

# Semantic Challenges for Volunteered Geographic Information

Andrea Ballatore

Center for Spatial Studies

University of California, Santa Barbara

RICH-VGI Workshop  
AGILE 2015, Lisbon

**spatial@ucsb**  
CENTER FOR SPATIAL STUDIES



# spatial@ucsb

CENTER FOR SPATIAL STUDIES

spatial.ucsb.edu

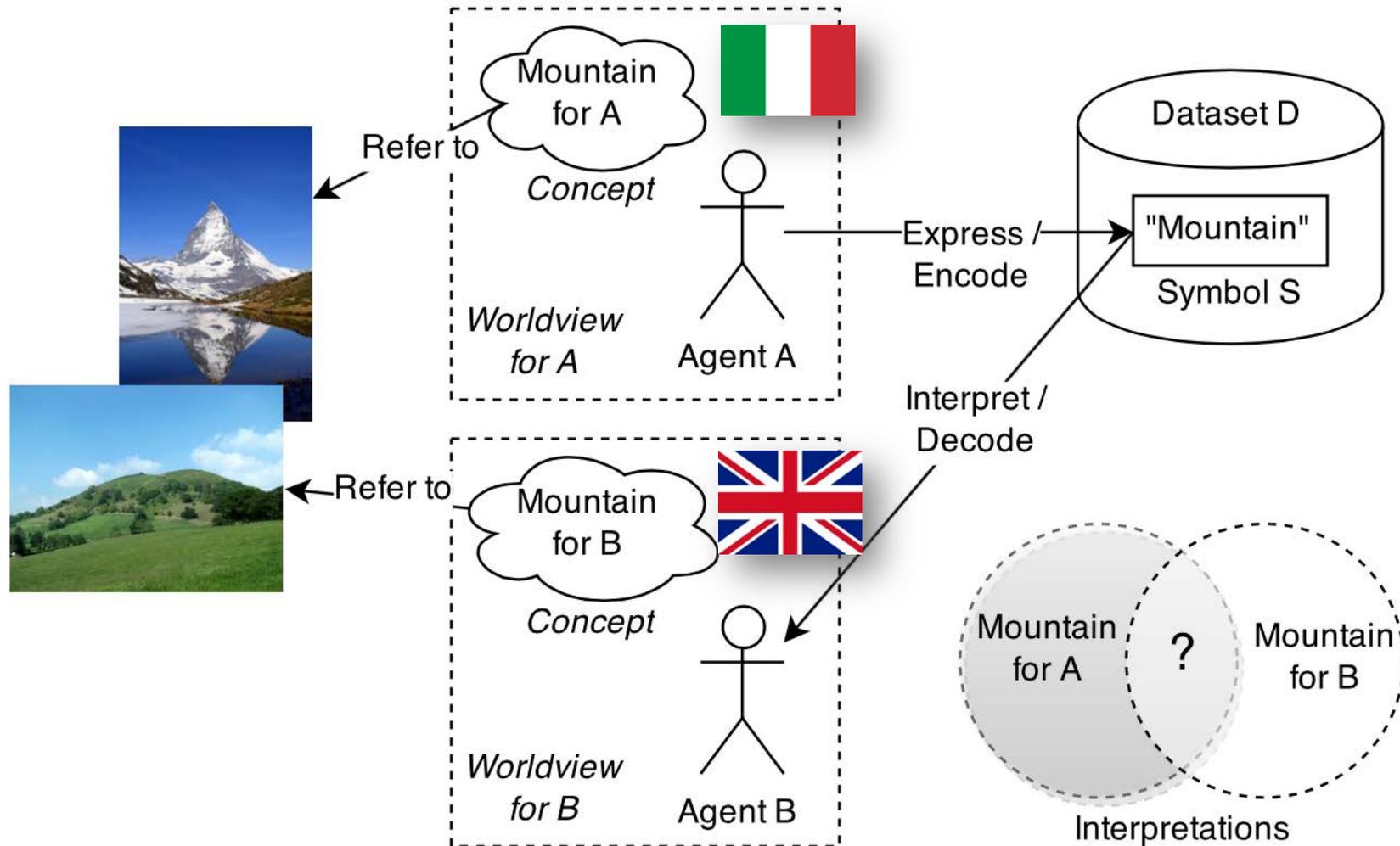


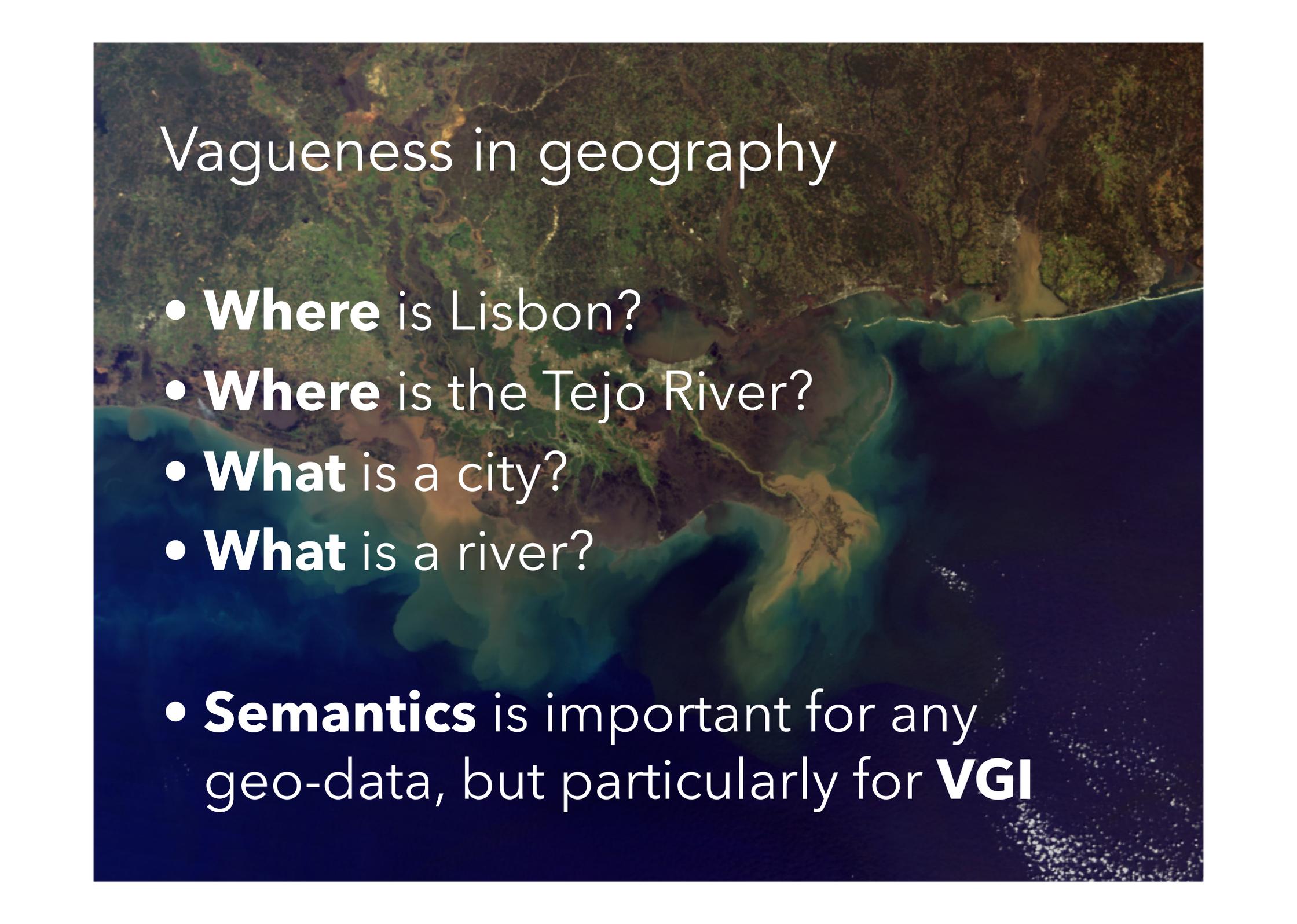


# Talk outline

1. The "Semantic Gulf"
2. Semantics of OpenStreetMap
3. Measuring Conceptual Quality

# The Semantic Gulf



An aerial satellite image of a coastal region, likely the Tejo River delta in Portugal. The image shows a complex network of waterways and land with dense green vegetation. The text is overlaid on the left side of the image.

# Vagueness in geography

- **Where** is Lisbon?
  - **Where** is the Tejo River?
  - **What** is a city?
  - **What** is a river?
- 
- **Semantics** is important for any geo-data, but particularly for **VGI**

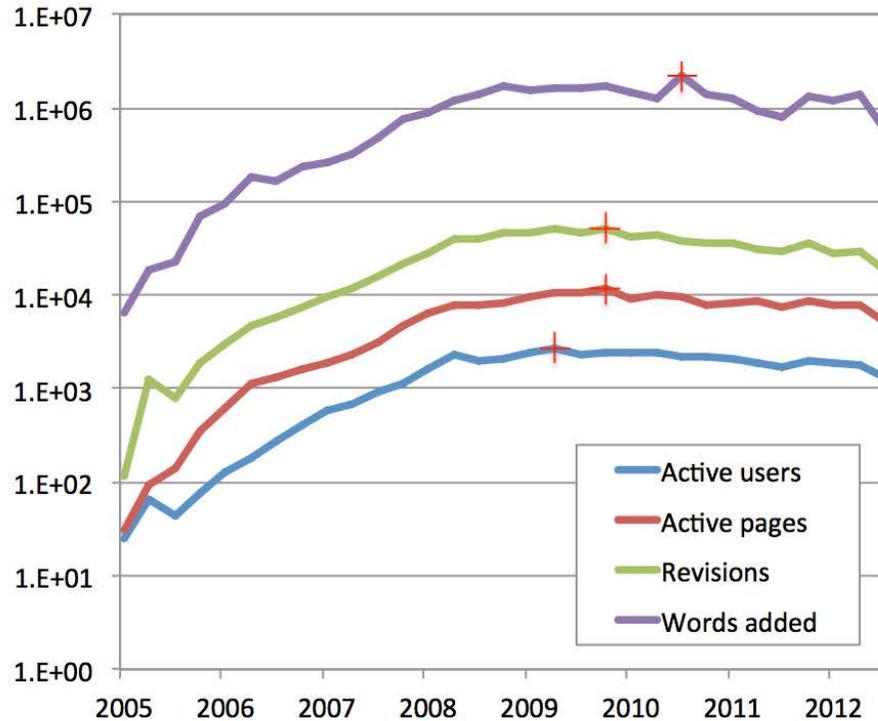


# The semantics in OSM

- **Shared conceptualization**
- Semi-structured **folksonomy** (tags)
- **Semantic ecosystem**: vector, wiki, mailing lists
- **Negotiation**: alternative | incommensurable conceptualizations

#amenity=bar #amenity=nightclub #amenity=hospital #amenity=post\_office  
#amenity=post\_box #amenity=fountain #amenity=fitness\_center #amenity=bbq  
#amenity=pharmacy #amenity=atm #amenity=bench #amenity=dentist  
#amenity=community\_centre #amenity=fast\_food #amenity=parking  
#amenity=grave\_yard  
#amenity=fuel #amenity=place\_of\_worship #amenity=cafe  
#amenity=school  
#amenity=police #amenity=bicycle\_parking #amenity=bicycle\_rental  
#amenity=restaurant #amenity=veterinary #amenity=marketplace  
#amenity=physical\_therapy #amenity=cinema #amenity=bus\_station  
#amenity=bank #amenity=courthouse #amenity=clinic #amenity=arts\_centre  
#amenity=prison #amenity=fire\_station #amenity=drinking\_water  
#amenity=pub #amenity=auto\_mechanic #amenity=library #amenity=toilets  
#amenity=rv\_dumpstation #amenity=public\_building #amenity=railway

# OSM Wiki website



(a) Global wiki activity (log scale)



(Forthcoming in IJGIS, **Ballatore & Mooney 2015**)

# OSM Wiki website



Relation	Top 1%	Top 10%
Revisions per contributor	39.1	81.5
Added words per contributor	48.6	87.9
Deleted words per contributor	51.0	90.3
Revisions per page	30.2	71.0
Added words per page	39.5	81.4
Deleted words per page	71.5	98.7

i.e., extreme contribution inequality!

# OSM Wiki website



Most negotiated terms

$N_p$	Key/tag page
1	highway
2	boundary
3	building
4	barrier
5	access
6	traffic calming
7	amenity=place of worship
8	waterway=riverbank
9	service
10	wheelchair
11	natural=tree
12	shop
13	bridge
14	highway=turning circle

The dimensions of negotiation

# Ontological negotiation

Mereo-topology

*Entities and their  
attributes, relations,  
and parts*



The dimensions of negotiation

# Cultural and linguistic negotiation

Conceptualization in British English

Problems of equivalence

Tension between universal and local



# The dimensions of negotiation



I don't give a flying monkey's for tag voting, automatic changebots, endless discussions, **categories**, or any of that crap, but prefer to get on and actually **do** stuff.

# Poisonous people

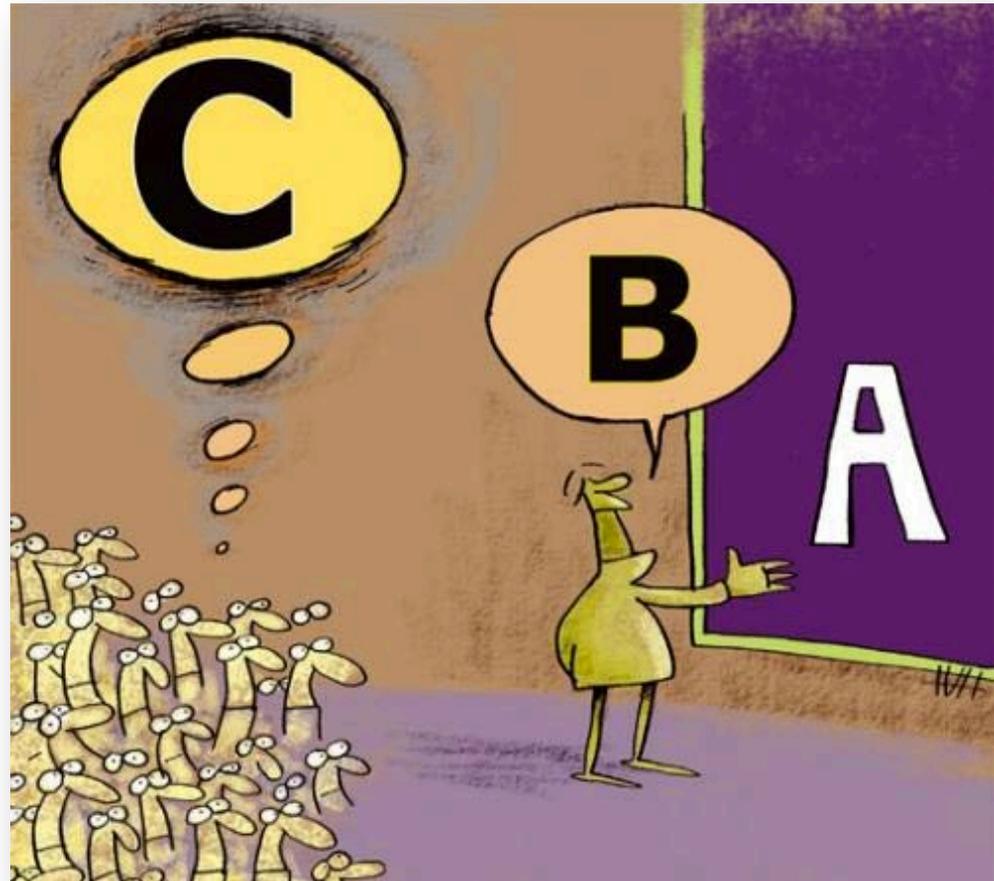
'disinfection' of the project from 'poisonous people' who "drain, paralyse, slow, cause needless infighting and destroy the attention and focus of a community . . .

[they] are wrecking the time, focus and goodwill of the majority of contributors, creating dissent out of nothing and even purposefully breaking our data"

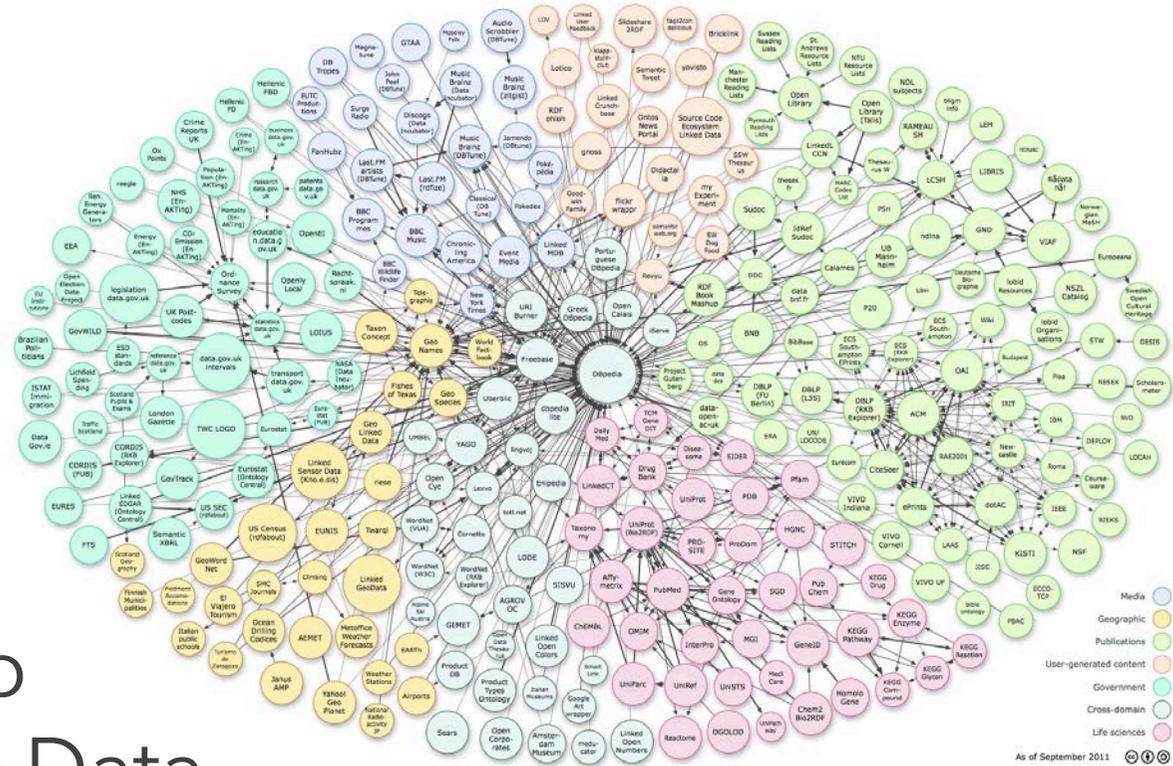
(Coast, 2010)

# EnRICHing: More explicit semantics

- Cooperation
- Exploration
- Retrieval
- Integration



# Semantic technologies (for VGI)



- Semantic Web
- Linked (open) Data
- E.g., **OSM Semantic Network**

# Conceptual quality

- **Information value** is not only in geometric and positional quality
- **Interpretation** and **intended meaning** of the data
- Operationalize conceptual quality to **measure the interpretability** of documentation, vocabularies, schemas, ontologies

# Conceptual quality

Accuracy

Granularity

Completeness

Consistency

Compliance

Richness

(Forthcoming in COSIT, Ballatore & Zipf 2015)

# Conceptual quality

**Accuracy:** Distance between conceptualization and domain knowledge. Degree of correctness in the classification of features.

**Granularity:** Level of thematic description present in the data, from abstract to specific concepts.

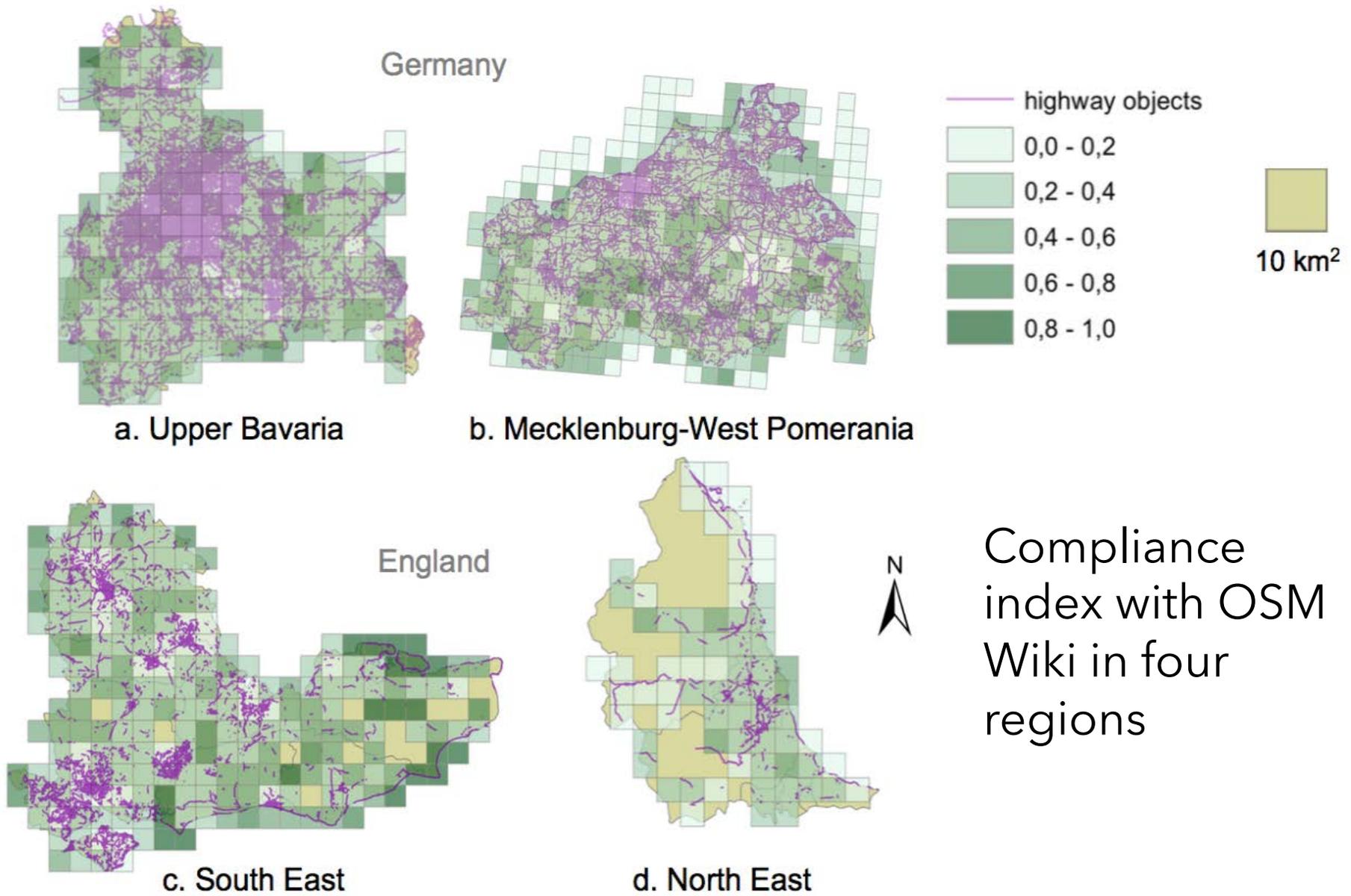
**Completeness:** Coverage in the conceptualization of the features of interest.

# Conceptual quality

**Consistency:** Degree of homogeneity in the descriptions of geographic features.

**Compliance:** Degree of adherence to a given source (standard, schema, etc).

**Richness:** Amount and variety of dimensions that are included in the description of the real-world entity.



# Thanks!

Andrea Ballatore

 [aballatore@spatial.ucsb.edu](mailto:aballatore@spatial.ucsb.edu)

 [@a\\_ballatore](https://twitter.com/a_ballatore)

**spatial@ucsb**  
CENTER FOR SPATIAL STUDIES