

KDI

A Methodology for Data Integration

Fausto Giunchiglia and Mattia Fumagalli

University of Trento





Overview of the Model



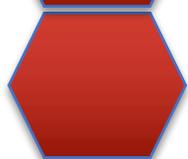
Generalized Queries



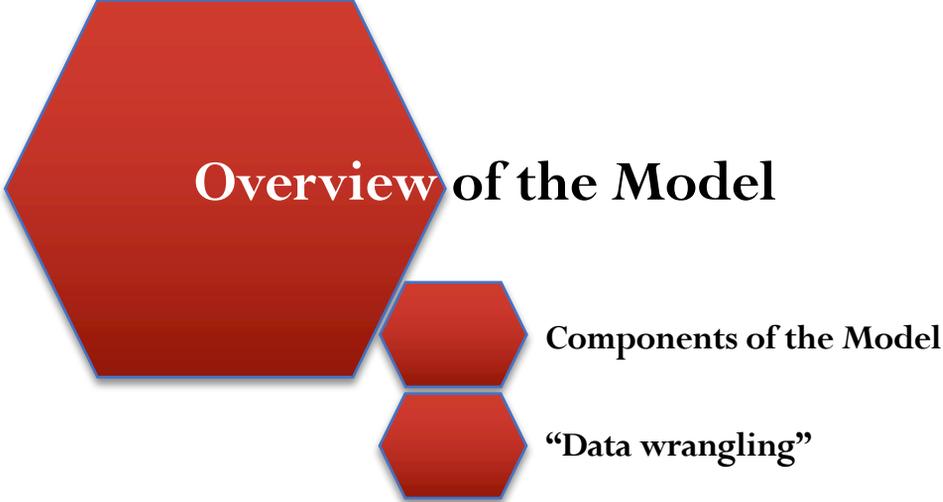
Etypes Model



Evaluation



Case Studies

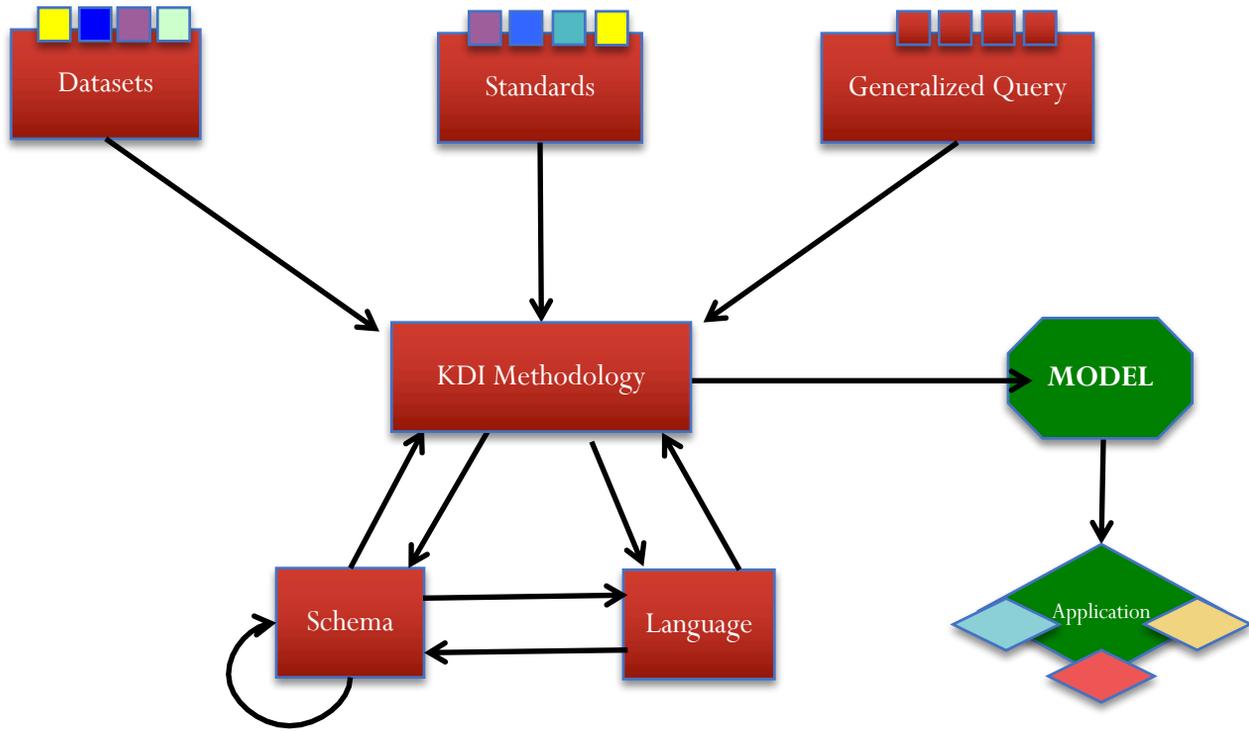


Overview of the Model

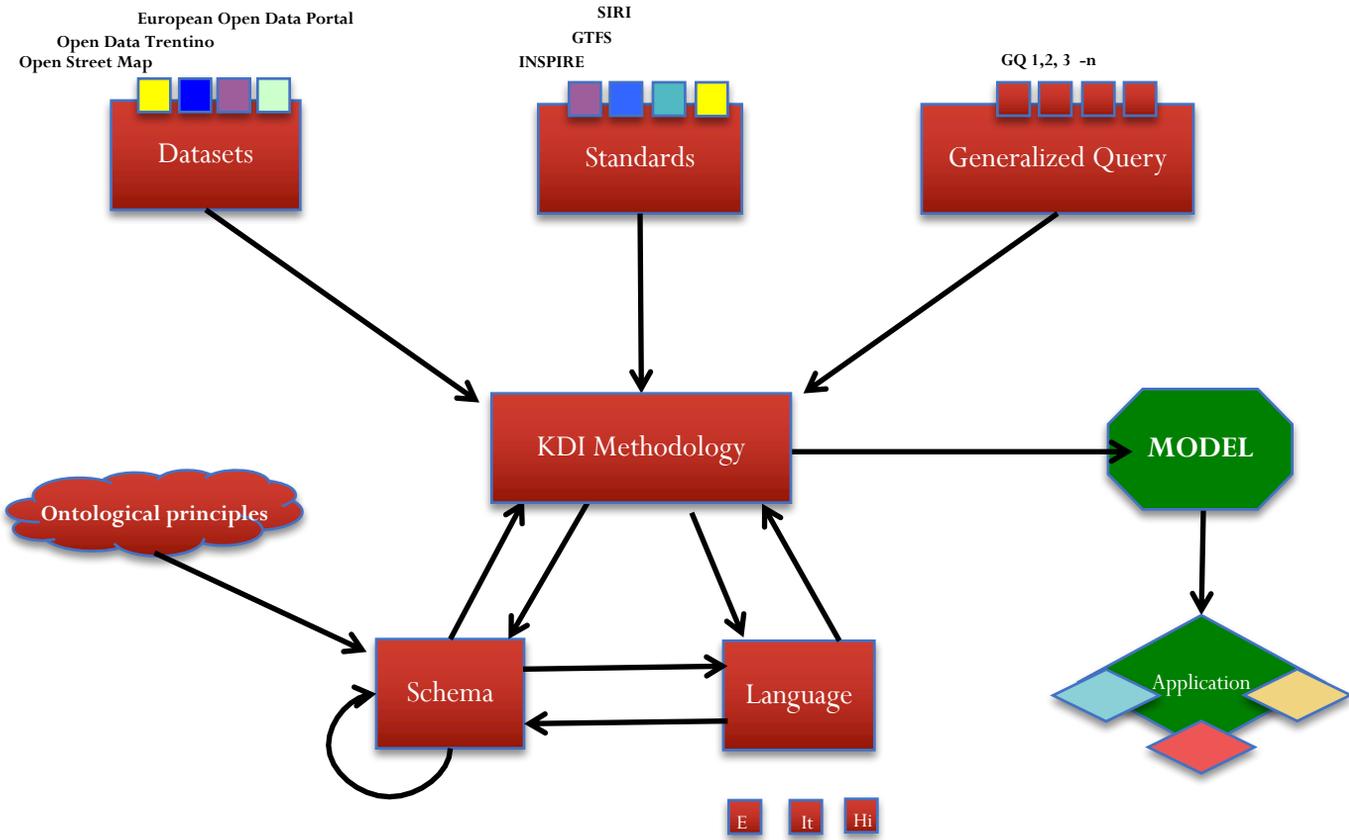
Components of the Model

“Data wrangling”

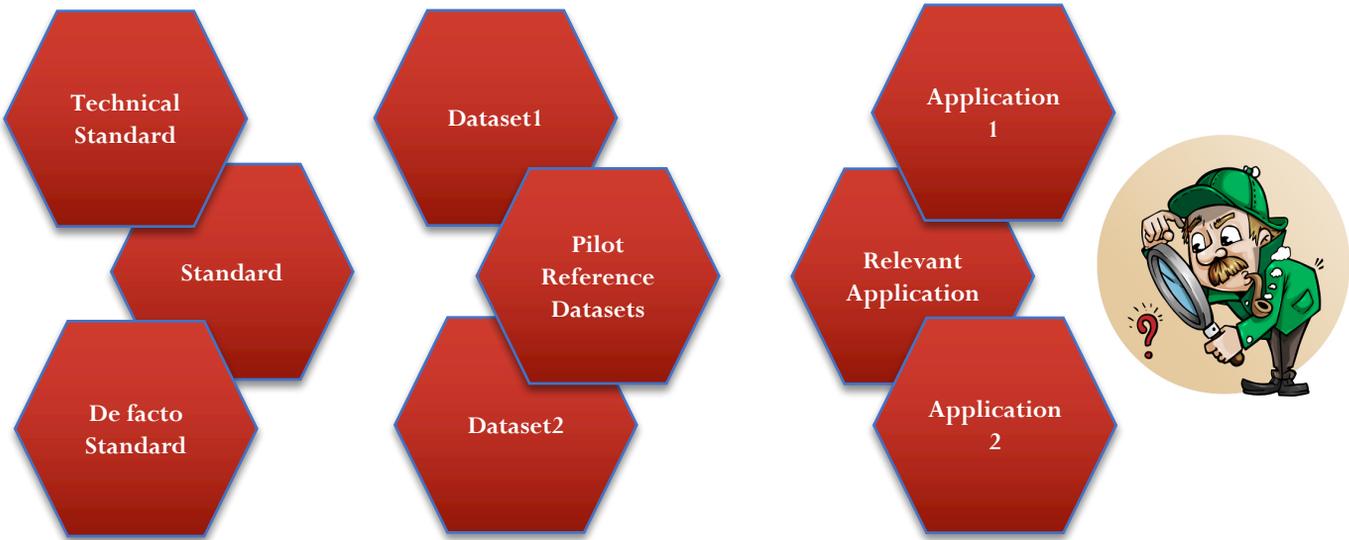
Components of the Model



Components of the Model



“Data wrangling”



Generalized Queries



Application Scenario



Identify the Concepts



Queries Collection Mechanism

Application Scenario

Choose the application scenario

Transport

Tourism



Start with a set of ground queries :

Given the application scenario, a set of queries will arise which place demands on an underlying ontology.

- Give a list all the Hotels in X City which has facility for disable ?
- Identification of general query pattern

Give me all X in Y AND WHERE.property.True

- Identification: Concepts and Properties

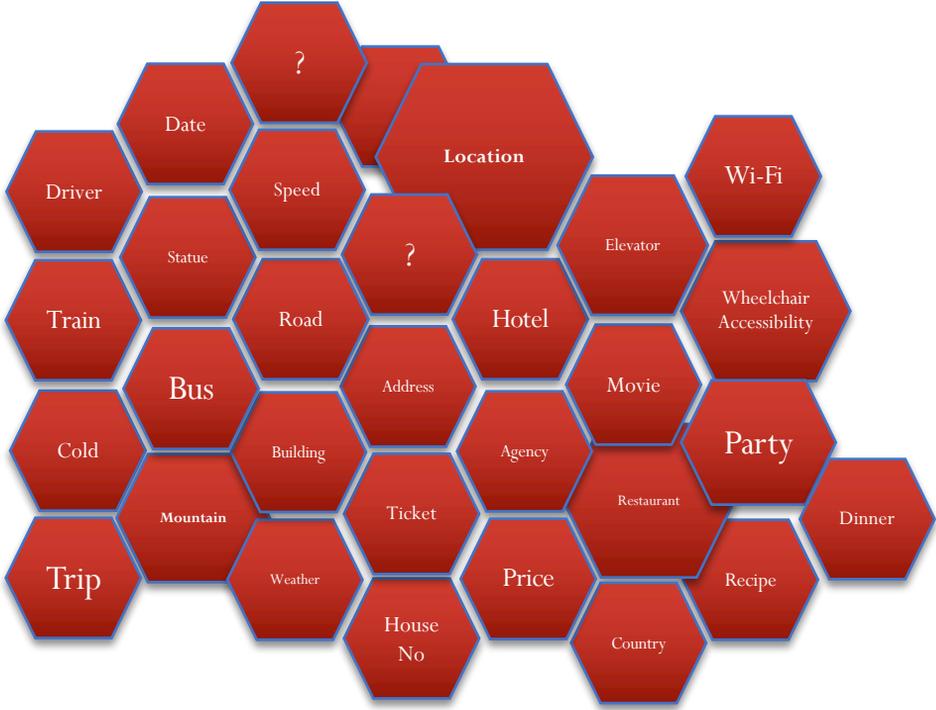
Entity: Hotel, City

Property: Hotel.name, City.name, facilityForDisable.

Boolean

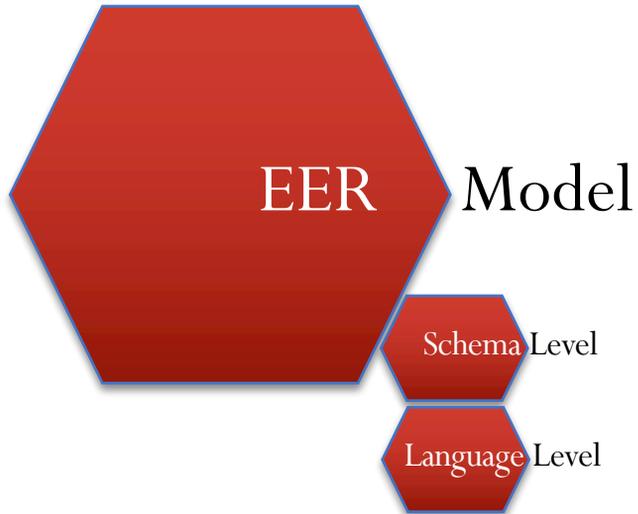
Identify the Concepts

Identify all the **core concepts** which are needed to answer the **generalized queries**.

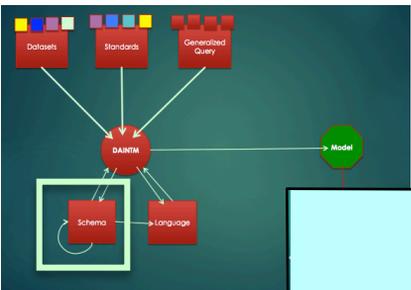


➤ Query generation methodology

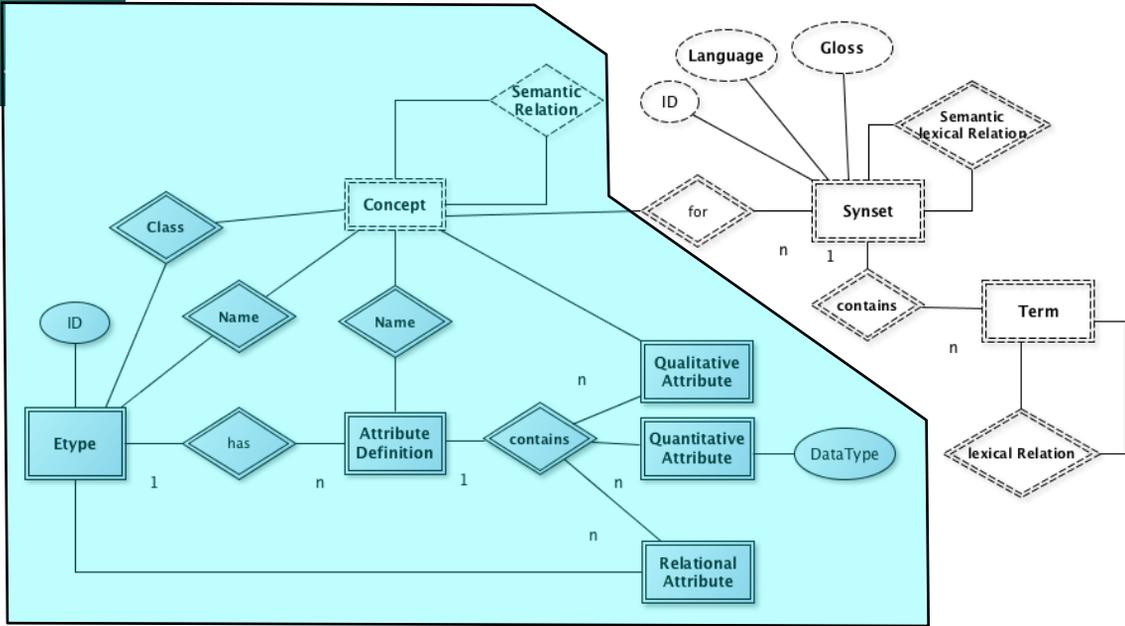
1. via a user study, for instance via questionnaires or focus group
2. via a benchmarking analysis of existing sites and data
3. heuristically based on the understanding of the domain developer
4. from datasets – (see **rapidminer** tree example... see also <http://quepy.machinalis.com/>)
5. a combination of the above



Schema Level

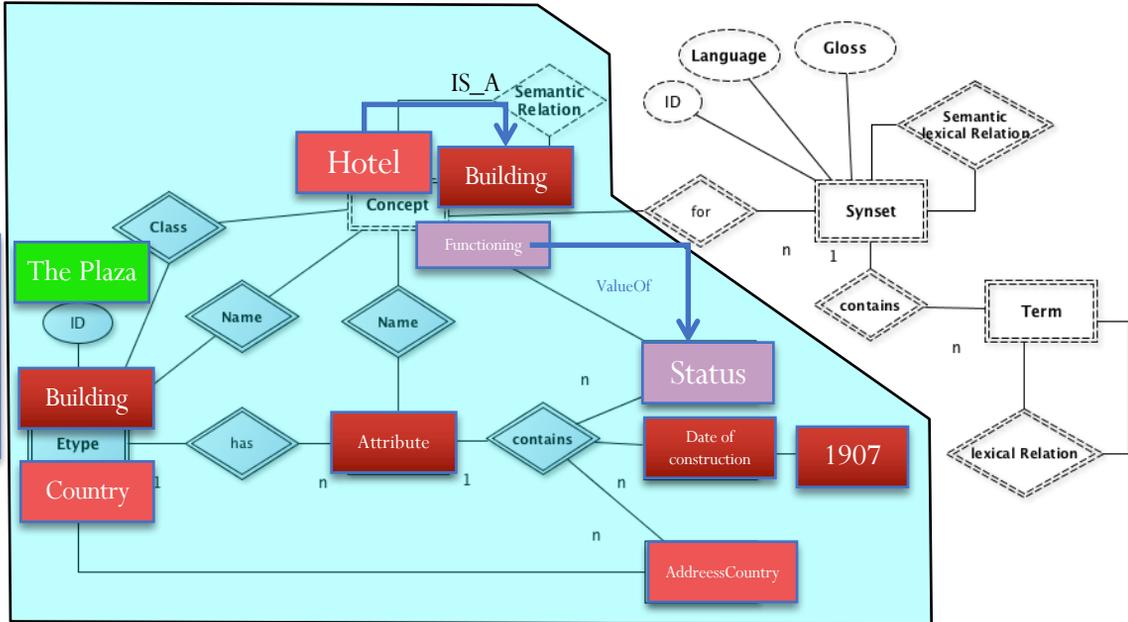


Schema

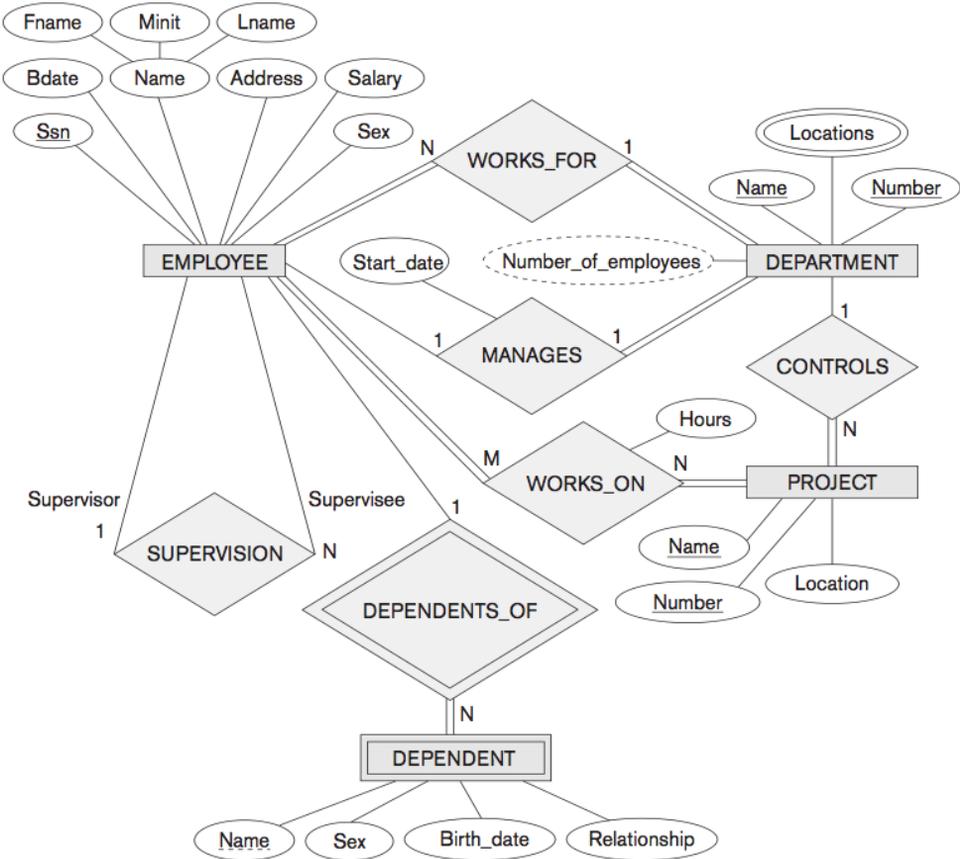


Schema Level

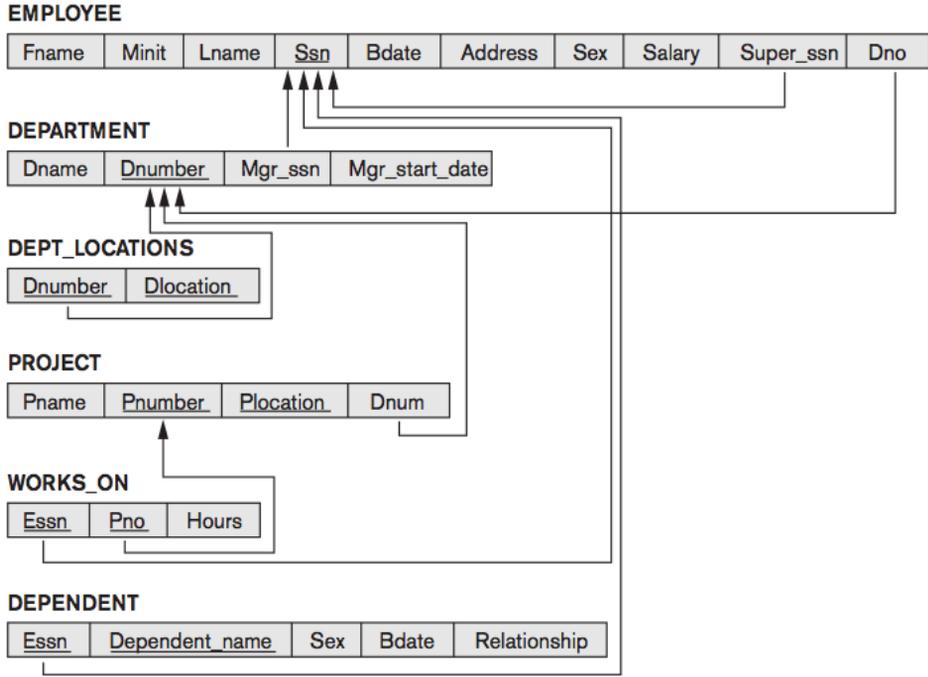
Schema Example



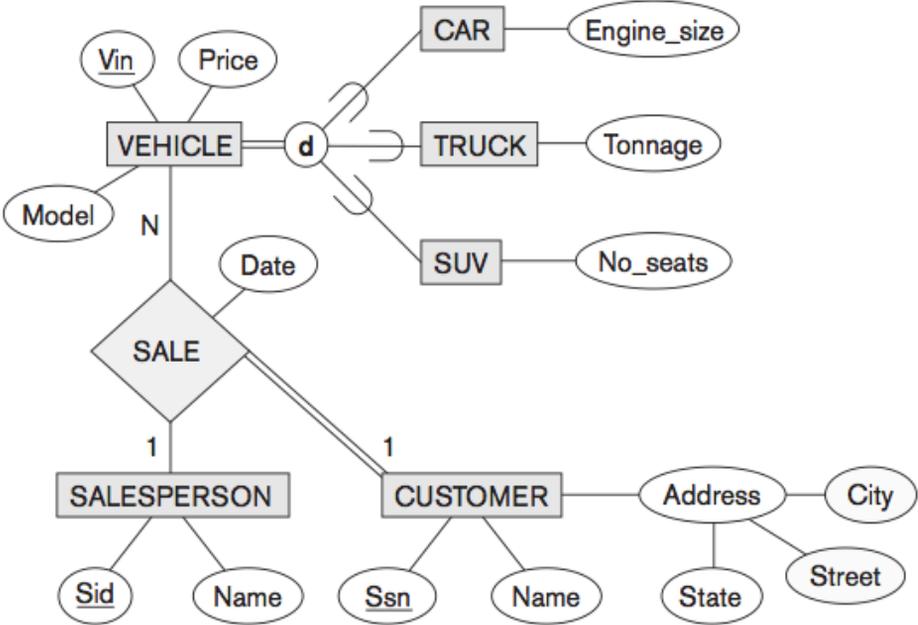
ER Model (example)



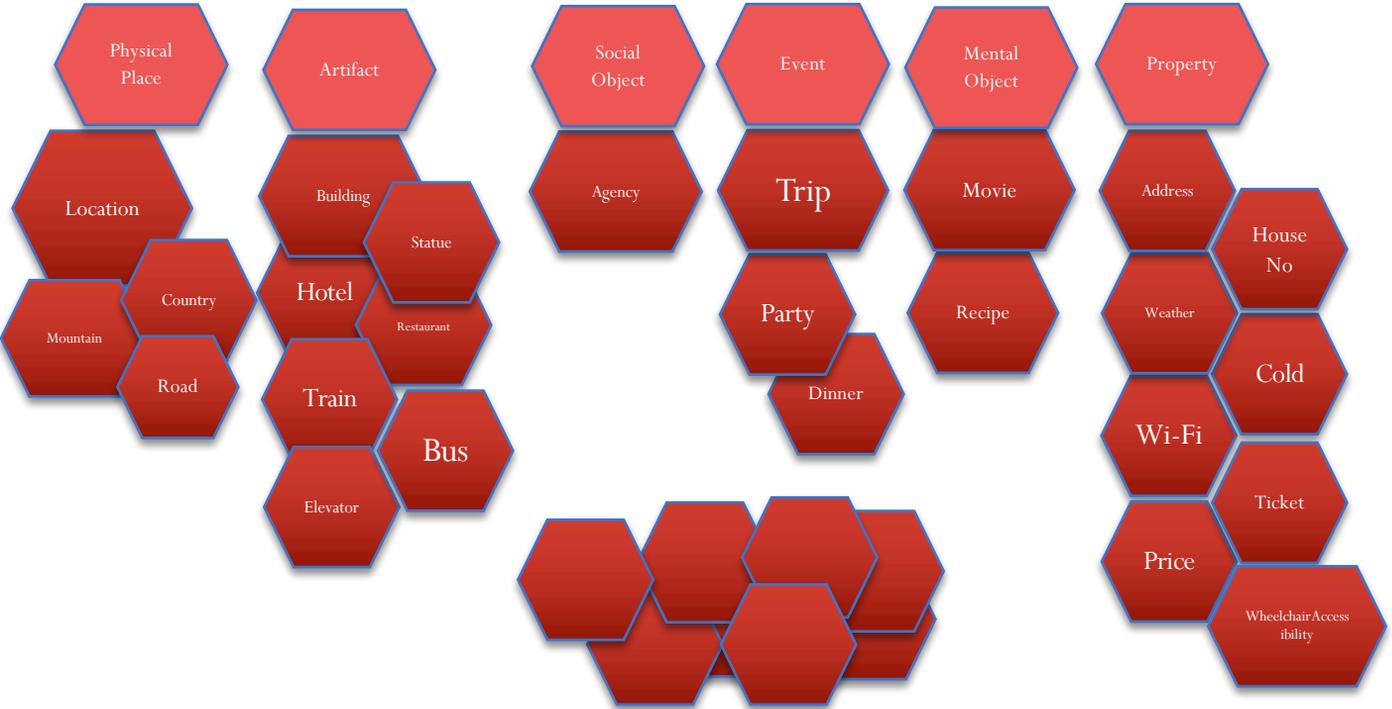
ER Model and Relational Database (example)



EER Model (example)



Alignment with Upper Ontology and Classification



Active Ontology | Entities | Individuals by class | DL Query | Object Properties |

Class hierarchy: owl:Nothing

- owl:Thing
 - ConsumableThing
 - DomainConcept
 - Fruit
 - NonConsumableThing
 - Region
 - ValuePartition
 - Vintage
 - VintageYear
 - Wine \equiv wine
 - WineDescriptor
 - Winery

Instances: owl:Nothing

Description: ChateauMargaux

Types: Margaux

Property assertions: ChateauMargaux

Object property assertions:

- hasMaker ChateauMargauxWinery
- locatedIn FrenchRegion
- locatedIn MargauxRegion
- locatedIn BordeauxRegion
- locatedIn MedocRegion

● building

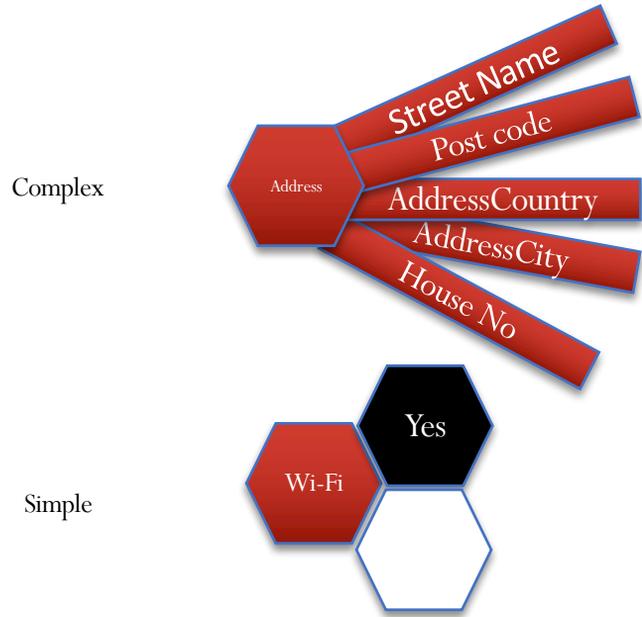
- agriculturalBuilding \equiv farmBuilding
- commercialBuilding
- educationalBuilding
- factory \equiv productionBuilding \equiv industrialBuilding
- farmBuilding \equiv agriculturalBuilding
- governmentBuilding
- healthcareBuilding \equiv healthcareFacility
- healthcareFacility \equiv healthcareBuilding
- industrialBuilding \equiv productionBuilding \equiv factory
- placeOfWorship \equiv religiousBuilding
- productionBuilding \equiv factory \equiv industrialBuilding
- religiousBuilding \equiv placeOfWorship
- residentialBuilding
- transportationBuilding



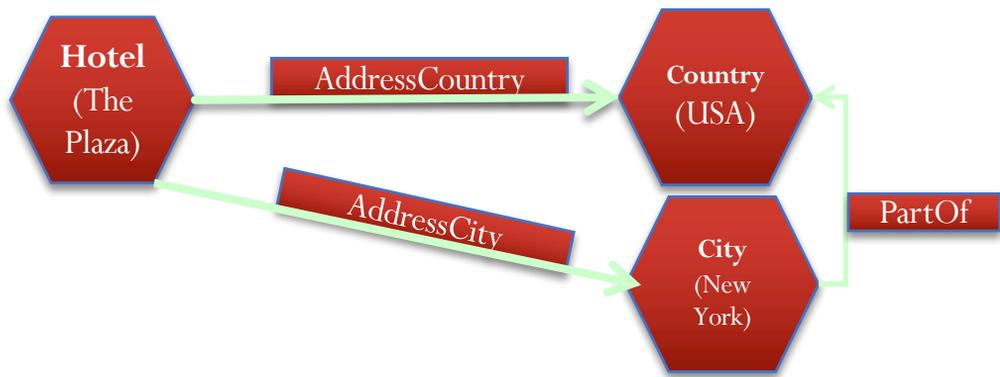
● administrativeDivision

- borough
- city
- country \equiv self_governing
- county
- federalDistrict \equiv unionTerritory
- prefecture
- region
- state
- town
- unionTerritory \equiv federalDistrict
- ward

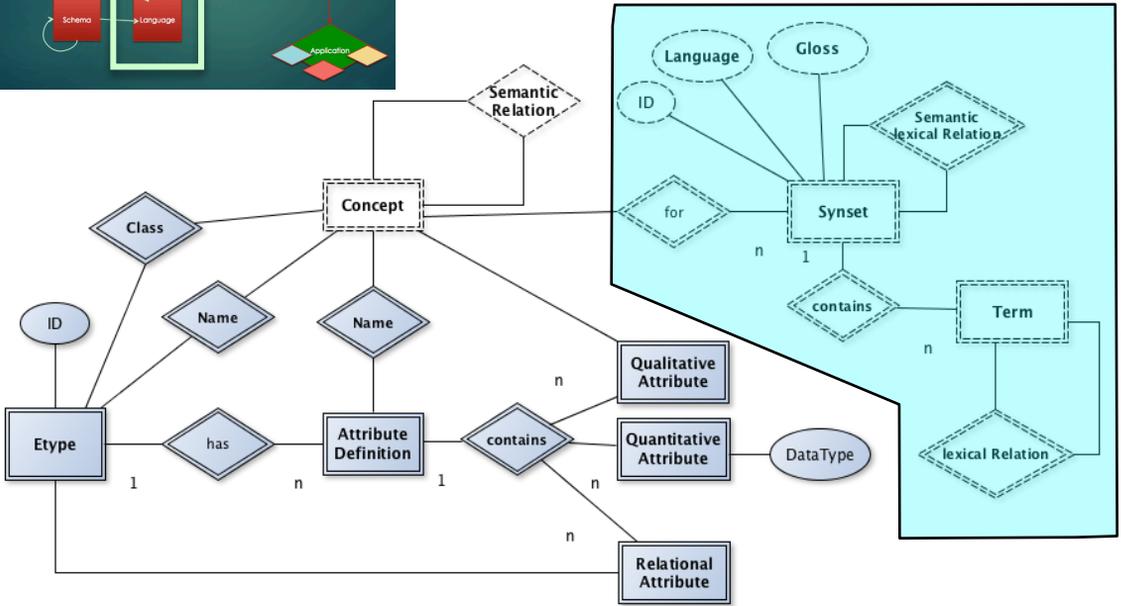
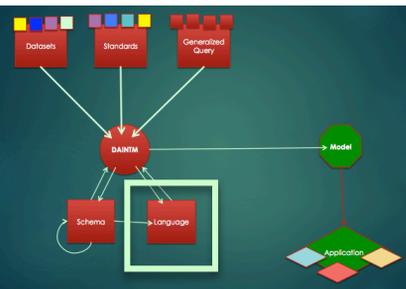
Issue_1:Attributes and DataProperties

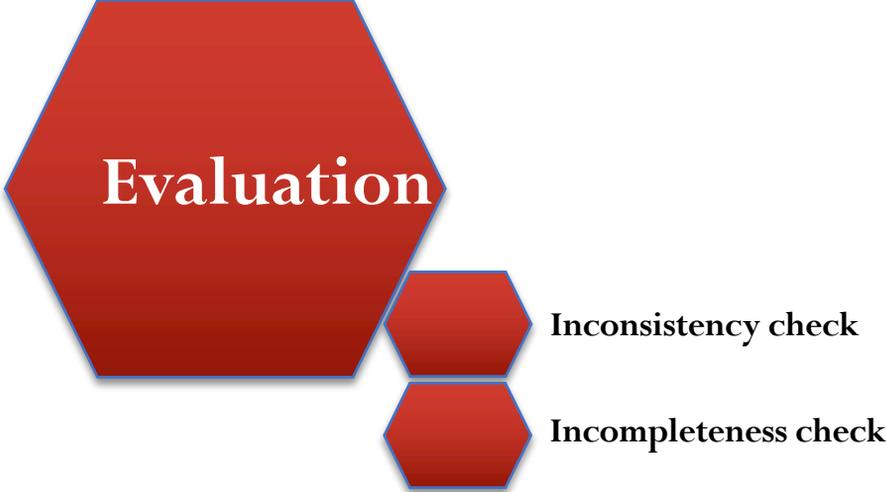


Issue_2: Relation and ObjectProperties



Language Level



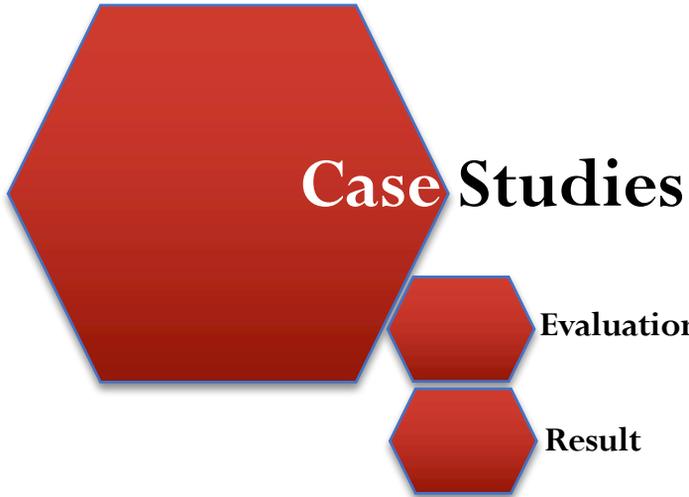


Evaluation

Inconsistency check

Incompleteness check

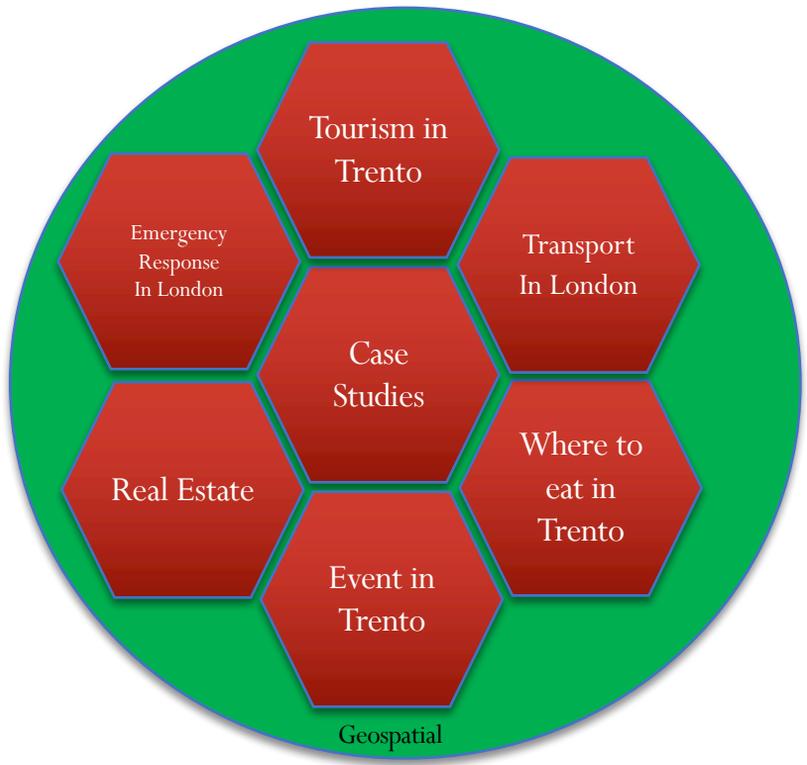
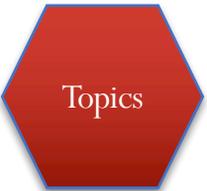
- **Inconsistency**
 - **circularity errors:** [ex. *Traveler subclassOf Person; Person subclassOfTraveler;*]
 - **semantic inconsistency errors:** [ex. *Airbus or Waterbus subclassOf Bus*]
 - **partition errors:** [ex. *Non stop Flight SubClassOf InternationalFlight and DomesticFlight* where *International* and *Domestic flight* are disjoint]
- **Incompleteness:** On traveling domain, if we classify only beach and mountain location, and we do not consider cultural heritage site
- **Redundancy**
 - **Identical formal definition of some class**
 - **Identical formal definition of instances**



Evaluation of Methodology

Result

Case Studies (example)



- **Technique**
 - **Used standard Human Computer Interaction (HCI) technique**
 - **Open Ended questions mixed with Likert scale closed questions**
- **How: Balanced Questioners**
- **Number of participant: 18**
- **Participants Information**
 - **Nationality: Italian, Indian, Germany, Brazil, Ukraine, Ethiopia, Mexico, Uganda, Cameroon**
 - **Gender: Male 13 Female 5**
 - **Age Range: 18-25 (14), 26-30 (4)**
 - **Level of education: Undergraduate (3) Postgraduate (15)**

Perspicuity: How easy it is to get familiar with the methodology

Efficiency: How effectively user can perform the process

Dependability: Can user control the process

Stimulation: Is it exciting and motivating

Novelty: Is it innovative and creative

Pros

- Well Structured
- programmatically durable
- It practically allows describe the world
- Provides methods to minimize the distance between the real world and the abstraction
- Helps finding out eventual defects of the ontology and helps correcting them :
taxonomic errors,
inconsistencies, reliability

Cons

- You need many practice to build something very well
- Needs more time to master
- difficult to identify class for to align with top level
- Necessary to write documentation to clarify choices and terms
- Formalizing DERA to DL

- ❑ Data on the Web Best Practices W3C Recommendation 31 January 2017
<https://www.w3.org/TR/dwbp/>
- ❑ Das, S., & Giunchiglia, F. (2016, October). GeoEtypes: Harmonizing Diversity in Geospatial Data (Short Paper). In *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"* (pp. 643-653). Springer International Publishing.
- ❑ Hlomani, H., & Stacey, D. (2014). Approaches, methods, metrics, measures, and subjectivity in ontology evaluation: A survey. *Semantic Web Journal*, 1-5.
- ❑ Giunchiglia, F., & Dutta, B. (2011). DERA: A FACETED KNOWLEDGE ORGANIZATION FRAMEWORK.
- ❑ Guarino, N., & Welty, C. A. (2009). An overview of OntoClean. In *Handbook on ontologies* (pp. 201-220). Springer Berlin Heidelberg.
- ❑ Gomez-Perez, A., Fernández-López, M., & Corcho, O. (2006). *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. Springer Science & Business Media.