



Data diversity

The problem and hints of a solution

Fausto Giunchiglia, Mayukh Bagchi,
Alessio Zamboni, Simone Bocca

[KnowDive](#)

DISI, University of Trento, Italy.

23.05.2022

Representation!

To represent the *world*, we use:

- Words
- Sounds
- Pictures
-



Representation(words) = Concepts + Language + Knowledge + Data

Example: Starry Night is an oil on canvas painting by Vincent van Gogh.

Representation Diversity

We define *Representation Diversity* to mean (approx.) *semantic heterogeneity*, as organized in the four layers:

- *(L1) Conceptual diversity*
- *(L2) Language diversity*
- *(L4) Knowledge diversity*
- *(L5) Data diversity*

Motivating Example

Car				
Nameplate	schema: speed	schema: fuelCapacity	schema: fuelType	schema: modelDate
FP372MK	150	62	Petrol	2020-11-25

Vettura		
Targa	Velocità	Tipo di corpo
FP372MK	158	Coupé

Vehicle			
vso:VIN	vso:feature	vso:modelDate	vso:speed
FP372MK	Armrest	2020-11-25	155.0

Stratification of Diversity: *Conceptual Diversity (L1)*

Car				
Nameplate	schema: speed	schema: fuelCapacity	schema: fuelType	schema: modelDate
FP372MK	150	62	Petrol	2020-11-25

Vettura		
Targa	Velocità	Tipo di corpo
FP372MK	158	Coupé

Vehicle			
vso:VIN	vso:feature	vso:modelDate	vso:speed
FP372MK	Armrest	2020-11-25	155.0

Concepts organized in multiple hierarchies, in terms of '*hypernym-hyponym*' links

Eg:- Car or Vehicle ?

Stratification of Diversity: *Language Diversity (L2)*

Car				
Nameplate	schema: speed	schema: fuelCapacity	schema: fuelType	schema: modelDate
FP372MK	150	62	Petrol	2020-11-25

Vettura		
Targa	Velocità	Tipo di corpo
FP372MK	158	Coupé

Vehicle			
vso:VIN	vso:feature	vso:modelDate	vso:speed
FP372MK	Armrest	2020-11-25	155.0

- Multiple Languages, e.g., *natural languages, namespaces and formal languages*.
- Phenomena like *polysemy and synonymy* allow for diverse representations of entities, even within the same language
- *Many-to-many mapping between words and concepts*, both within the same language and across languages

Eg:- vso:VIN, Nameplate, Targa

Stratification of Diversity: *Knowledge Diversity (L4)*

Car				
Nameplate	schema: speed	schema: fuelCapacity	schema: fuelType	schema: modelDate
FP372MK	150	62	Petrol	2020-11-25

Vettura		
Targa	Velocità	Tipo di corpo
FP372MK	158	Coupé

Vehicle			
vso:VIN	vso:feature	vso:modelDate	vso:speed
FP372MK	Armrest	2020-11-25	155.0

Knowledge diversity arises from the *many-to-many mapping between etypes and the properties* employed to describe them,

Eg:- Same etype Car, but associated with different groups of properties.

Stratification of Diversity: *Data Diversity (L5)*

Car				
Nameplate	schema: speed	schema: fuelCapacity	schema: fuelType	schema: modelDate
FP372MK	150	62	Petrol	2020-11-25

Vettura		
Targa	Velocità	Tipo di corpo
FP372MK	158	Coupé

Vehicle			
vso:VIN	vso:feature	vso:modelDate	vso:speed
FP372MK	Armrest	2020-11-25	155.0

Data diversity arises because *the mapping between entities and the property values used to describe them is many-to-many.*

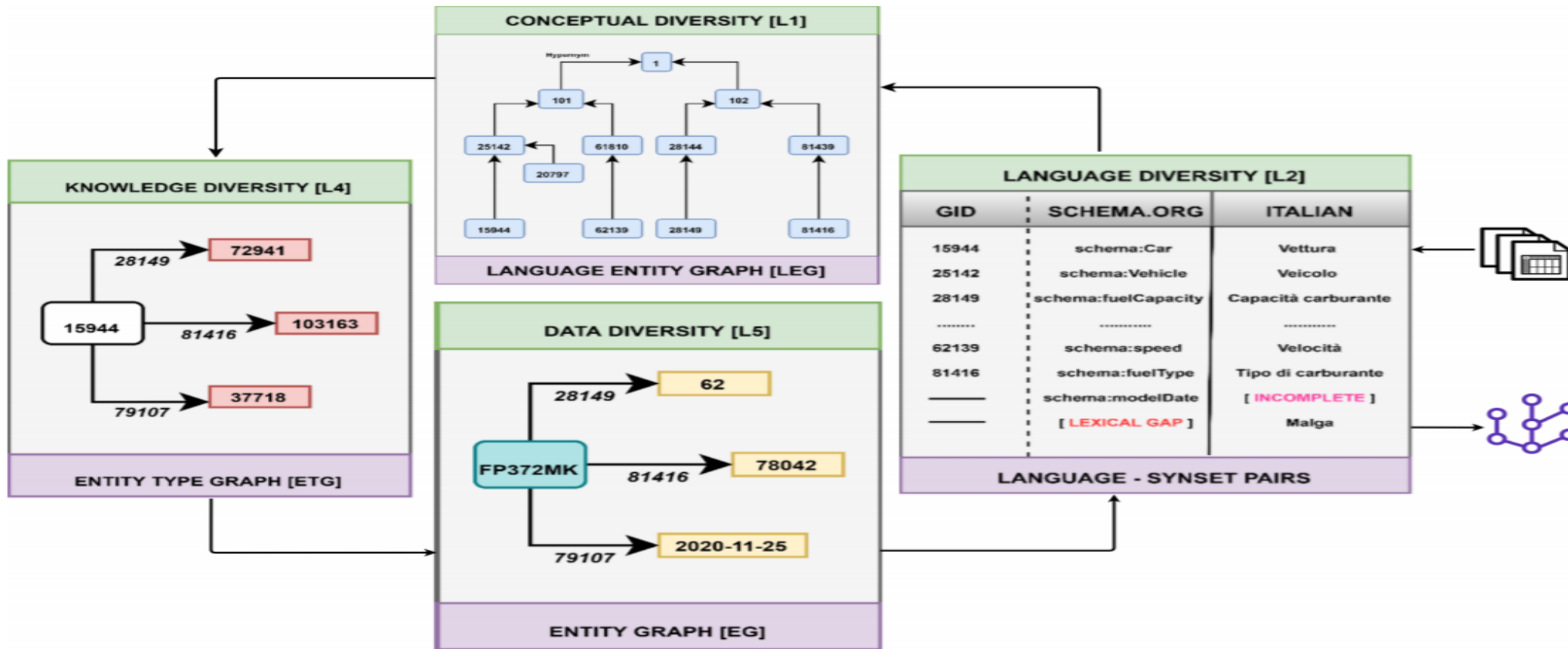
Eg:- Same entity car 'FP372MK' but with different velocities

Stratification of Diversity - Contd.

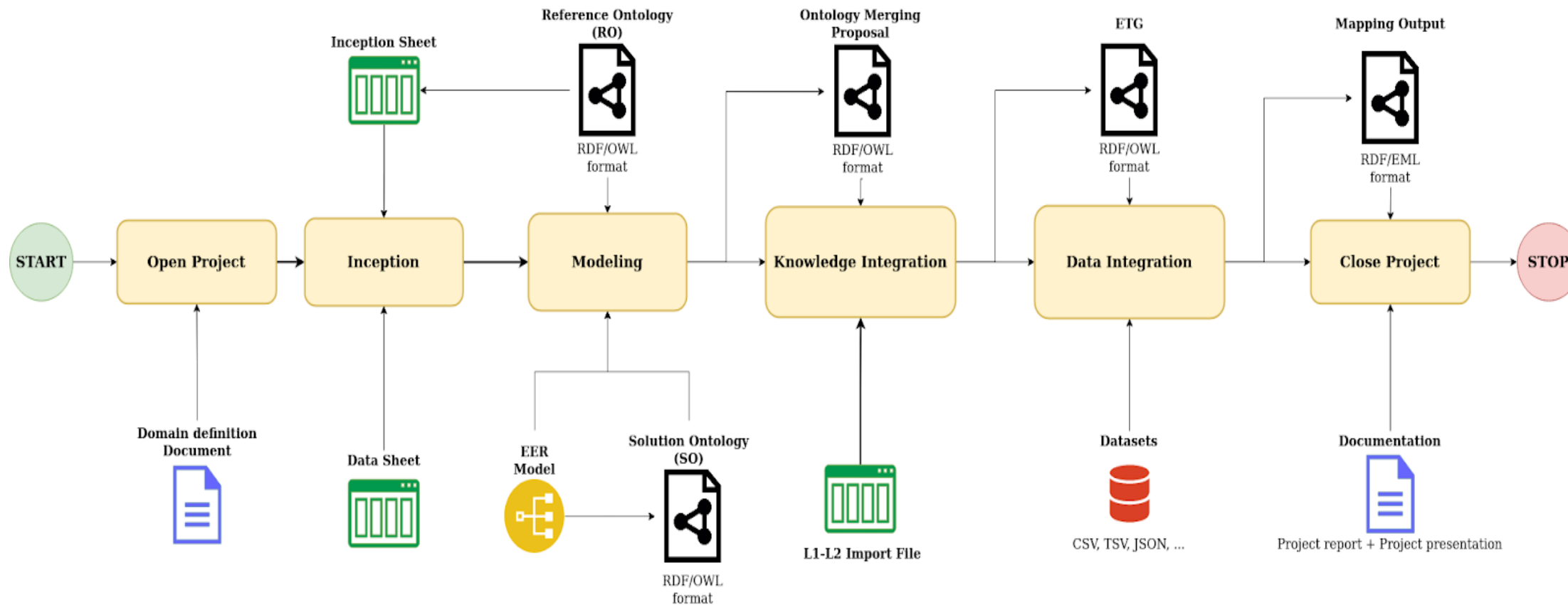
The stratification of diversity presented has the following *crucial characteristics*:

- each layer models a different type of phenomenon and the corresponding type of diversity.
- diversity which appears in one layer is *independent* of how it appears in any other layer

The Representation Diversity Architecture



iTelos: Stratified Data Integration Methodology



Note : An evaluation step is considered within the execution of each activity. If the evaluation is not acceptable, the process will return to the previous activity.



UNIVERSITY
OF TRENTO - Italy

Dipartimento di Ingegneria e Scienza dell'Informazione



Thank You!

Fausto Giunchiglia, Mayukh Bagchi,
Alessio Zamboni, Simone Bocca

[KnowDive](#)

DISI, University of Trento, Italy.

23.05.2022

Online material

KDI (next year: KGE) Course portal: <http://knowdive.disi.unitn.it/teaching/kdi/>

LiveSchema portal: <http://liveschema.eu/>

DataScientia Portal (under development): <http://datascientia.disi.unitn.it/>

Stratified Data Integration paper: <https://arxiv.org/abs/2105.09432>

LiveSchema Demo: https://drive.google.com/file/d/16yZr2Tla1FJZZ6-JVKYRYcK0_6w6Hx3a/view?usp=sharing