KDI : **Knowledge and Data Integration**

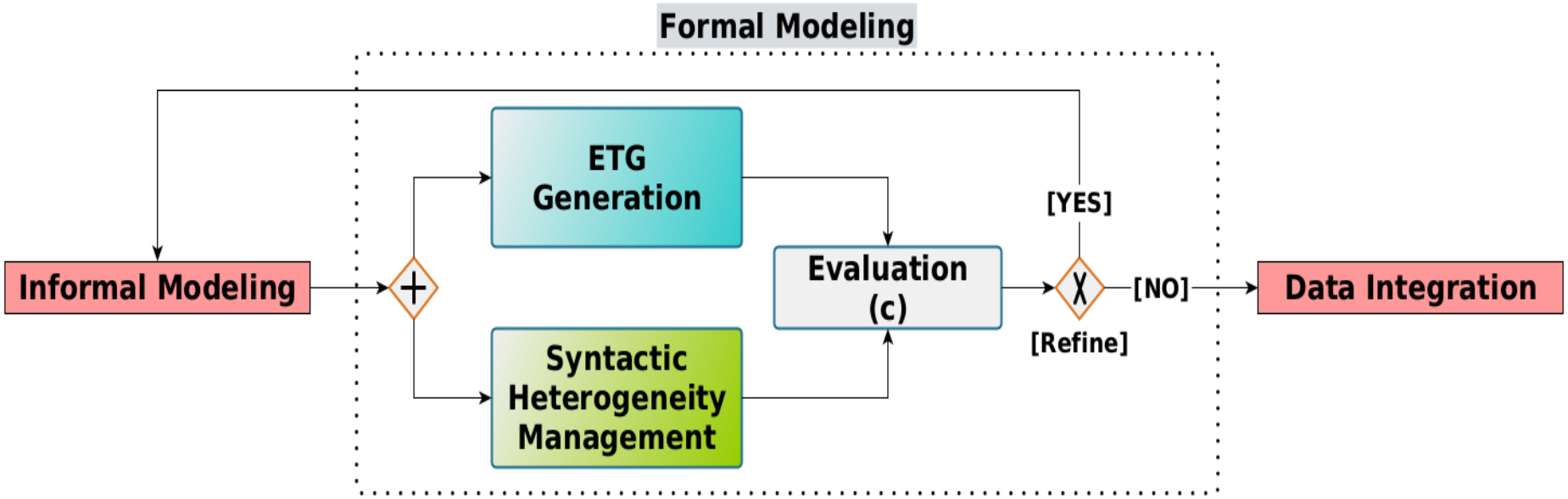
Ontology Selection

iTelos Formal Modeling Phase

Fausto Giunchiglia, Mayukh Bagchi

Contents

Formal Modeling phase



Formal Modeling objective

Formal Modeling is the third iTelos phase

Inputs:

- ER model.
- Selected datasets.
- Reference ontologies.

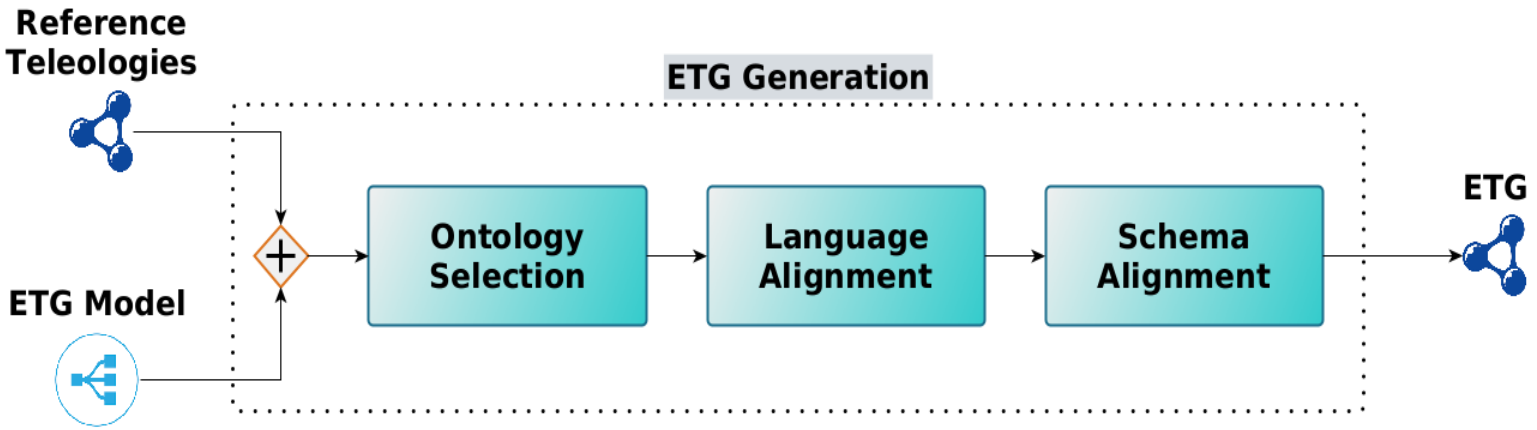
Outputs:

- ETG.
- Dataset syntactically aligned.

The Formal Modeling phase aims to:

- **Knowledge layer**: generate the ETG as shareable as possible reusing the reference ontologies, and as much as possible aligned with the ETG model.
- **Data layer**: handle the syntactic heterogeneity within the datasets.

ETG Generation Activity



The ETG Generation activity is internally defined by three sub activities:

- **Ontology Selection**: selection of those ontologies which includes appropriate concepts which can be reused to model the ER.
- **Language Alignment**: identification, and import in Knowledge Base (see ETG generation in practice), of concepts and terms to be used to build the ETG.
- **Schema Building**: composition of the ER schema designed with the foundational teleology.

Contents

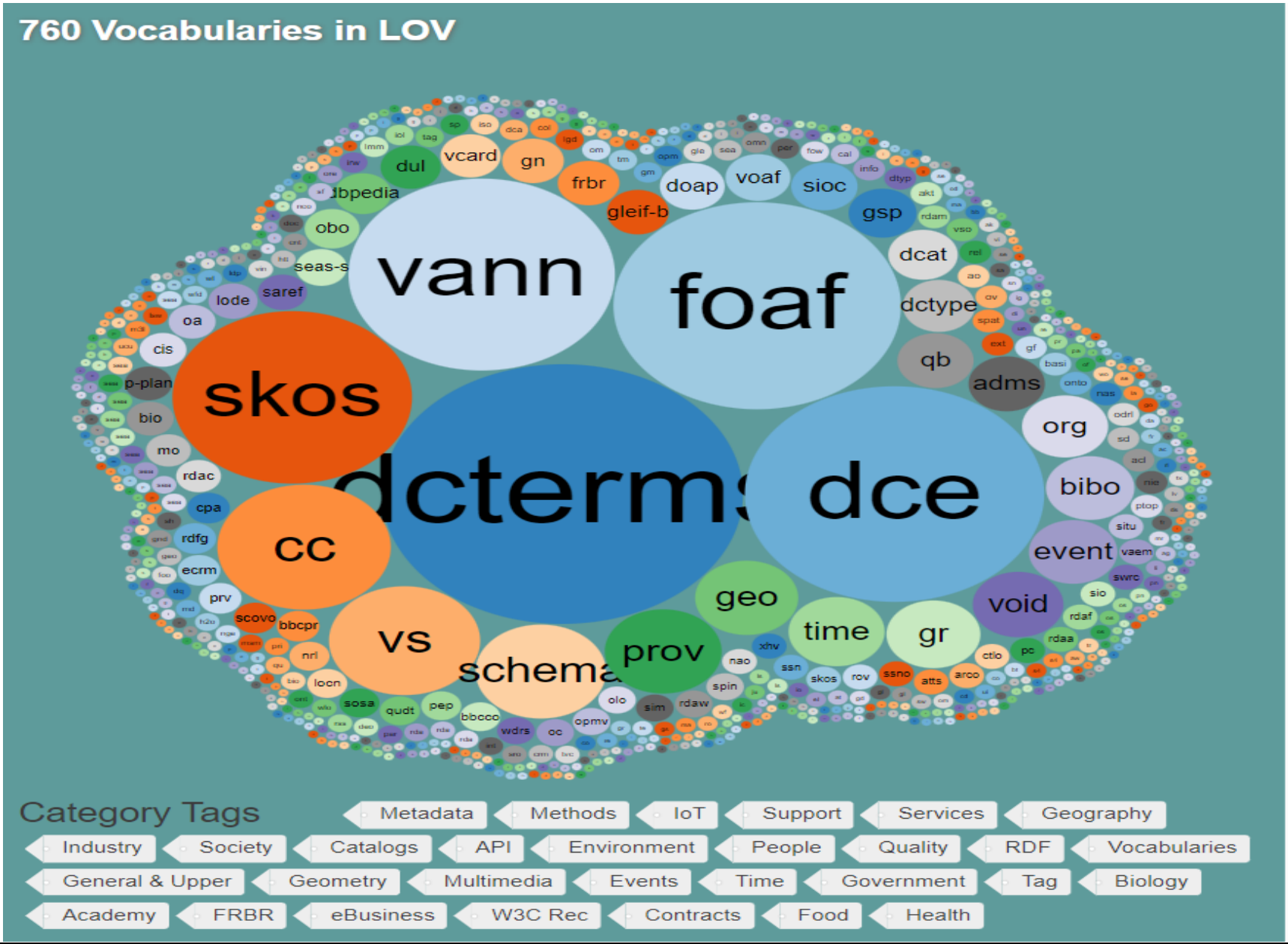
Motivation

- In the informal modelling phase, we learnt about the methodology for developing the ER Model in accordance to the theory of teleology
- Ontology Selection, as the first step of *ETG generation* activity, is focused on *reusing ontology elements* (classes, object properties, data properties) from state-of-the-art ontologies which are *semantically synonymous* with concepts which model the ER model
- The key observation is the fact that the *reused concepts* are *recast and aligned* in terms of - objects, functions and actions - explicit in the ER, something which is absent in the source ontology from where they are reused

Motivation (Contd.)

- The *recast and reuse* can be from amongst any state-of-the-art *general purpose ontology*, *domain-specific ontology* or *application-specific ontology* found in ontology repositories such as LiveSchema, LOV etc.
- For example, for common concepts like space and time, schemas like GTFS (for space) and iCal, W3C Time (for time) can be *reused*
- Terms from *general purpose ontologies* such as *schema.org* and *DBpedia* can also be suitably reused as required
- Ontology Selection is *key* to the two unique feature of *iTelos* - *reuse* and *shareability* (facilitated by metadata)

Linked Open Vocabulary (LOV)



BioPortal

Welcome to BioPortal, the world's most comprehensive repository of biomedical ontologies

Search for a class

Enter a class, e.g. Melanoma

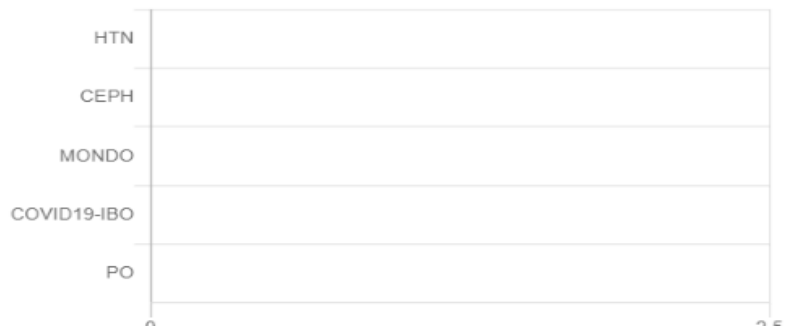
Advanced Search

Find an ontology

Start typing ontology name, then choos

Browse Ontologies

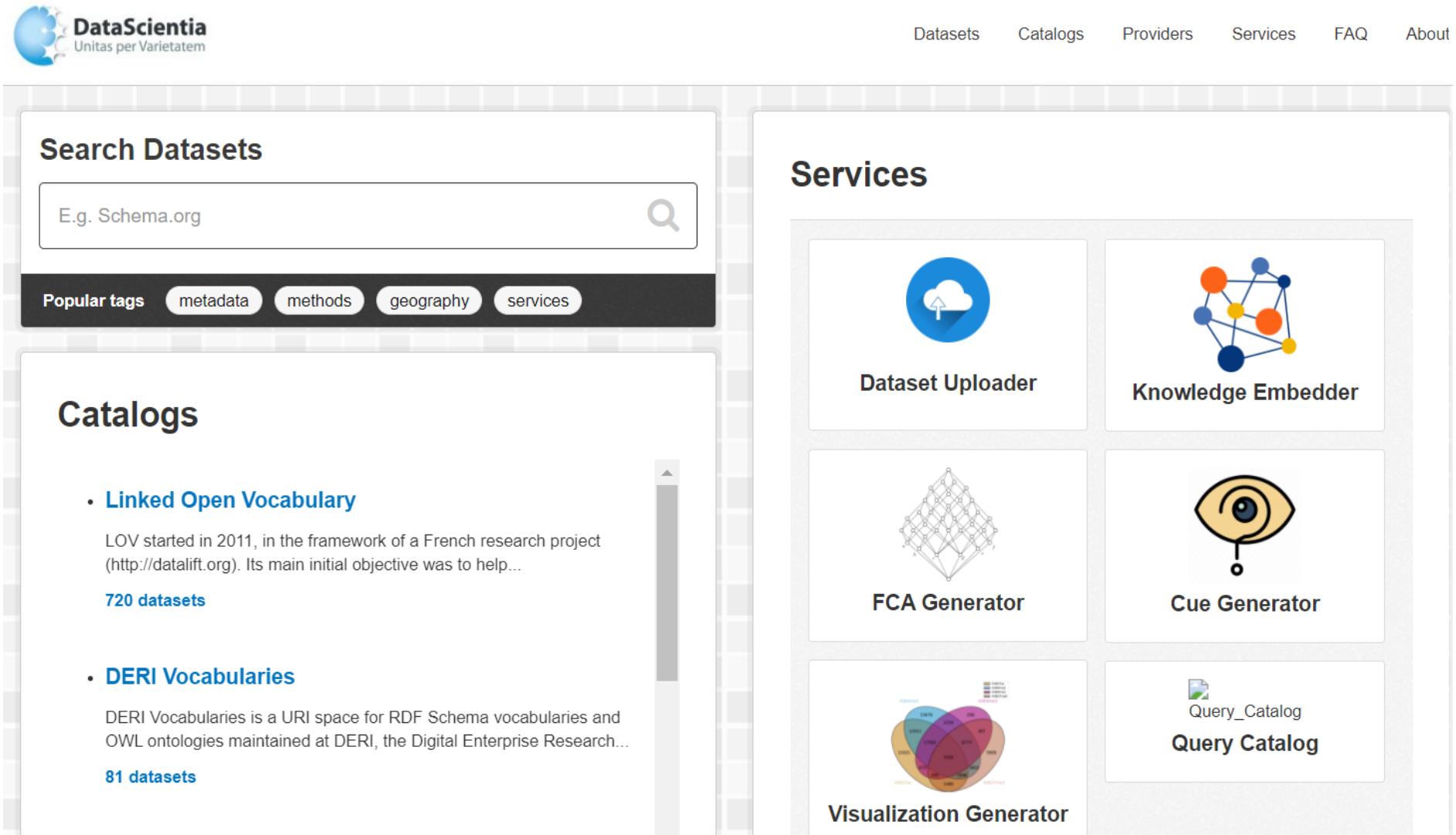
Ontology Visits (October 2021)



BioPortal Statistics

Ontologies	934
Classes	13,395,455
Properties	36,286
Mappings	55,648,584

LiveSchema


DataScientia
Unitas per Varietatem
Datasets
Catalogs
Providers
Services
FAQ
About
Search Datasets
E.g. Schema.org
Popular tags
metadata
methods
geography
services
Catalogs
Linked Open Vocabulary
LOV started in 2011, in the framework of a French research project (http://datalift.org). Its main initial objective was to help...
720 datasets
DERI Vocabularies
DERI Vocabularies is a URI space for RDF Schema vocabularies and OWL ontologies maintained at DERI, the Digital Enterprise Research...
81 datasets
Services
Dataset Uploader
Knowledge Embedder
FCA Generator
Cue Generator
Visualization Generator
Query_Catalog
Query Catalog

Approach

We follow the ***Common*** $\rightarrow$ ***Core*** $\rightarrow$ ***Contextual*** ordered classification of concepts throughout the following ontology selection activity -

1. Firstly, we consider the potential concepts (objects, functions, actions, data properties, object properties) in the common category from existing ontologies, which can be reused
2. Secondly, we consider the concepts from the core category which can be reused
3. Finally, we consider the concepts from the contextual category which can be reused

The essence of the *iTelos methodology* is to (ideally) achieve maximum reuse of concepts for the core, common and contextual categories, roughly in a 20:60:20 proportion

Contents

Example

The following concept hierarchy fragment is taken from *schema.org* -

- FoodEstablishment -
  - Bakery
  - BarOrPub
  - Brewery
  - CafeOrCoffeeShop
  - Distillery
  - FastFoodRestaurant
  - IceCreamShop
  - Restaurant
  - Winery

Example (Contd.)

- Reusing the above conceptual hierarchy (either fully or partially) depends upon the modelling requirements of the project.
- One approach can be to consider *FoodEstablishment* to be a *Producer - Consumer* of the object *Establishment* having several specialized functions such as *Bakery, BarOrPub, Brewery etc ...* in the context of the domain of food and accommodation as the *reference context*.
- Other (equally valid) approaches can be to utilize the hierarchy *as it is* (i.e. as an object hierarchy) if the modelling requirements (formalized in the ETG model) demands so.

Contents

Summary

- We learnt about the motivation behind the ontology selection activity and why it is key to iTelos methodology
- We understood the general approach to be followed for selecting and reusing concepts from ontologies to formally design the project teleology
- We illustrated via a small example of food establishment from schema.org, the possibilities which it offers for formal modelling
- THANK YOU !!!

Contents

Glossary

- **Ontology**: An ontology is a formal, explicit specification of a shared conceptualization
- **Teleology**: Teleologies are ontologies with the proviso that teleologies focus on function and on how a chosen representation fits a certain purpose, this being the basis for a general model for the diversity of knowledge
- **ER Model**: An entity–relationship model describes interrelated things of interest in a specific domain of knowledge. It is composed of entity types and specifies relationships that can exist between entities

Glossary (Contd.)

- **ETG**: An Entity Type Graph (ETG) is a fully formal, schema-level directed acyclic graph (DAG) which models concepts and interrelationships amongst such concepts for any domain of interest. Each concept in an ETG is *alinguistic* and is formally identified by a unique GID assigned in the UKC.

- **EG**: An Entity Graph (EG) is a data-level knowledge graph generated by populating an ETG with the entities extracted from the input datasets.

KDI : Knowledge and Data Integration

Fausto Giunchiglia, Mayukh Bagchi

Ontology Selection

iTelos Formal Modeling Phase