

Enabling Data Analytics from Knowledge Graphs

Henrique Santos

SCIENTIFIC DATA COLLECTION



- LACK OF METADATA, NOT CLEAN, CAN'T BE DIRECTLY COMPARED OR COMBINED
- DATA PREPARATION TAKES AROUND 80% OF WHOLE ANALYTICAL PROCESS (PATIL, 2012)
- HOW TO PROVIDE ACCESS TO SCIENTIFIC DATA POINTS IN KGs?
- HOW TO EXPLOIT KNOWLEDGE TO PROVIDE ACCESS TO THE NEEDED SUBSET OF DATA?



SCIENTIFIC DATA ANNOTATION



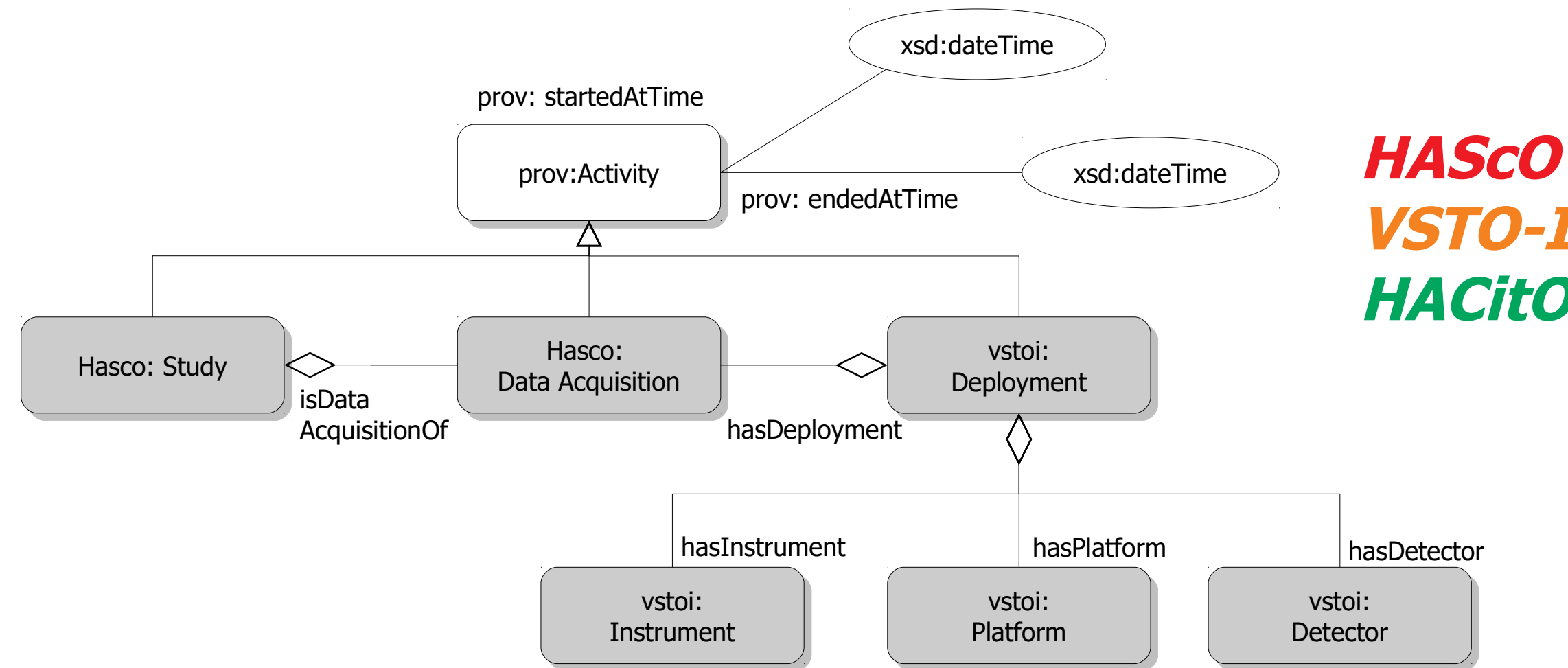
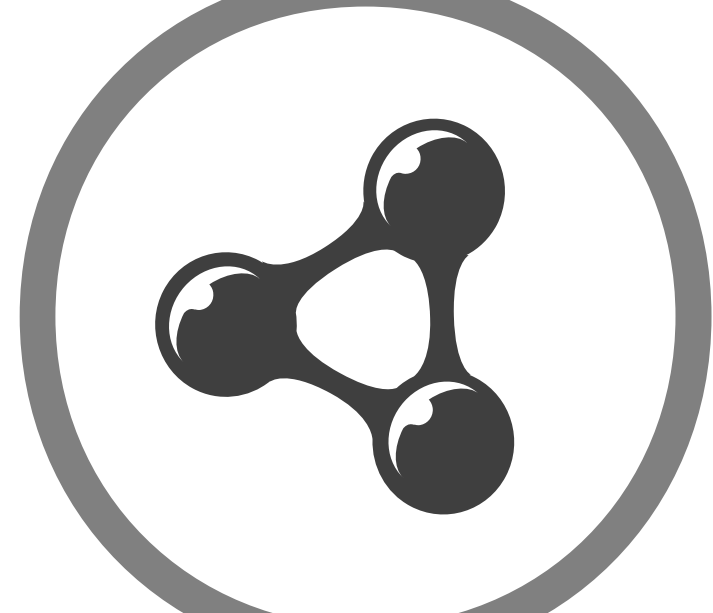
- CONTEXTUALIZED COMMA SEPARATED VALUES
- CONNECTION BETWEEN CSV DATA AND KNOWLEDGE IN SCIENTIFIC KNOWLEDGE GRAPHS

```
<data-collection>
a hasneto:DataCollection;
a time:Interval;
prov:startedAtTime [some.timestamp]^xsd:dateTime .

<dataset>
a vstoi:Dataset;
prov:wasGeneratedBy <data-collection>;
hasneto:hasMeasurementType <mt0> .

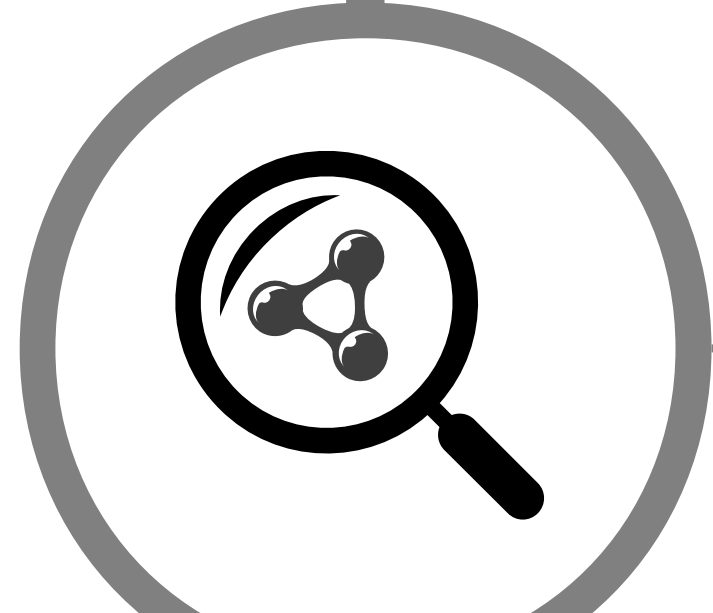
<mt0>
a oboe:Measurement;
a time:Instant;
time:inDateTime <ts0>;
ccsv:atColumn 1;
oboe:ofCharacteristic [some.characteristic];
oboe:usesStandard [some.standard] .
```

KNOWLEDGE GRAPH BUILDING



Field Facets	Facet 1	Facet 2
Company	<ul style="list-style-type: none"> Cearense (15174) Dragão do Mar (15174) Empresa Aliança (15174) Fortaleza (15174) Fretcar (15174) Maraponga (15174) Santa Cecília (15174) Santa Marta (15174) Silva Grande (15174) São José (15174) Terra Luz (15174) Vega (15174) Vilaça Urbana (15174) missing (0) 	<p>Id: d3p8tb9ib98pzddevsy5endoiv5fsqk69z3</p> <p>PF_codigo_empresa: 35</p> <p>PF_nome_empresa: Vega</p> <p>PF_codigo_onibus: 43343</p> <p>timestamp: Sun Feb 01 07:47:28 UTC 2015</p> <p>value: 1</p> <p>unit: Binary</p> <p>entity: Bus</p> <p>characteristic: E</p> <p>PF_codigo_ponto: 98</p> <p>PF_nome_ponto: Estrada do Jabobá/G Pereira/P Barbosa</p> <p>location: -3.80272300,-38.62784400</p> <p>_version_: 1500540294601375744</p> <p>score: 1.0</p>
Bus	<ul style="list-style-type: none"> 43007 (3653) 43010 (3653) 43015 (3653) 43017 (3653) 43020 (3653) 	<p>Id: h128gp39f9o5c2tjdj68s7d9xrg44273o3w</p> <p>PF_codigo_empresa: 35</p> <p>PF_nome_empresa: Vega</p> <p>PF_codigo_onibus: 43343</p> <p>timestamp: Sun Feb 01 06:41:08 UTC 2015</p> <p>value: 1</p> <p>unit: Binary</p> <p>entity: Bus</p> <p>characteristic: E</p> <p>PF_codigo_ponto: 98</p> <p>PF_nome_ponto: Estrada do Jabobá/G Pereira/P Barbosa</p> <p>location: -3.80272300,-38.62784400</p> <p>_version_: 1500540294601375745</p> <p>score: 1.0</p>
Checkpoint	<ul style="list-style-type: none"> Osório de Paiva, Gal./M Lobato A. Malveira (42120) Adalberto Malveira/Bamoso/Jari (32994) Pitangueiras/O de Paiva/E Mendes (23688) Estrada do Jabobá/G Pereira/P. Barbosa (21060) Siqueira (14040) missing (63180) 	

KNOWLEDGE GRAPH BROWSING

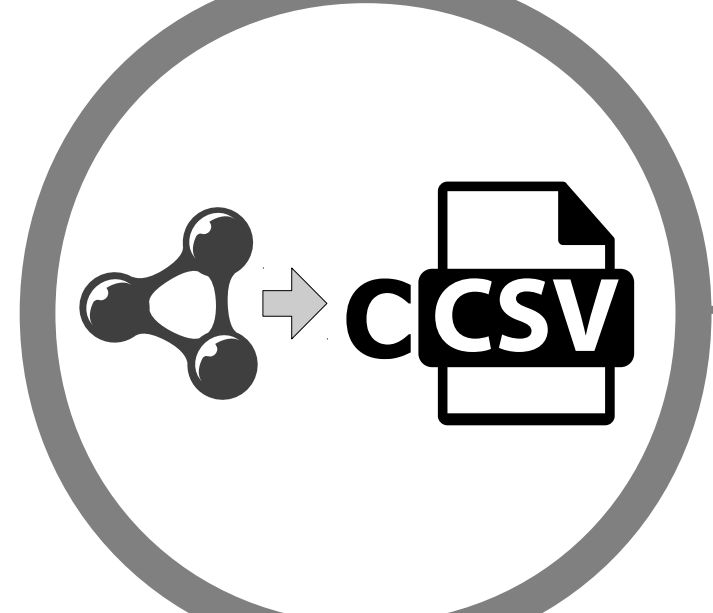


- FACETED-SEARCH PROVIDES A USER-FRIENDLY INTERFACE FOR BROWSING THE DATA IN THE KG
- FACETS ARE DEFINED BY SCIENTIFIC CONCEPTS AND DOMAIN ENTITIES IN THE ONTOLOGIES
- APACHE SOLR IS USED TO PROVIDE FACETING FEATURES

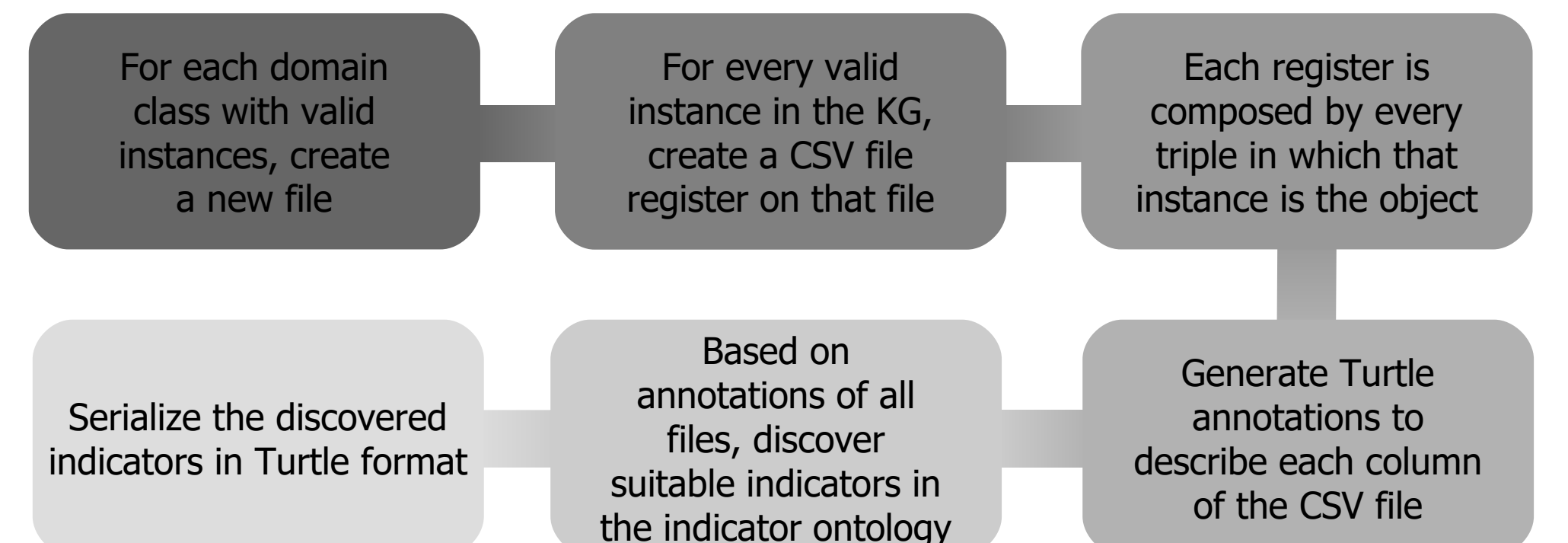
SANTOS, H.; FURTADO, V.; PINHEIRO, P.; MCGUINNESS, D. L. Contextual Data Collection for Smart Cities. In: *Proceedings of the Sixth Workshop on Semantics for Smarter Cities*. Bethlehem, PA, USA. 2015.



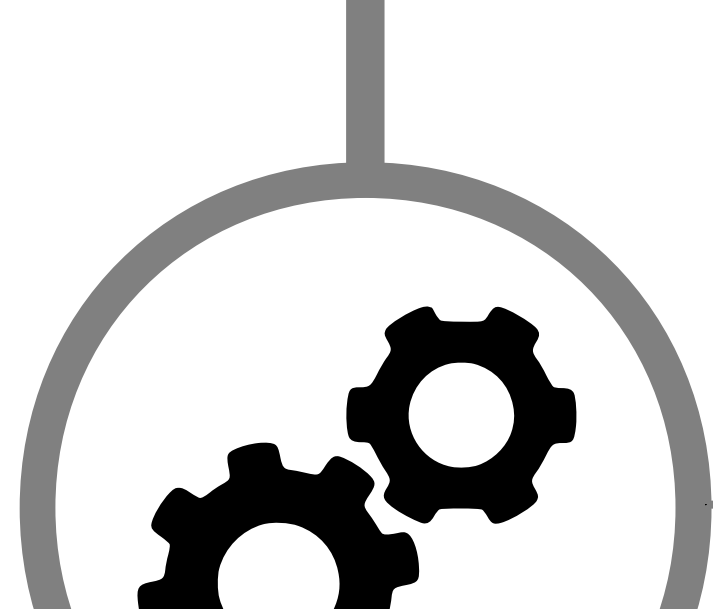
KNOWLEDGE GRAPH SERIALIZATION



- GENERATES DATASETS FROM SUBSETS OF THE KG
- EASILY COPE DATA FROM KGs WITH ROUTINE DATA TOOLS (R, PYTHON, BI)

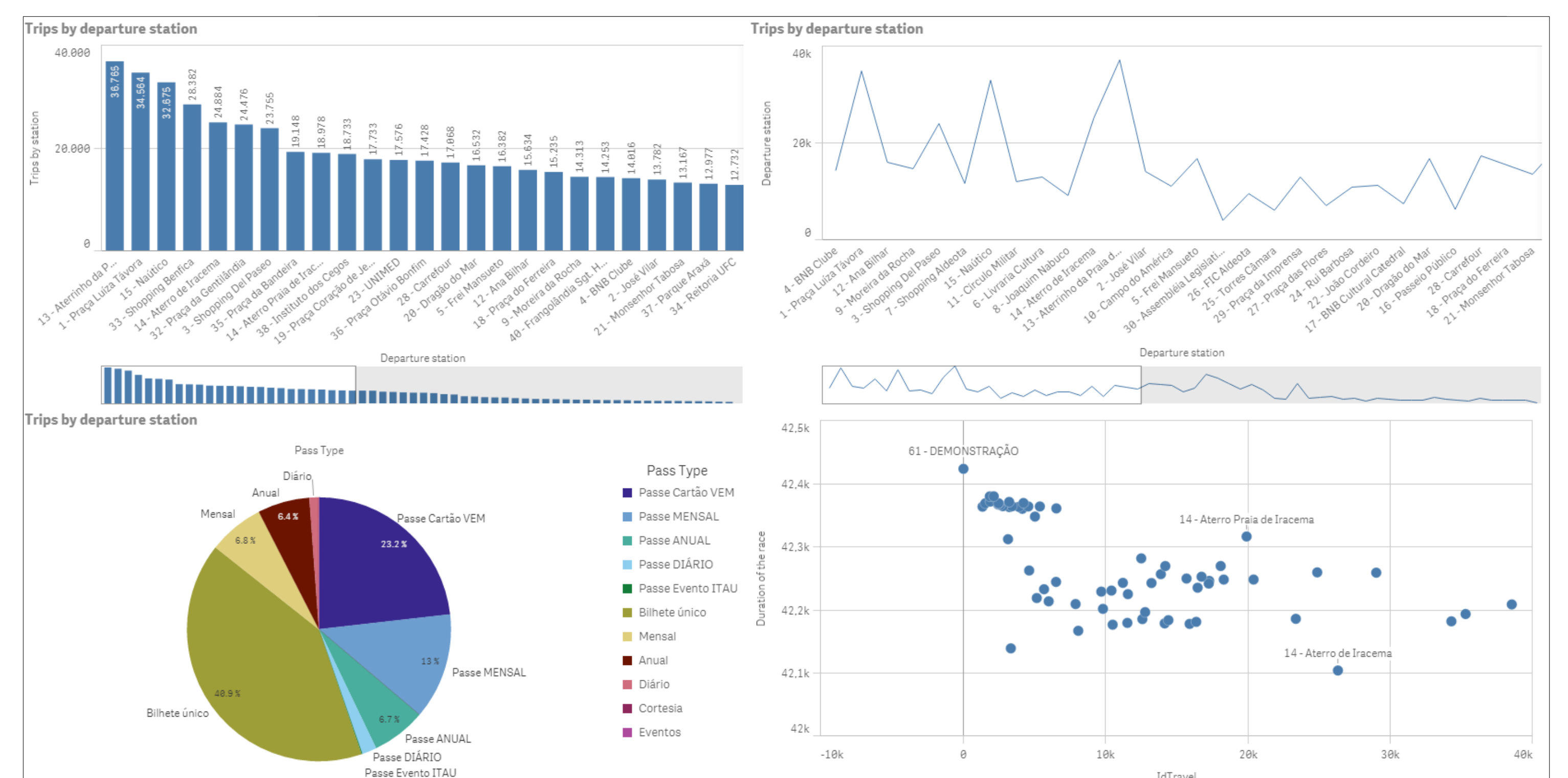


INTELLIGENT APPLICATIONS

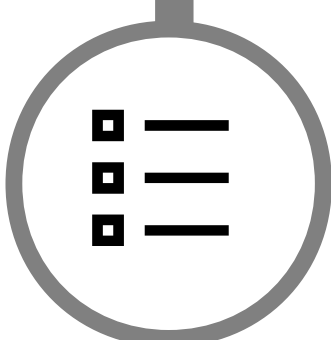


- ATTACHED KNOWLEDGE IS USED TO FOSTER DATA ANALYSIS

SANTOS, H.; DANTAS, V.; FURTADO, V.; PINHEIRO, P.; MCGUINNESS, D. L. From Data to City Indicators: A Knowledge Graph for Supporting Automatic Generation of Dashboards. In: *The Semantic Web - Proceedings of the 14th Extended Semantic Web Conference (ESWC 2017)*. Portorož, Slovenia. 2017.

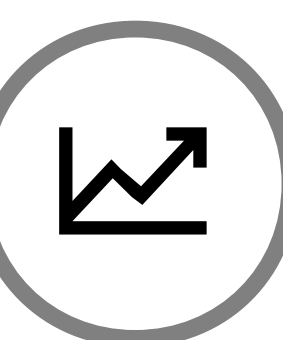


ROADMAP



- Expand indicators to more complex calculations (procedures and algorithms)
- Prepare data for analysis based on queries

EVALUATION



- State of the art KG evaluation approaches will be used to evaluate our KG
- Work together with data scientists to build a KG using their data and provide tools for querying and preparing datasets for their data analytics tasks