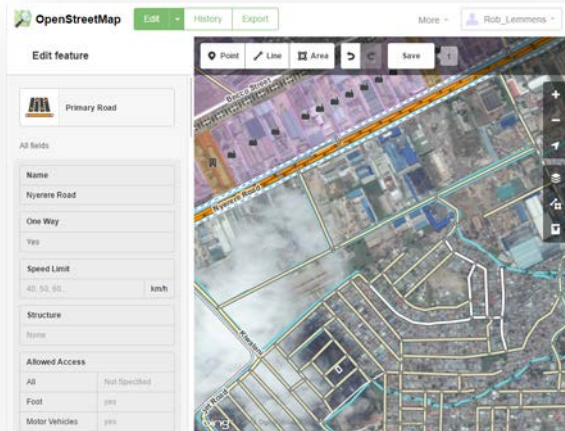
[Accuracy](#) [Active Sensing Task](#) [Activity](#) [Actor](#) [Algorithm](#) [Analysis](#) [Analyst](#) [Annotations](#) [Application](#)  
[Application Field](#) [Article](#) [Assesment](#) [Automated Tagging](#) [Book](#) [Book Chapter](#) [Characterization](#) [Claim](#)  
[Collaborative Human-Computer Analytical Activity](#) [Comparison](#) [Contributor](#) [Daily Activity](#) [Data](#) [Data acquisition](#)  
[Data Analysis](#) [Data description](#) [Data Mining Algorithm](#) [Data quality](#) [Data source](#) [Data type](#) [Dataset Analysis](#)  
[Decision or Policy Making](#) [Definition](#) [Description](#) [Discourse Element](#) [Education](#) [Emergency management](#)  
[End user task](#) [Example](#) [Experimentation](#) [Folksonomy](#) [Geographic Feature Description](#) [Geographic Information](#)  
[Geographical Accuracy](#) [Geographical Object](#) [Geographical precision](#) [Geometry Acquisition](#) [Georef Image](#)  
[Georef Scalar](#) [Georef Structured](#) [Georef Text](#) [GeoreferencedData](#) [GeoSpatial Entity](#) [GIS](#) [Government Agency](#)  
[Hardware](#) [Human Activity](#) [Human sensor](#) [Hypothesis](#) [In-situ sensor](#) [Individual Actor](#) [Information](#)  
[Information type](#) [Institution or Agency](#) [Knowledge Resource](#) [Machine Learning](#) [Mapping](#) [Mapping Activity](#)  
[Measurement Accuracy](#) [Method](#) [Methodology](#) [Mobile application](#) [Monitoring](#) [Mountain](#) [Municipality](#)  
[National Mapping Agency](#) [Navigation](#) [NLP](#) [NonGeoreferenced Data](#) [Ontology](#) [Passive Task](#)  
[Pedestrian Navigation](#) [Positional Accuracy](#) [Precision](#) [Processing](#) [Project](#) [Querying](#)

VGI system: OpenStreetMap



Is of Type

VGI publication



A conceptual model for quality assessment of VGI for the purpose of flood management

Livia Castro Degrossi  
University of São Paulo  
São Carlos, Brazil  
Heidelberg University  
Heidelberg, Germany  
degrossi@icmc.usp.br

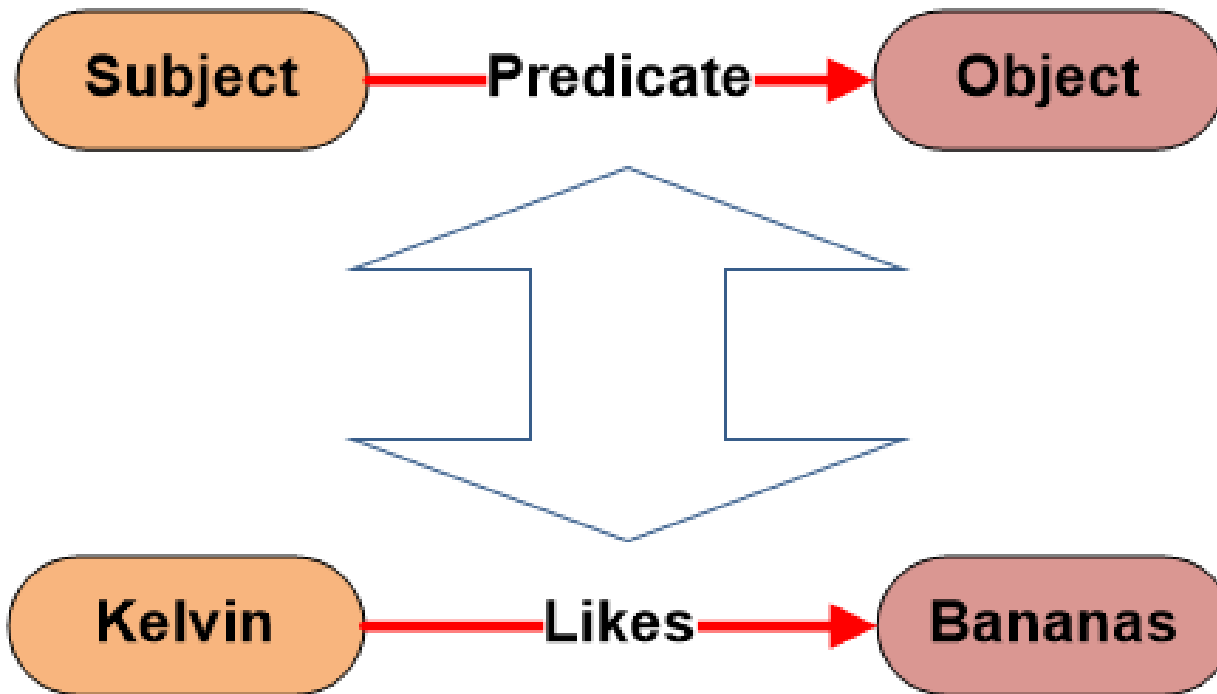
João Porto de Albuquerque  
University of Warwick  
Warwick, England  
GIScience Chair, Heidelberg University  
Heidelberg, Germany  
j.porto@warwick.ac.uk

Hongchao Fan, Alexander Zipf  
GIScience Chair, Heidelberg  
University  
Heidelberg, Germany  
hongchao.fan@uni-heidelberg.de,  
zipf@uni-heidelberg.de

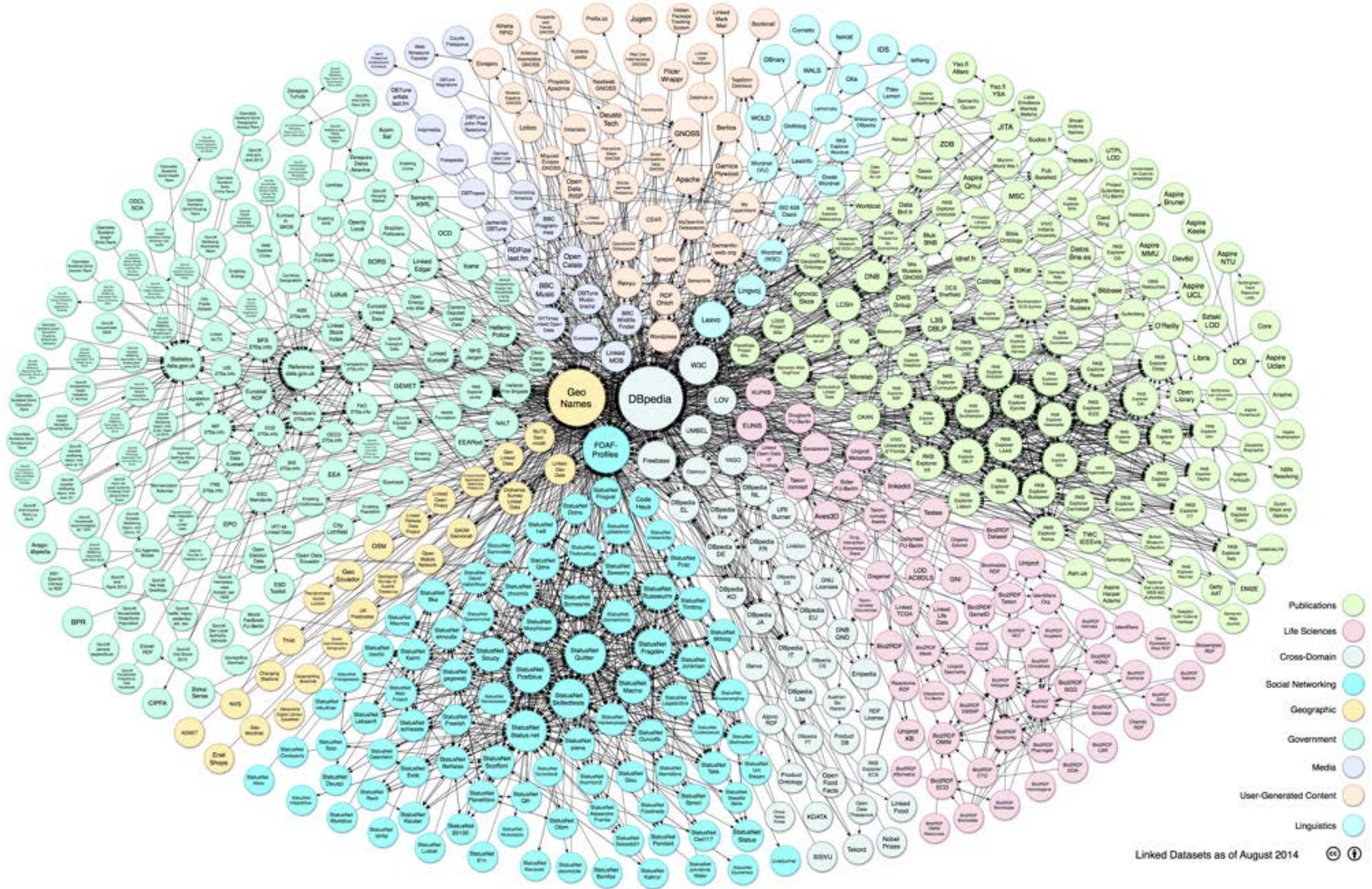
### Abstract

Volunteered Geographic Information (VGI) has emerged as a potential source of geographic information for different domains. Despite the many advantages associated with it, such information lacks of quality assurance, since it is provided by individuals with different motivations and backgrounds. In response to this, several methods have been proposed to assess the quality of volunteered geographic information of different platforms. However, there has been little investigation aimed at explaining how cross-platform data could be used for quality

# Linked Data – RDF triples

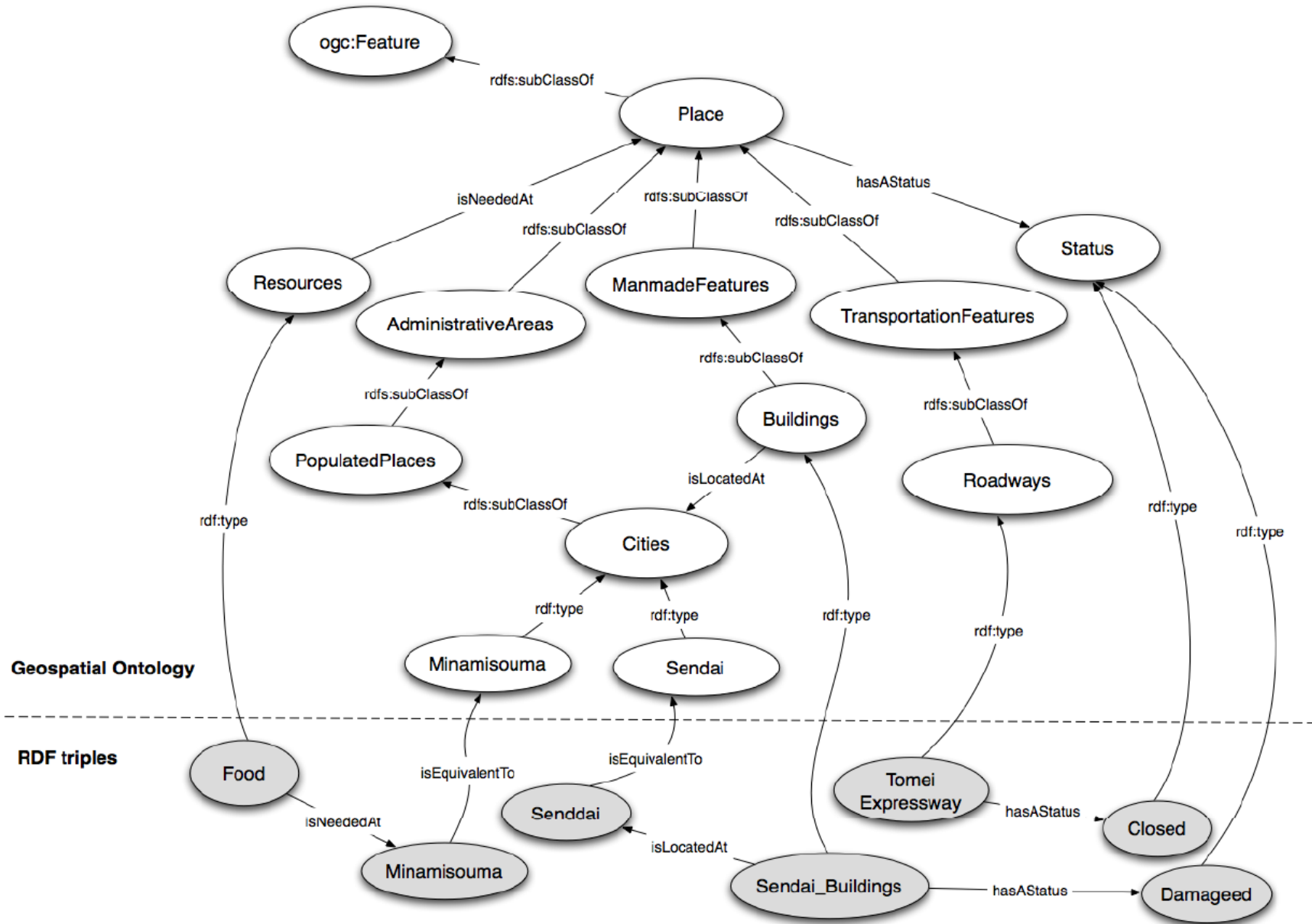


# Linked Open Data cloud



Linked Datasets as of August 2014

# Ontologies & Linked Data



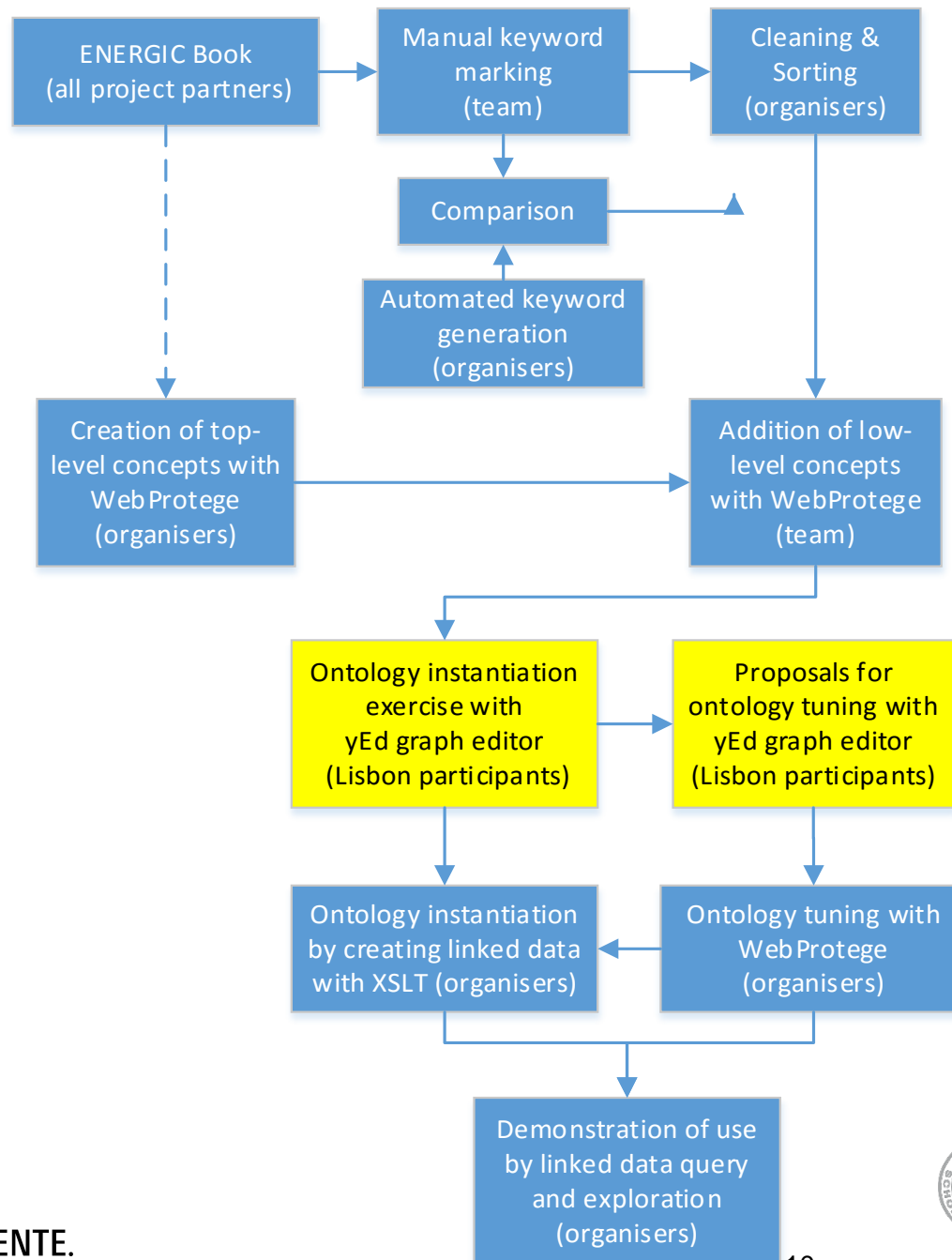


# ENERGIC VGI Ontology - Objectives

1. Common understanding of VGI concepts (e.g., for research, outreach and education) – delimiting the field
  - broad coverage, human-readable definitions
2. Create applications and database schema for the development of new VGI applications
  - centered on VGI data structures
3. Usage of ontology of tasks for evaluating of data quality
  - centered on tasks
4. Semantic enrichment of VGI-based systems and indexing literature for VGI source discovery and integration
  - rich set of properties to describe resources
5. Create relationships between VGI and other domains, such as GIS, SDI, etc.

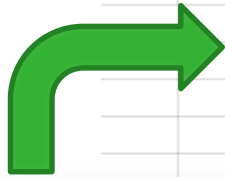
# ENERGIC VGI Ontology

## Workflow



# ENERGIC VGI Ontology Cleaned keyword list

				communicating
				communication
			giscience	community
		urban	walking	community
				community-oriented
	positional	accuracy	and	completeness
			data	completeness
				completeness
				completeness
				complexity
inherent	spatial	and	temporal	component
		collaborative	cloud	computers
			cloud	computing
				computing
			shared	concepts
				conceptualization
	temporally	accurate	road	conceptualizations
				conditions
				confidence
				confidentiality
				connection
				connections
				connections
			d-sourcing	conrmation
				consciously
				consensusproducing
				consistency
			logical	consistency
				consistency



various environments, which might not necessary be caused by or towards these environments. In the following, we extend existing research, and illustrate how social media data can be harnessed to extract people's affective responses to environments. Particularly, we focus on **geotagged photos** in Flickr.

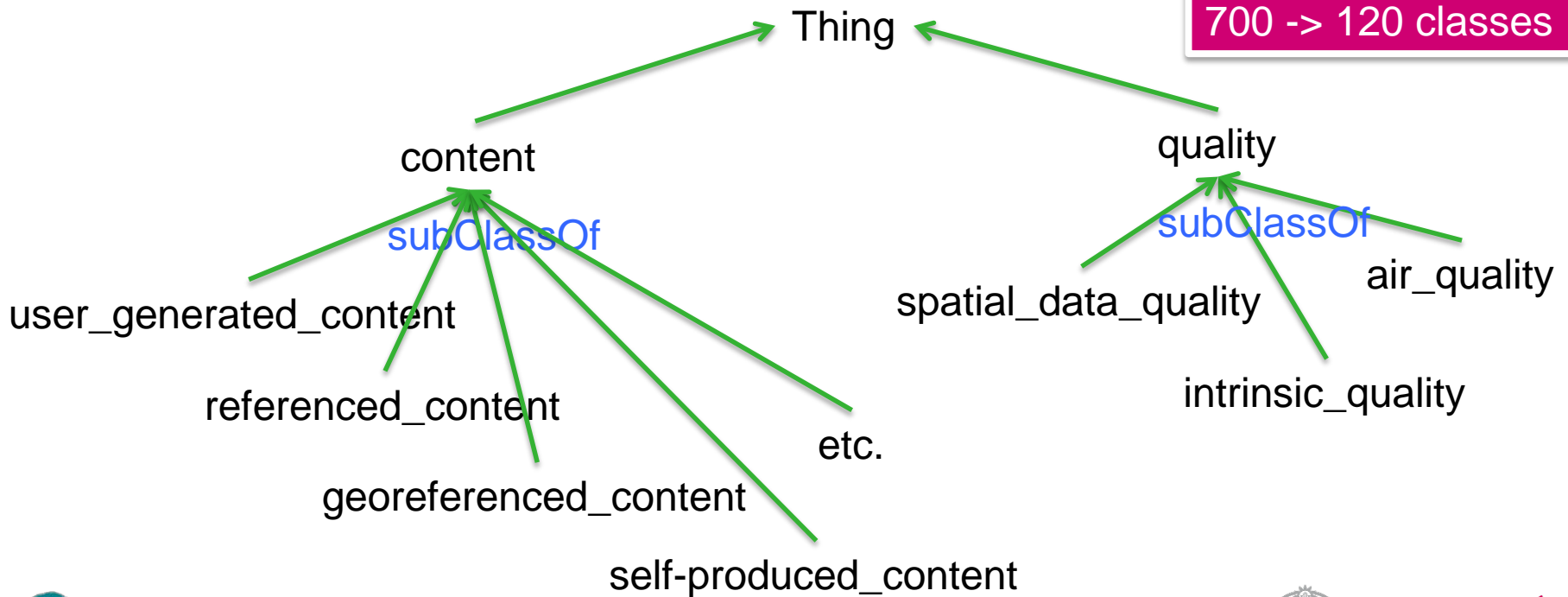
For extracting affective responses from social media data, we apply **sentiment analysis** technique. Sentiment analysis (or **opinion mining**) is a **natural language processing (NLP)** technique, and aims to determine an author's attitudes, opinions or sentiments with respect to the topic written about. Different methods have been proposed for sentiment analysis, among which lexicon-based method is one of the most popular ones. Lexicon-based sentiment analysis employs NLP techniques to tokenize,

# From key terms to organized concepts

- remove single-word terms (e.g. ajax)
- normalize the terms (replace "of", "ing" forms)
- remove n-word terms when the last word appears only once
- group the n-word terms under their last word

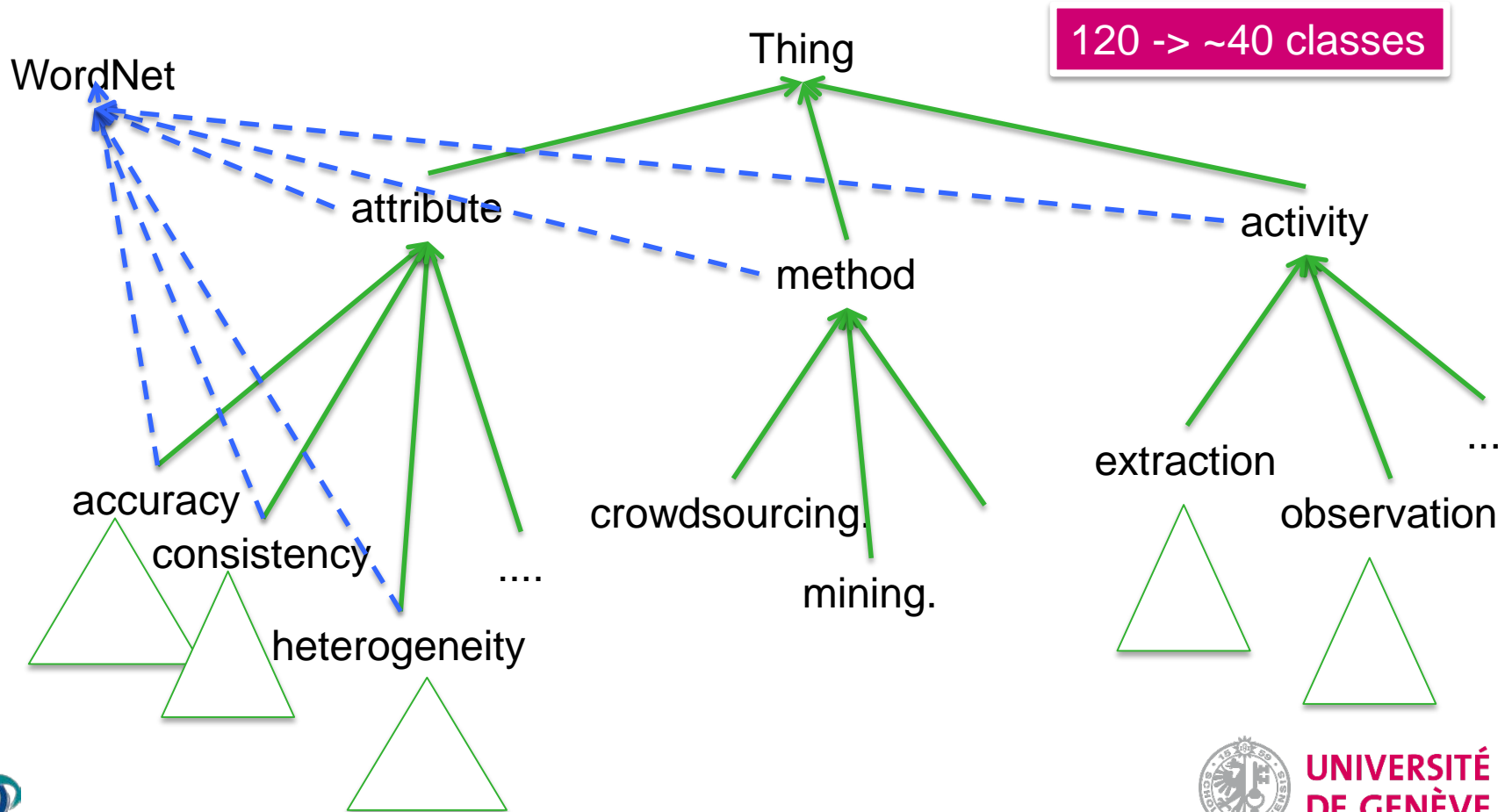
2000 -> 700 classes

700 -> 120 classes



# From key terms to organized concepts

- align with WordNet concepts
- add common superclasses from WordNet



# ENERGIC VGI Ontology Collaborative editing with WebProtege

The screenshot displays the WebProtege web interface for editing the ENERGIC VGI Ontology. The top navigation bar includes 'Project', 'Share', 'Rob Lemmens', and 'Help'. Below this, the 'Classes' tab is active, showing a tree view of the ontology classes. The 'Data content class' is selected, and its description is shown in the main panel. The description includes a display name, IRI, annotations, and properties. A discussion panel on the right shows a topic titled 'Or Datatype?' started by a guest, with a comment from Rob Lemmens. A project feed at the bottom right shows a recent edit by GillesFalquet.

**Classes**

- owl:Thing
  - Data type
  - VGI Entity 2
    - Actor
      - Analyst
      - Contributor
      - Individual Actor
      - Institution or Agency
    - Application Field
      - Decision or Policy Making
      - Education
      - Mapping
    - Data content class 3
    - Data quality
    - Data source
    - Sensor 2
      - Human sensor
      - In-situ sensor
      - Remote sensor
    - Knowledge Resource
    - Method 1
      - Automated Tagging
      - NLP
      - Validating
    - Project
    - Related domain
    - Relevant theoretical concept
    - Similar domain
    - Task

**Class description for Data content class**

**Display name**  
Data content class

**IRI**  
<http://webprotege.stanford.edu/RyWubZpV7KOhYDNDv1IOZ6>

**Annotations**

rdfs:label	Data content class	lang	X
rdfs:comment	Link to classes like Building, River, Road, from an existing external ontology	lang	X
Enter property	Enter value	lang	

**Properties**

Enter property	Enter value	lang
----------------	-------------	------

**Discussions for Data content class**

Post new topic

**Or Datatype ?**  
started by guest  
unresolved

guest 9 days ago  
Should we have a class 'Datatype' to describe the different types of data that appear as input and output in VGI processes. Possible subclasses could be

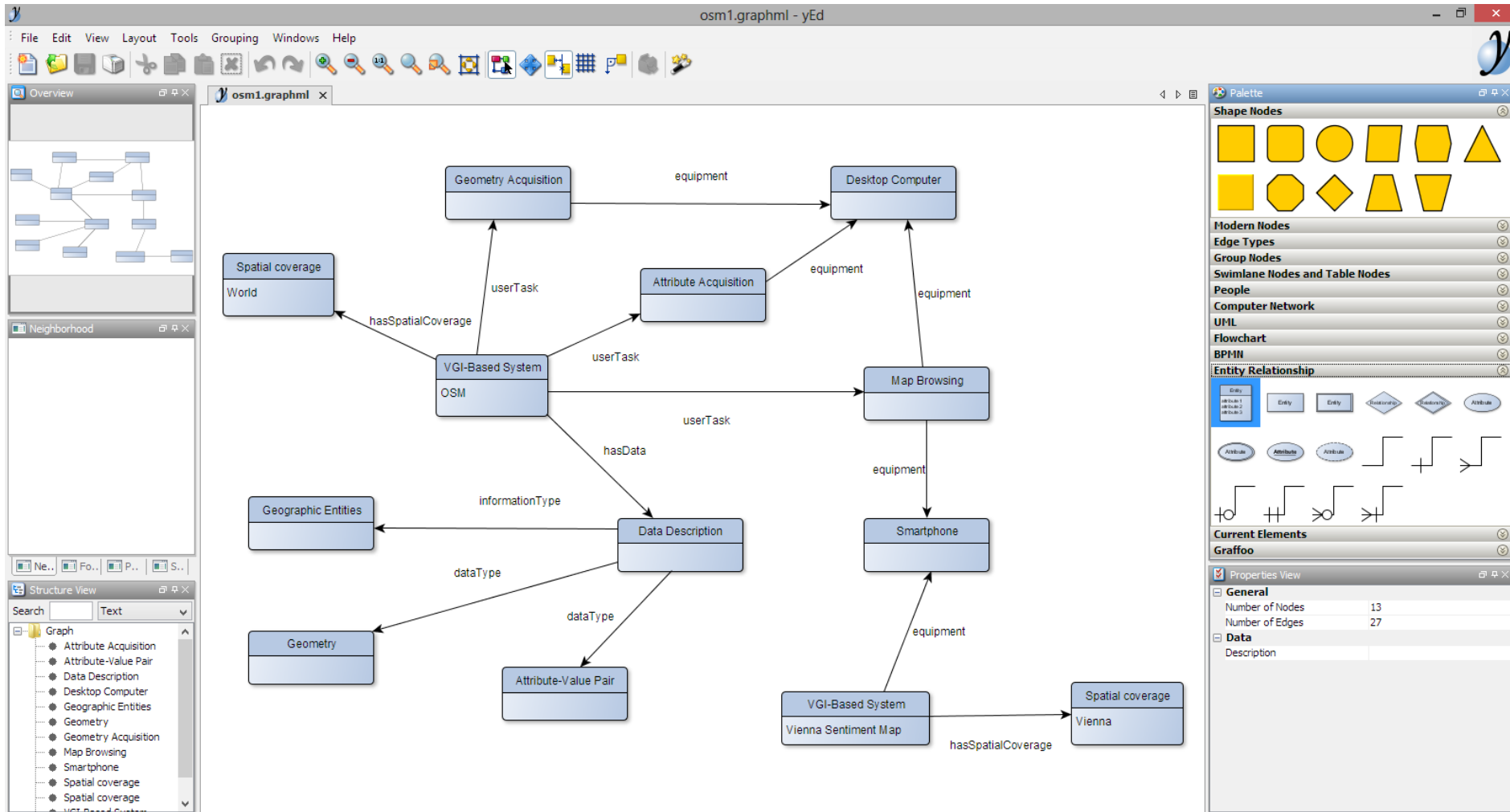
- GeoreferencedData
- Georef Scalar (Nominal, Ordinal, Interval, Ratio)
- Georef Structured
  - Georef Text
  - Georef Image
- NonGeoref Data
- ..

Rob Lemmens 8 days

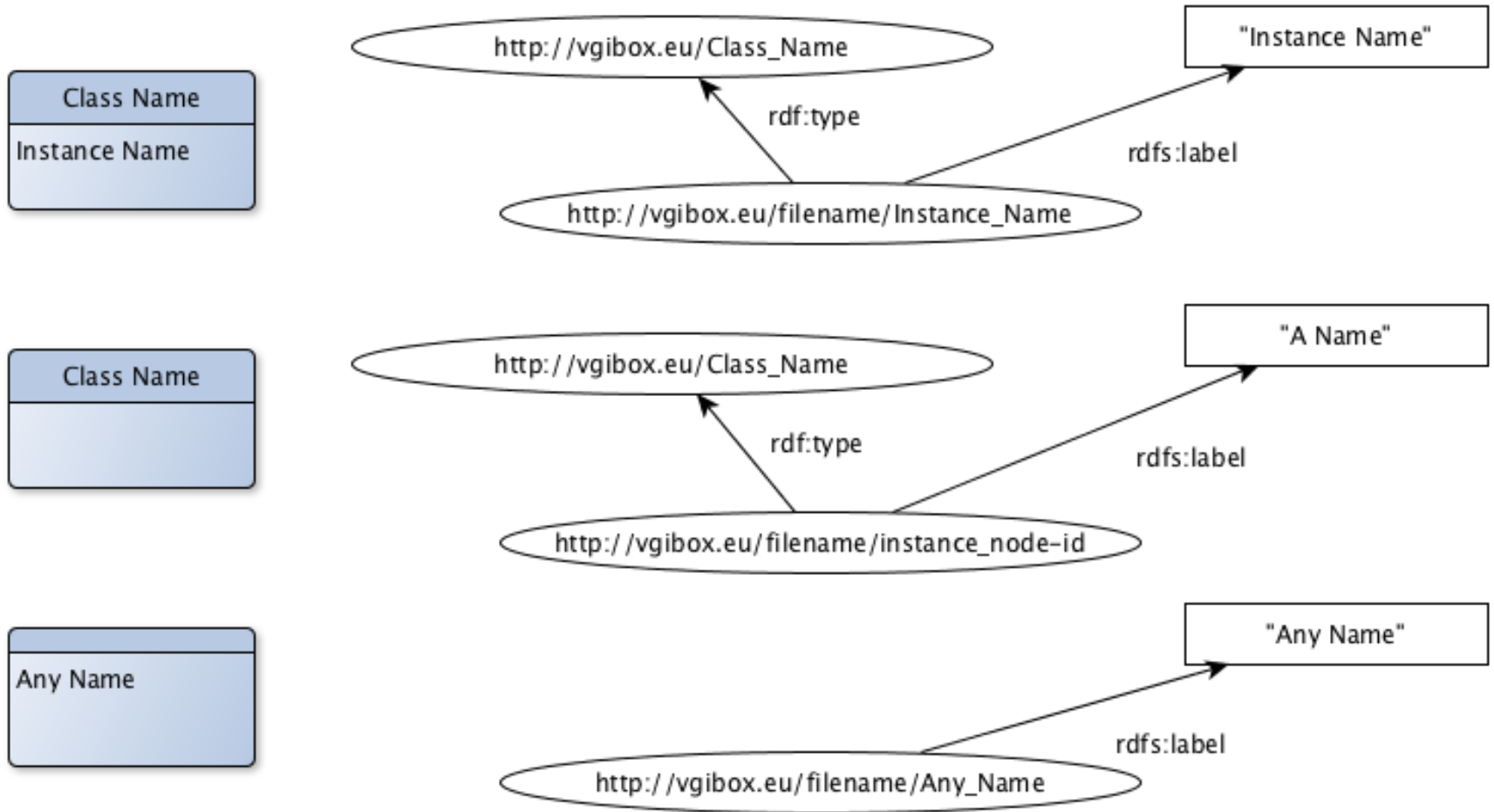
**Project feed**

GillesFalquet edited:  
Class: VGI-based System  
yesterday

# ENERGIC VGI Ontology Instantiation with yEd graph editor

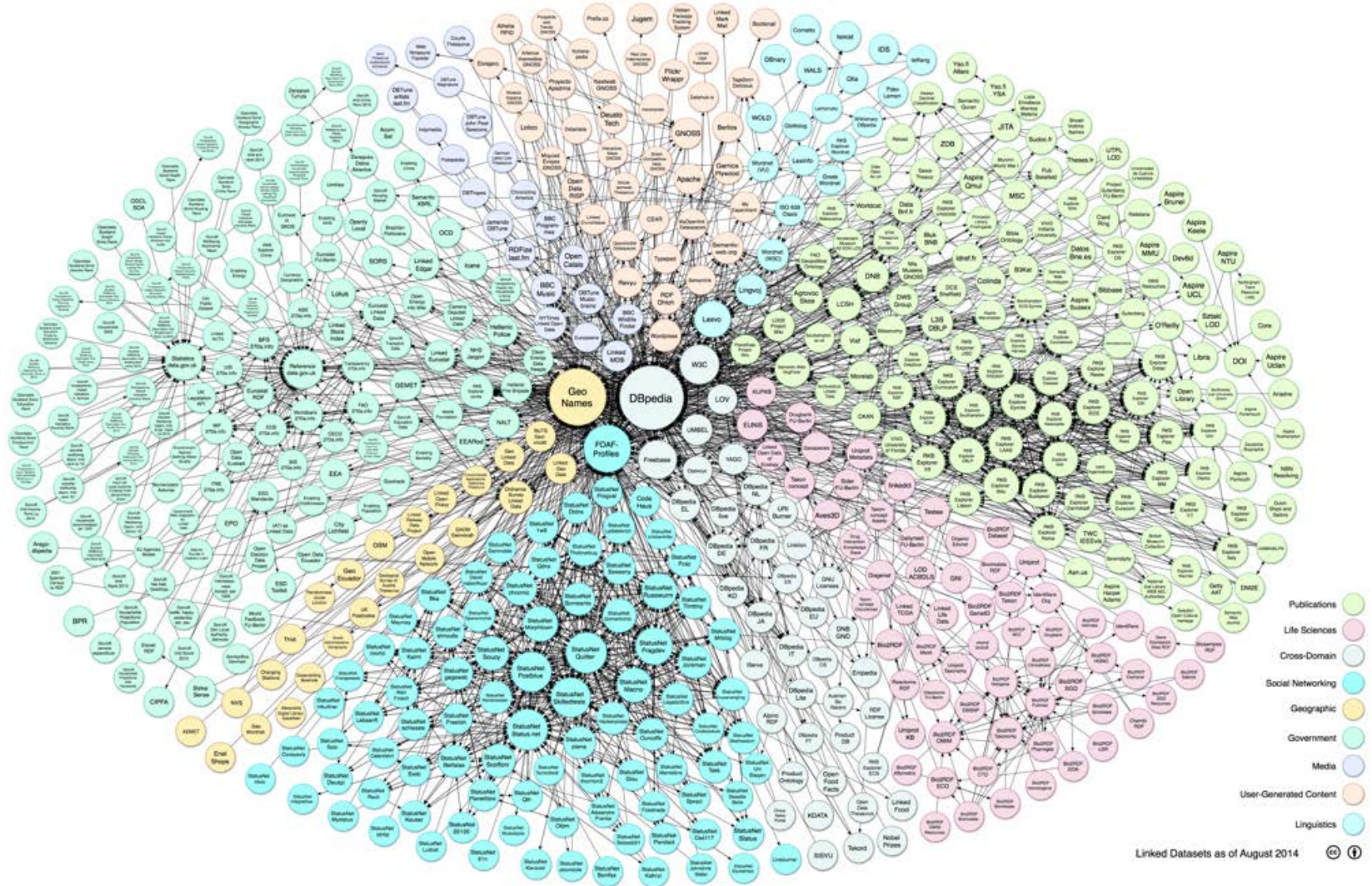


# From E-R diagrams to RDF

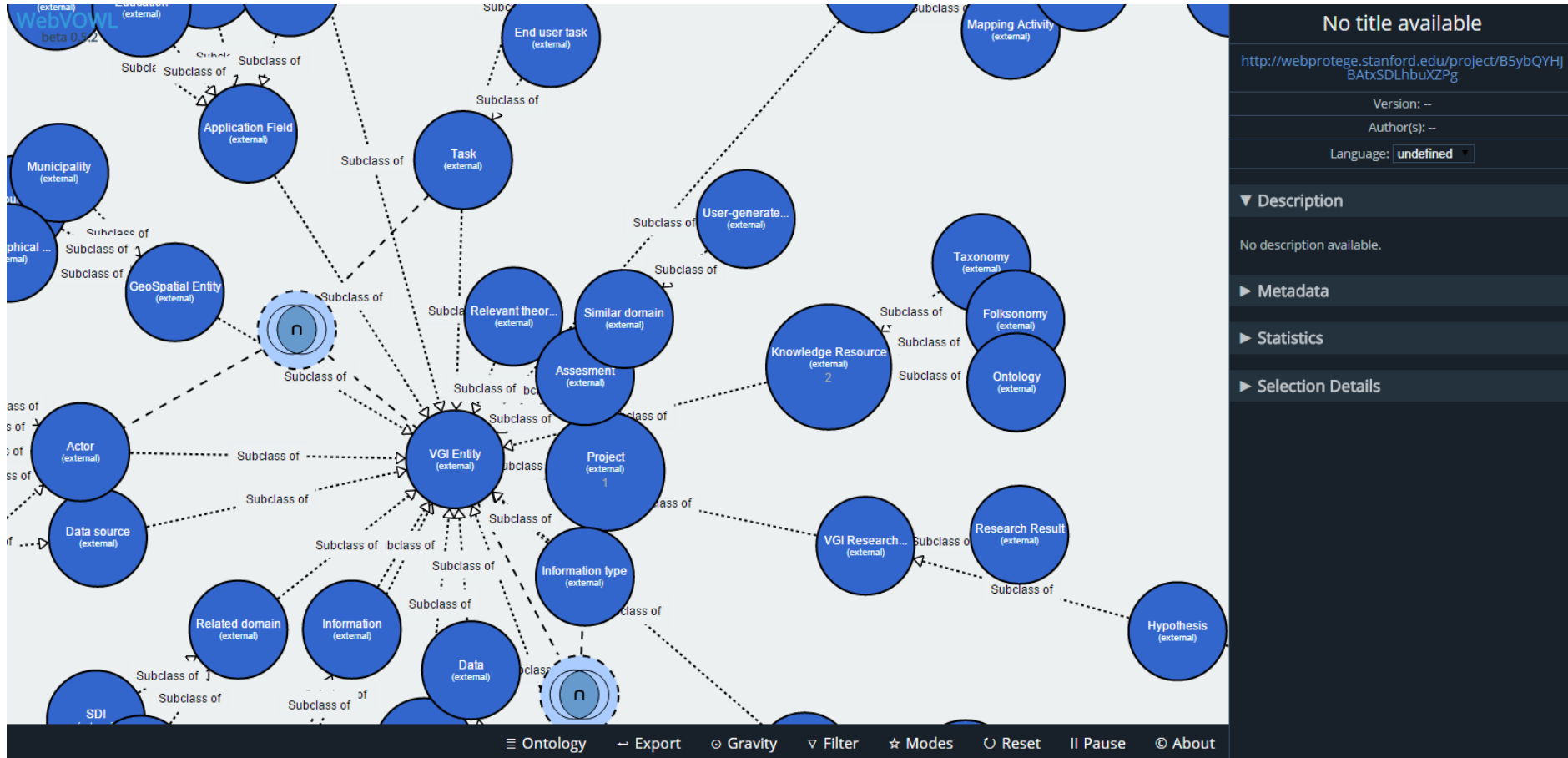




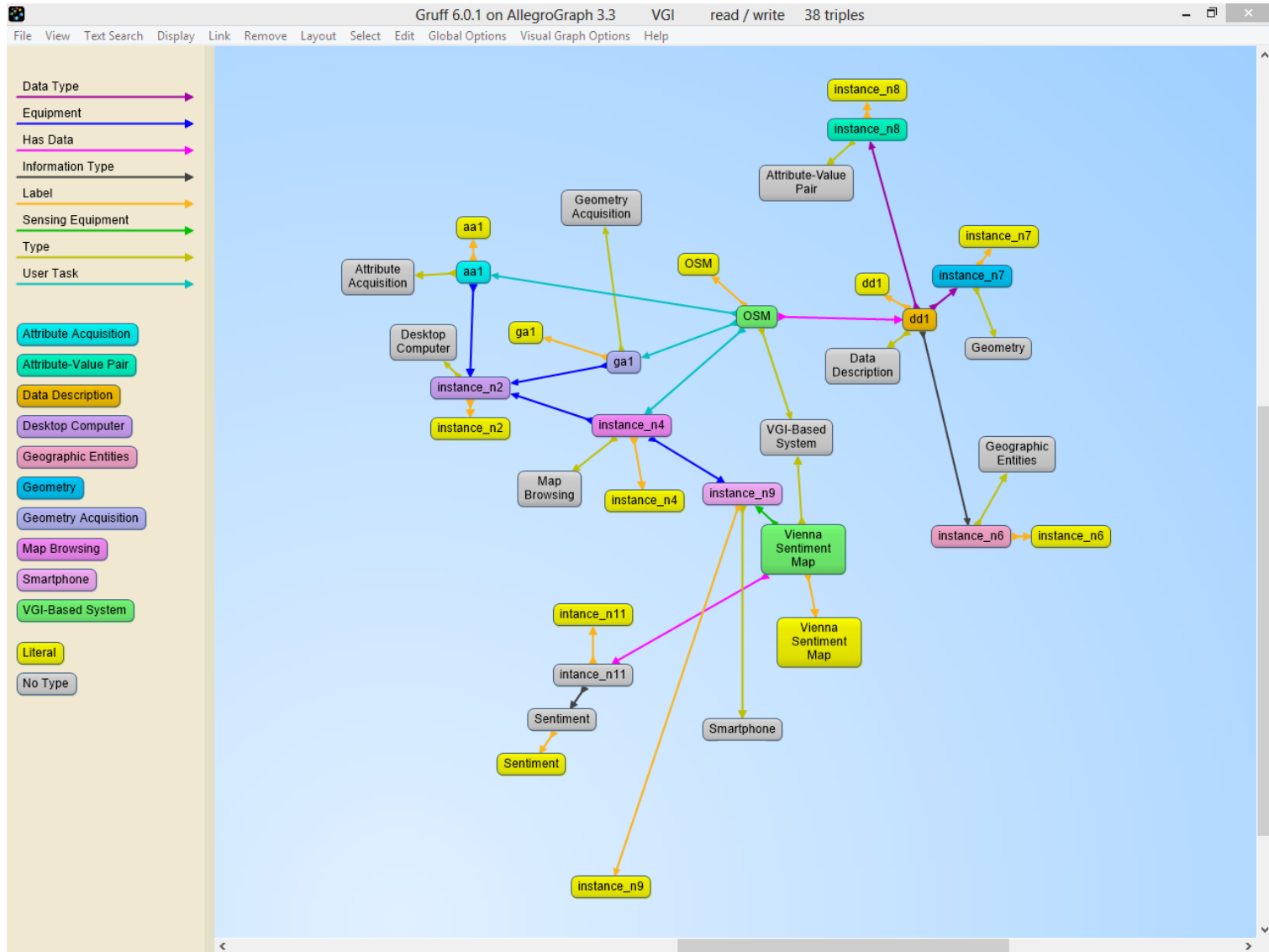
# Linked Open Data cloud



# Visualization with WebVOWL



# Visualization with AllegroGraph - Gruff



# Exploratory querying with Linked Data

**SPEX** Spatio-temporal content explorer Endpoint:

Clear Map    Clear Timeline

16 17 18 19 20 21 22  
November 2015

I am looking for:   
 (2 classes available)  
 Things of a kind  
 Subclasses  
 Instances

VGIbox:Desktop\_Computer  
 VGIbox:Smartphone

Specify relationship from other things (1)

?VGIbox:VGI-Based_System	?VGIbox:Spatial_coverage	?VGIbox:Map_Browsing	?var3
OSM	World	instance_n4	instance_n9
OSM	World	instance_n4	instance_n2

```

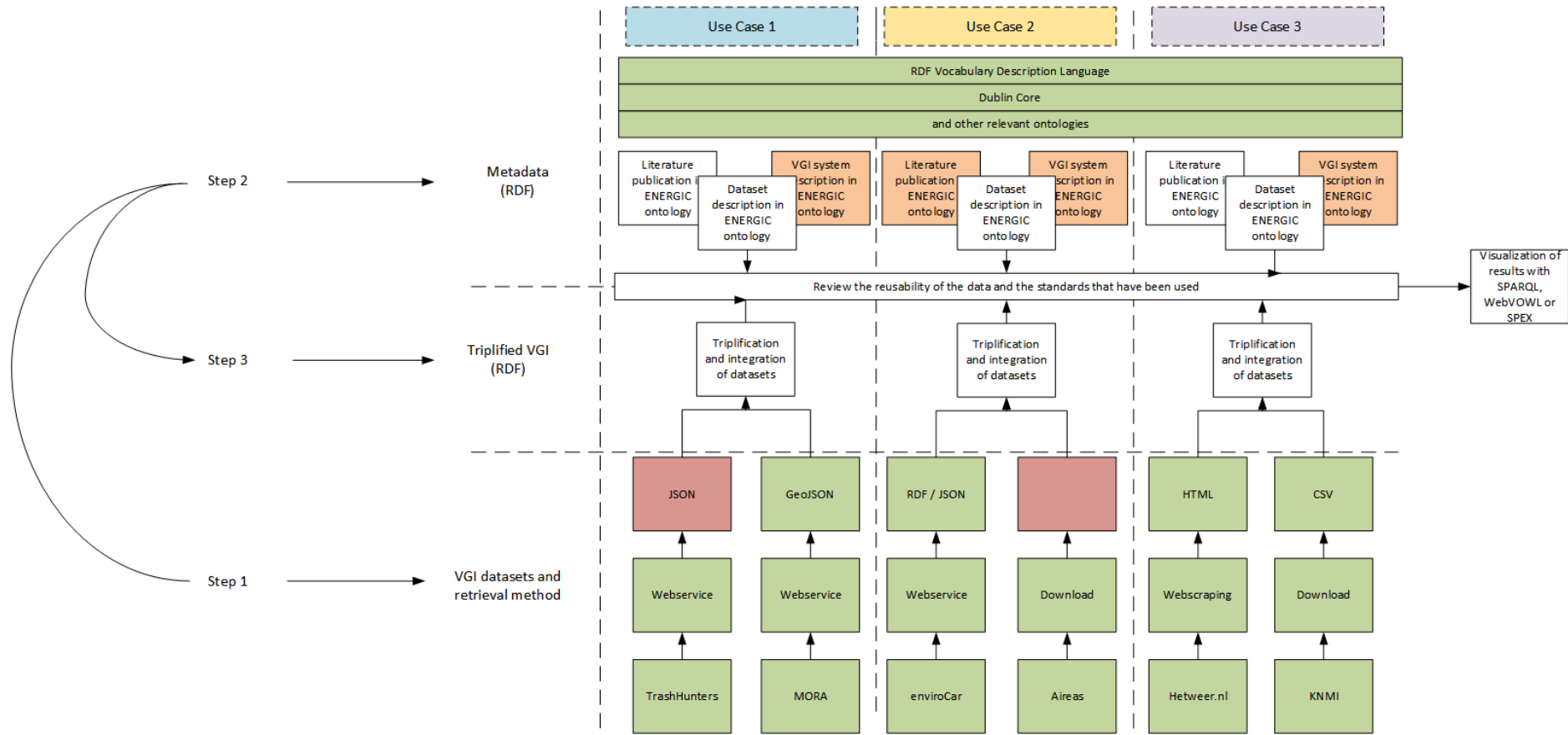
WHERE {
  ?var0 a VGIbox:VGI-Based_System.
  ?var0 VGIbox:hasSpatialCoverage ?var1.
  ?var0 VGIbox:userTask ?var2.
  ?var1 a VGIbox:Spatial_coverage.
  ?var2 a VGIbox:Map_Browsing.
  ?var2 VGIbox:equipment ?var3.
  
```

# ENERGIC Datathon - <http://vgibox.eu/>

Demonstrate the potential of transformations of data to knowledge:

- Challenges:
  1. Analyse spatio-temporal VGI sources (Telecom, Tourist, Transport) with other VGI and authoritative data
  2. Expanding and test ENERIGIC ontology
    - a. Create relevant VGI system descriptions and detailed VGI publication references based on the ENERIGIC ontology and publish these descriptions online.
    - b. Create and execute (Geo)SPARQL queries.
  3. Align the ENERIGIC ontology with other ontologies, such as Geonames, OSM and DBPedia -> create a triple store execute (Geo)SPARQL queries.
  4. Triplify VGI and connect to other Linked Data.
- Input: VGI ontology and other ontologies, VGI sources
- Deliverable (31 July 2016): Report and recorded pitch on results, sources, methods, novelty.
- Prize: Trip to the final COST Energic meeting in London on 12-13 October 2016.

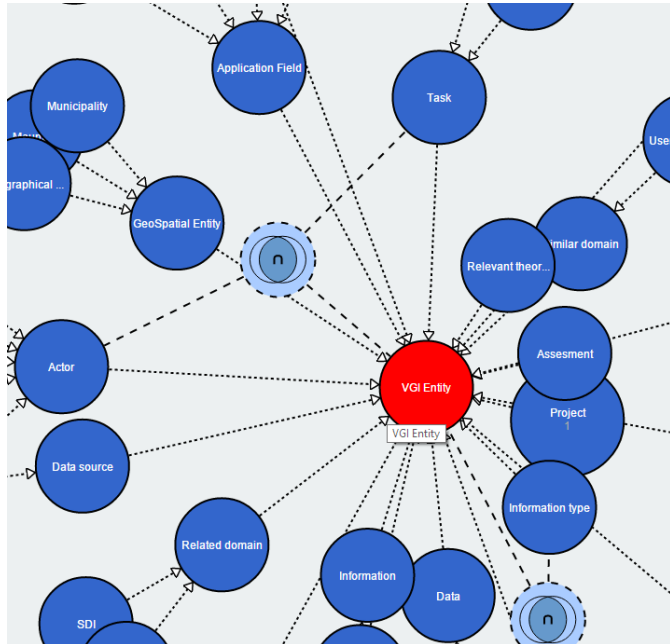
# VGI Reuse



# VGI Reuse

	Findable	Downloadable/Scrapable	Available via webservice or API	Semantically enriched with VGI ontology	Metadata is aggregated via a VGI portal
*	✓	✓			
**	✓	✓	✓		
***	✓	✓	✓	✓	
****	✓	✓	✓	✓	✓

# AGILE Link-VGI workshop, Helsinki 14 June 2016



## Towards Linked Data and ontology development for the semantic enrichment of volunteered geo-information

Rob Lemmens

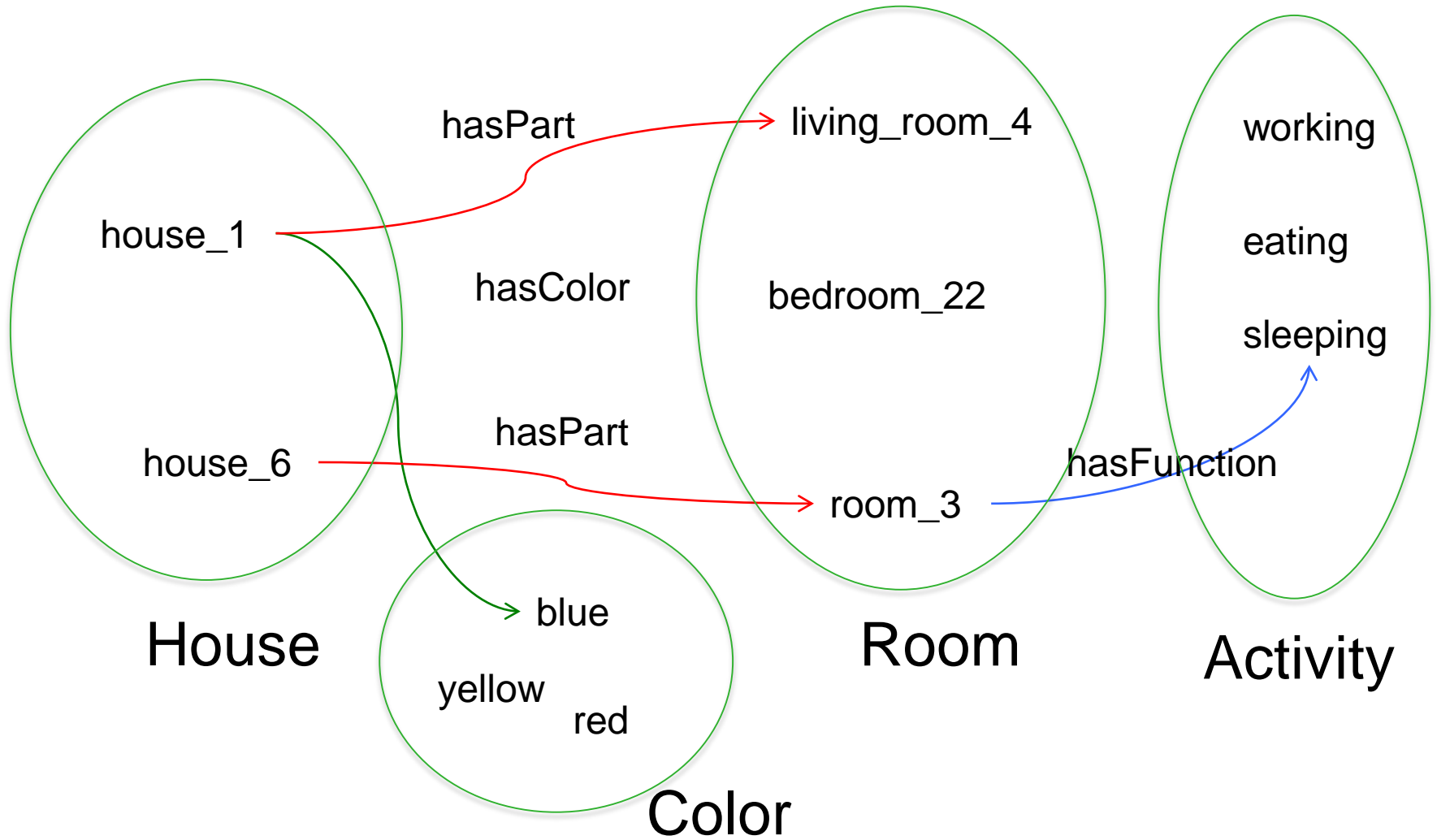
Gilles Falquet,  
Claudine Métral

University of Twente,  
Faculty of Geo-Information  
Science and Earth  
Observation (ITC)  
Enschede, The Netherlands  
r.l.g.lemmens@utwente.nl

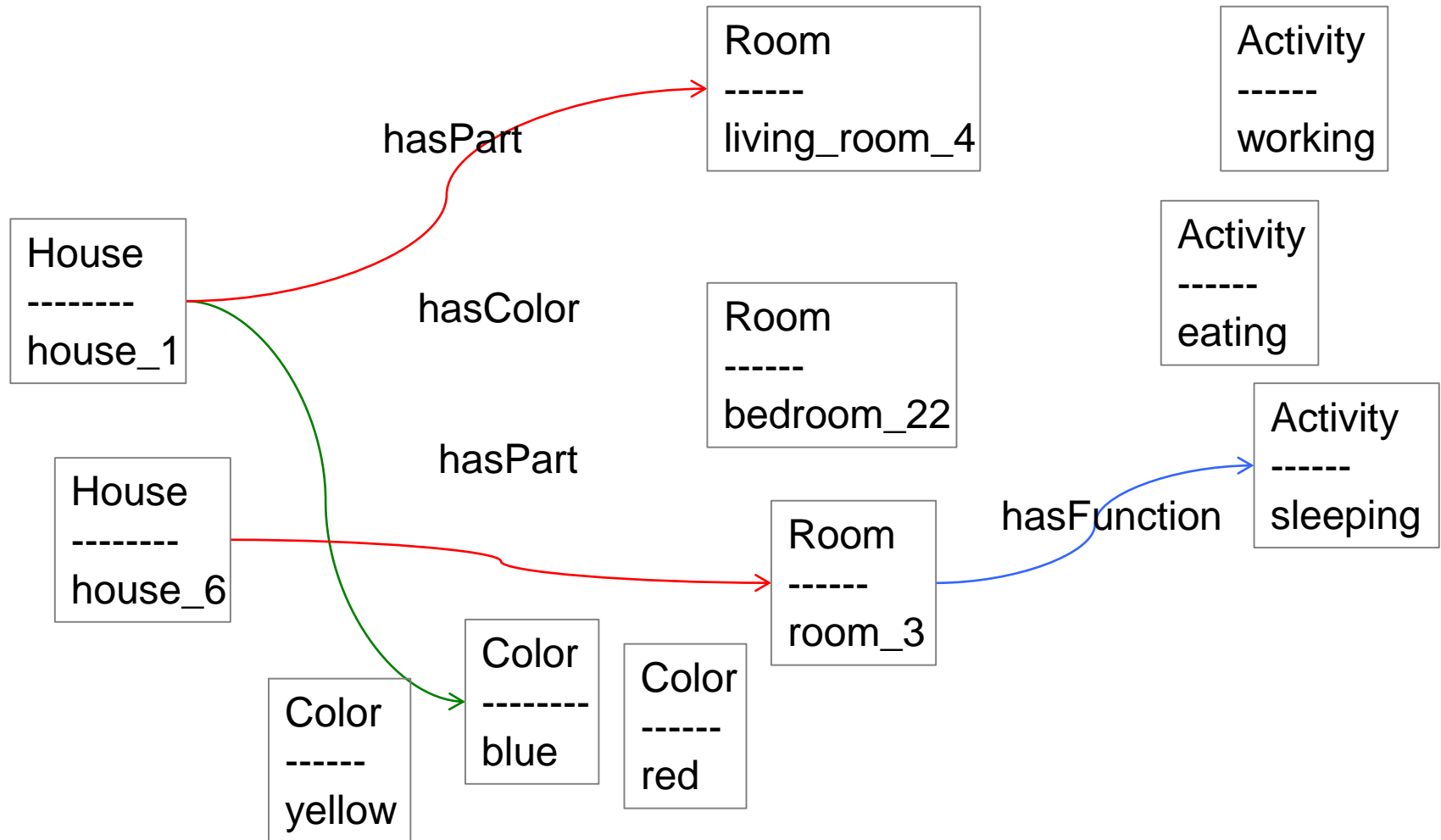
University of Geneva,  
Faculty of economics and  
social sciences,  
Geneva, Switzerland  
{Gilles.Falquet;  
claudine.metr}@unige.ch



# Instances live in classes

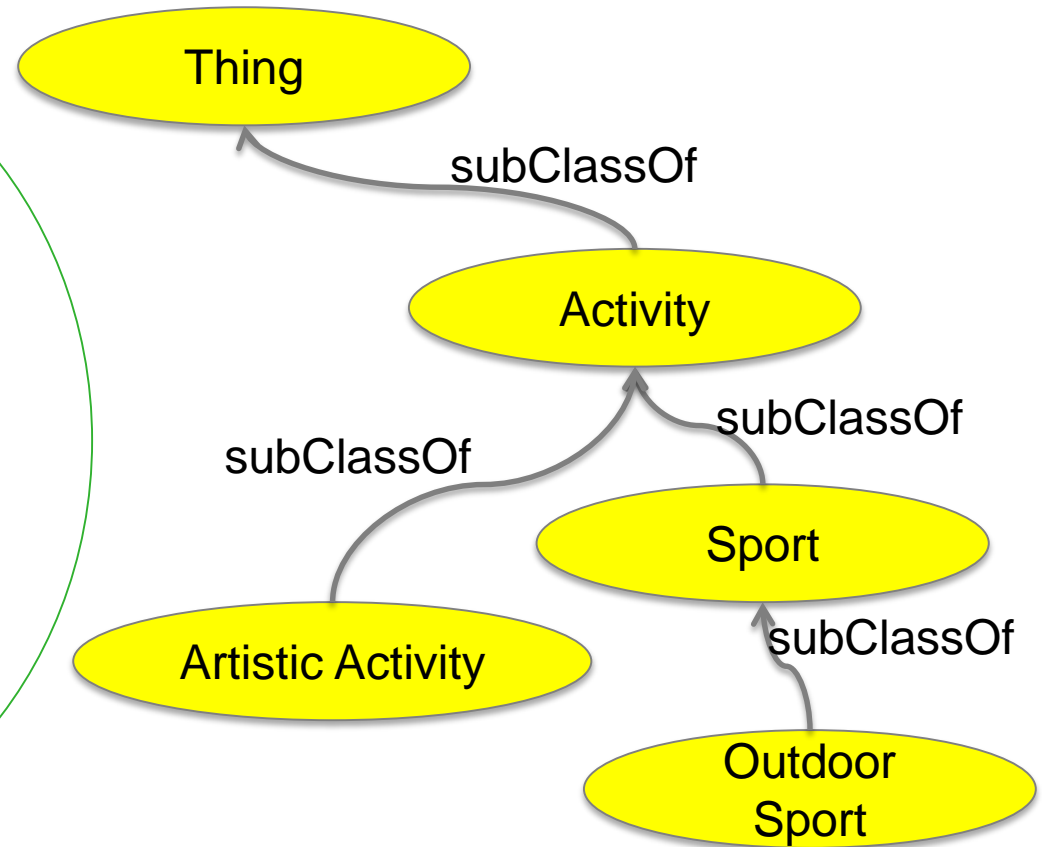
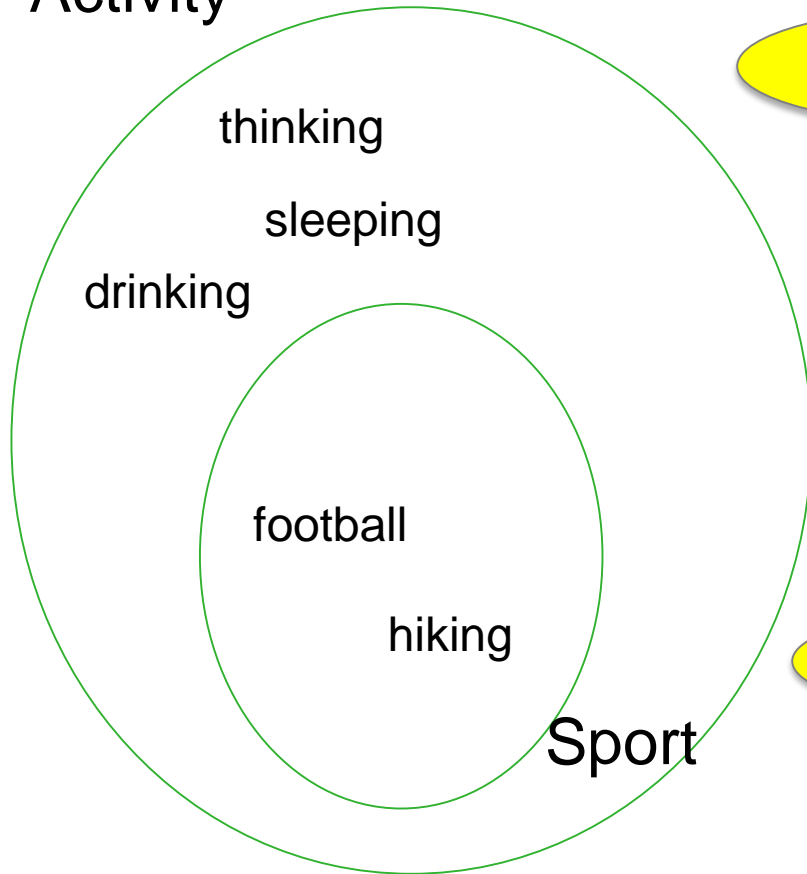


# Another notation

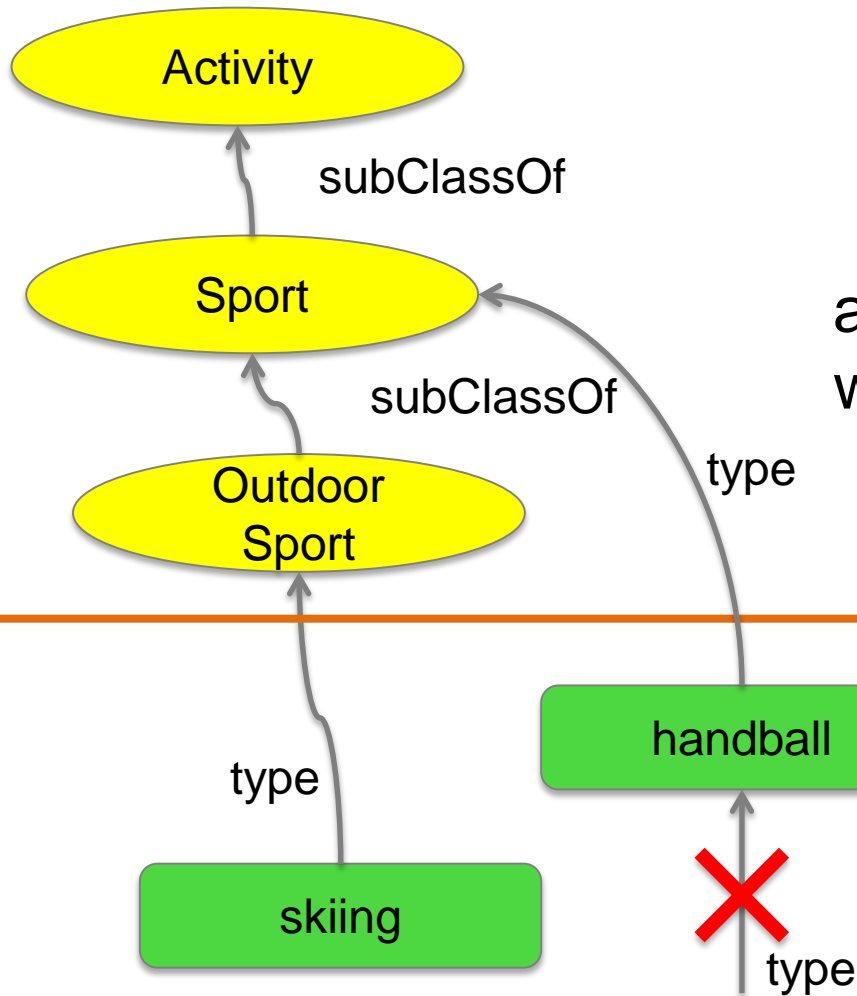


# Classes form an inclusion hierarchy

Activity



Individuals and classes reside at different, clearly separated, levels



# CLASSES

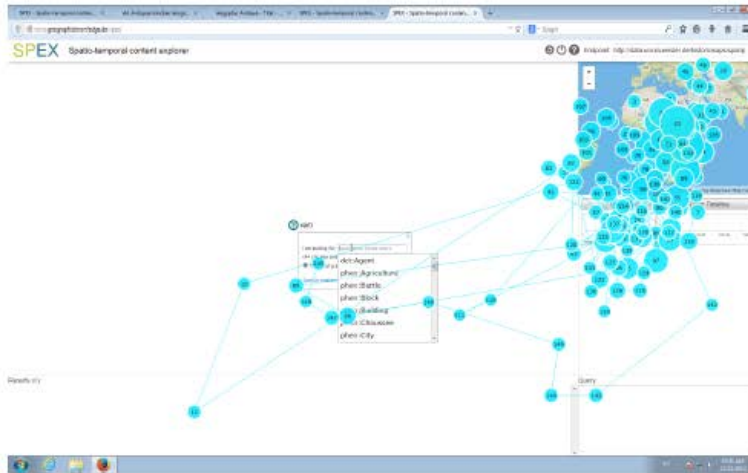
a difficult ontology design decision:  
where to put this limit?

# INSTANCES

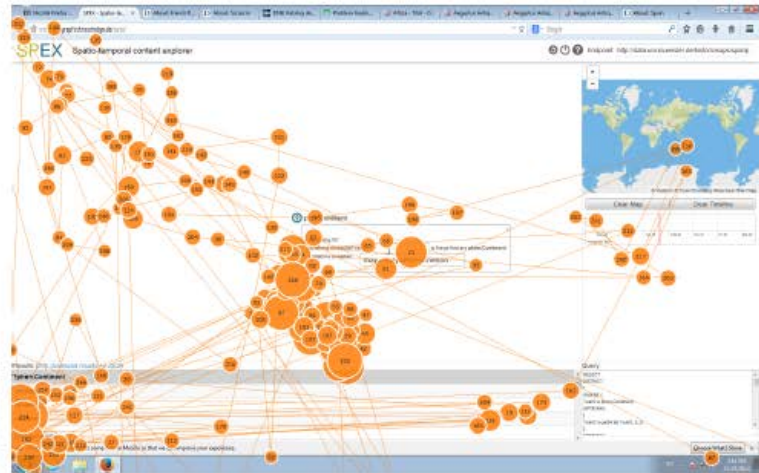
# Usability analysis



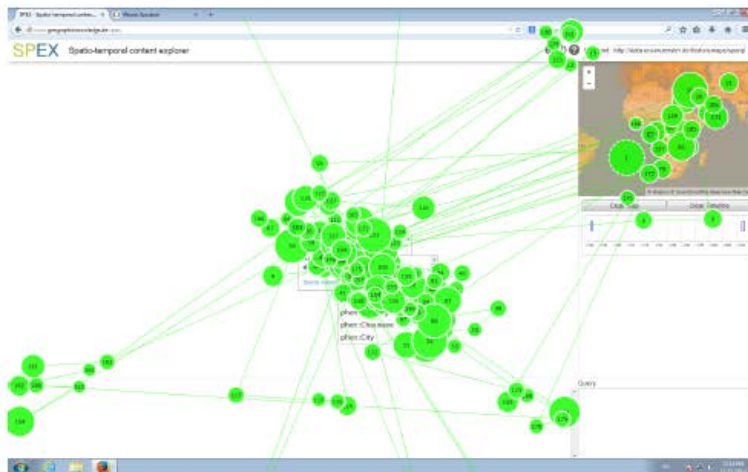
# Usability analysis



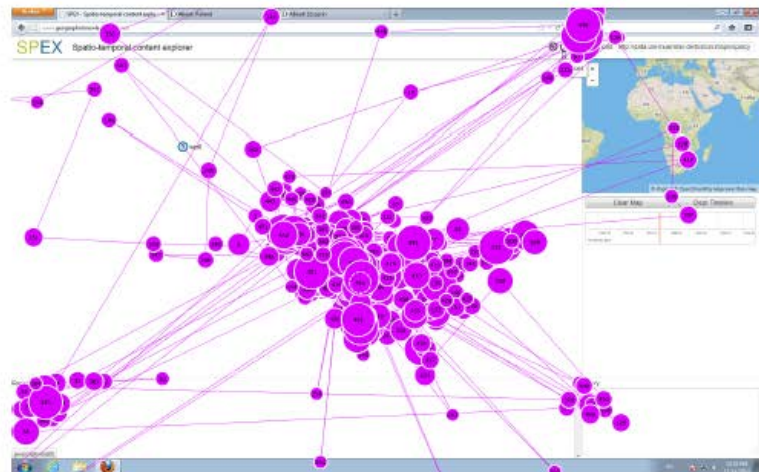
(a) A gaze plot of user 1 for Q 4a.



(b) A gaze plot of user 4 for Q 4a.



(c) A gaze plot of user 3 for Q 4a.



(d) A gaze plot of user 5 for Q 3b.