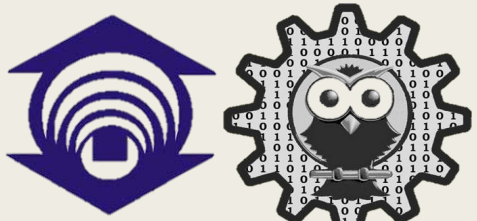

# FROM DATA TO CITY INDICATORS: A KNOWLEDGE GRAPH FOR SUPPORTING AUTOMATIC GENERATION OF DASHBOARDS

Henrique Santos[1], Victor Dantas[1], Vasco Furtado[1], Paulo Pinheiro[2] and Deborah L. McGuinness[2]

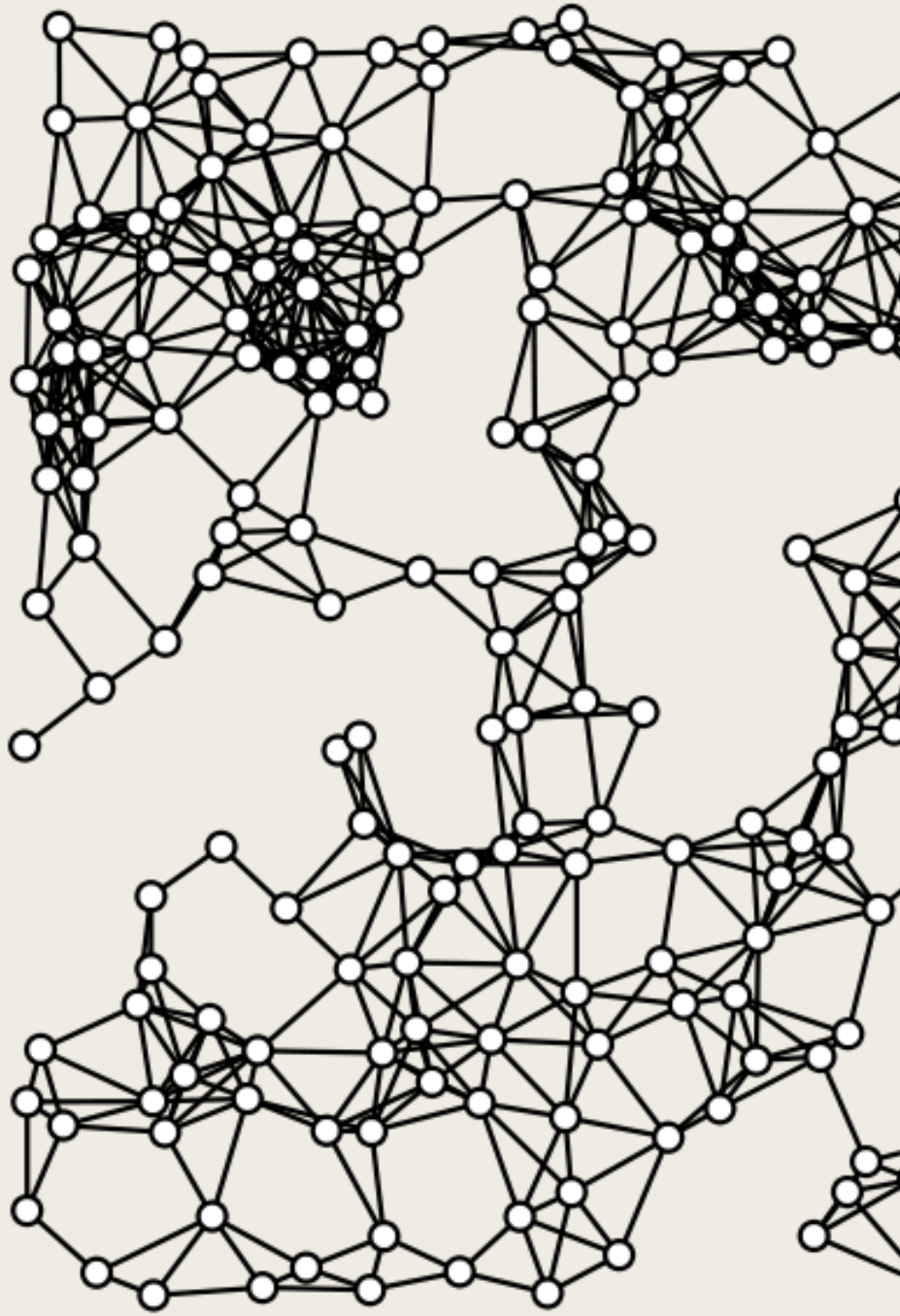
[1]Universidade de Fortaleza, Fortaleza, CE, Brazil
[2]Rensselaer Polytechnic Institute, Troy, NY, USA



14th Extended Semantic Web Conference (ESWC 2017)
Portorož, Slovenia, 31 May 2017

# Agenda

- Motivation
- City indicators
- City Knowledge Graph
  - *Metadata ontologies*
  - *Domain and Indicator ontologies*
- Use-case on BICICLETAR (Fortaleza bicycle-sharing system)
  - *Data pipeline*
  - *SBIG (Semantic BI Generator)*
- Conclusions and ongoing work



# Assessing city performance

- Why is it important to measure city performance?
  - *Compare to other cities*
  - *Better decision making*
  - *Better budget allocation*
- ISO 37120:2014 – Sustainable development of communities – Indicators for city services and quality of life
  - *Standardized set of indicators for measuring city performance*
- But how to calculate city indicators?
  - *"Get the needed data… but what data do I need?"*
  - *"I think I have the data… but I can't understand it"*
  - *High quality metadata is crucial when calculating trustworthy indicator values*

# City indicator requirements

01 Temporal coverage

02 Entities of interest

03 Provenance

04 Contextual knowledge

05 (Geo-)Location

06 Support easy visualization

# Metadata ontologies

**VSTO-I**

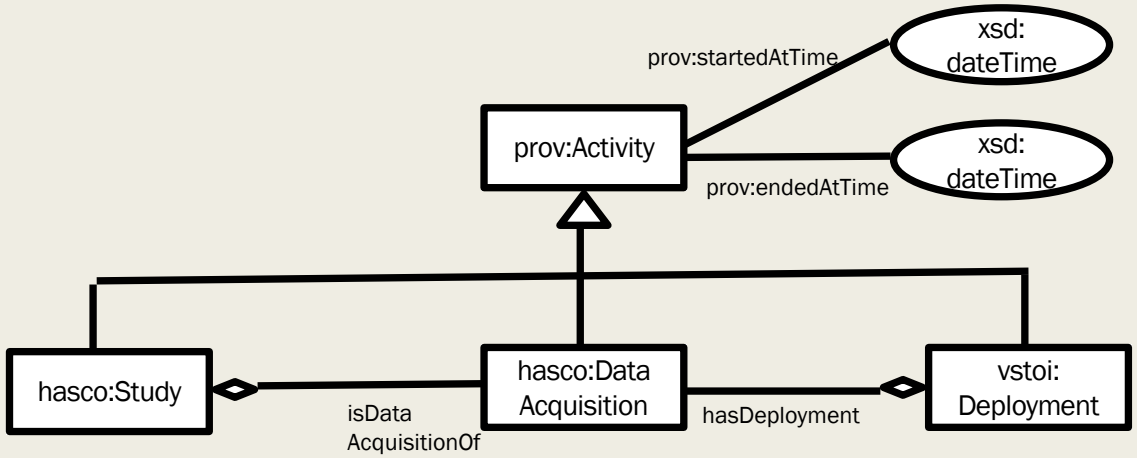
- Instruments and sensing devices
- Activities performed over instruments

**HAScO**

- Scientific activities
- Alignment of VSTO-I and PROV

**HACitO**

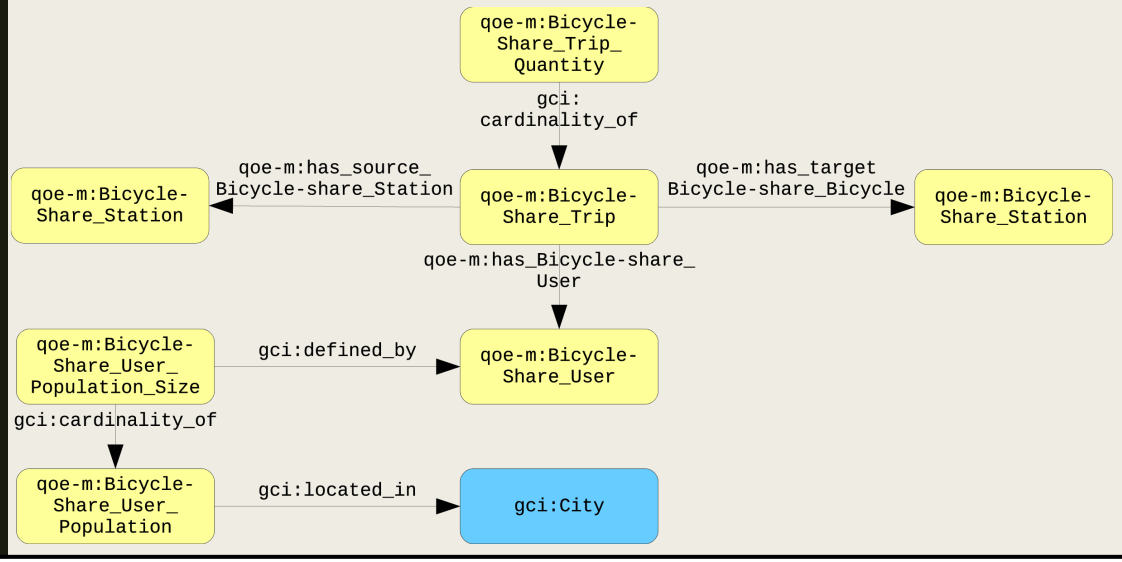
- Specialization of HAScO concepts for cities



# Domain and indicator ontologies

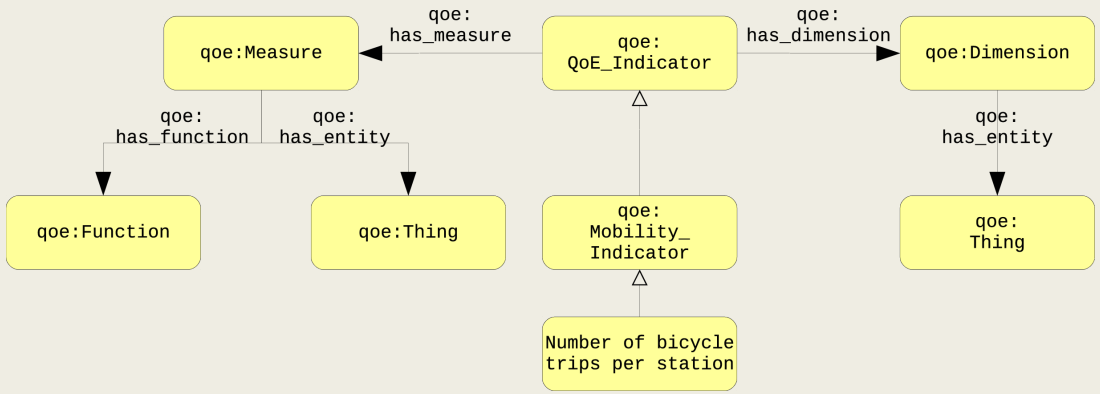
## QoE Domain

- Entities and relations of interest
- Quantifications




## QoE Indicators

- Relations/procedures between/over domain quantifications



FUNDAÇÃO EDSON QUEIROZ
UNIVERSIDADE DE FORTALEZA
ENSINANDO E APRENDENDO

# BICICLETAR: The bicycle-sharing system of Fortaleza, Brazil

Number of journeys performed

000159553 1

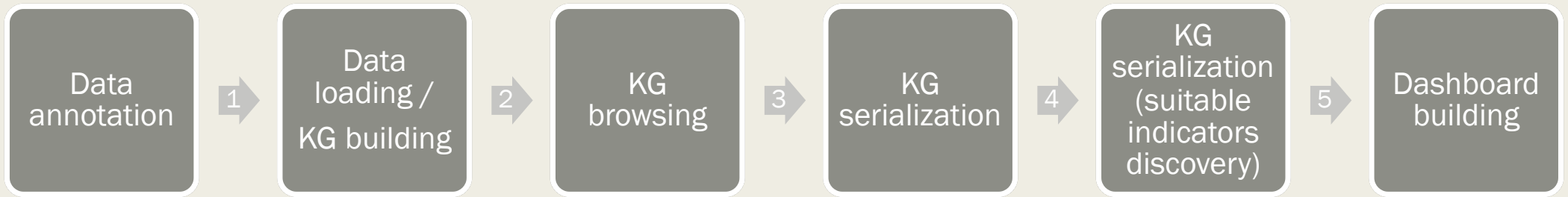
Tons of $CO_2$ saved

0000574.36

- Dataset containing list of bicycle stations
- Dataset containing list of journeys performed during May 2016

As of May 2017

# Data pipeline steps

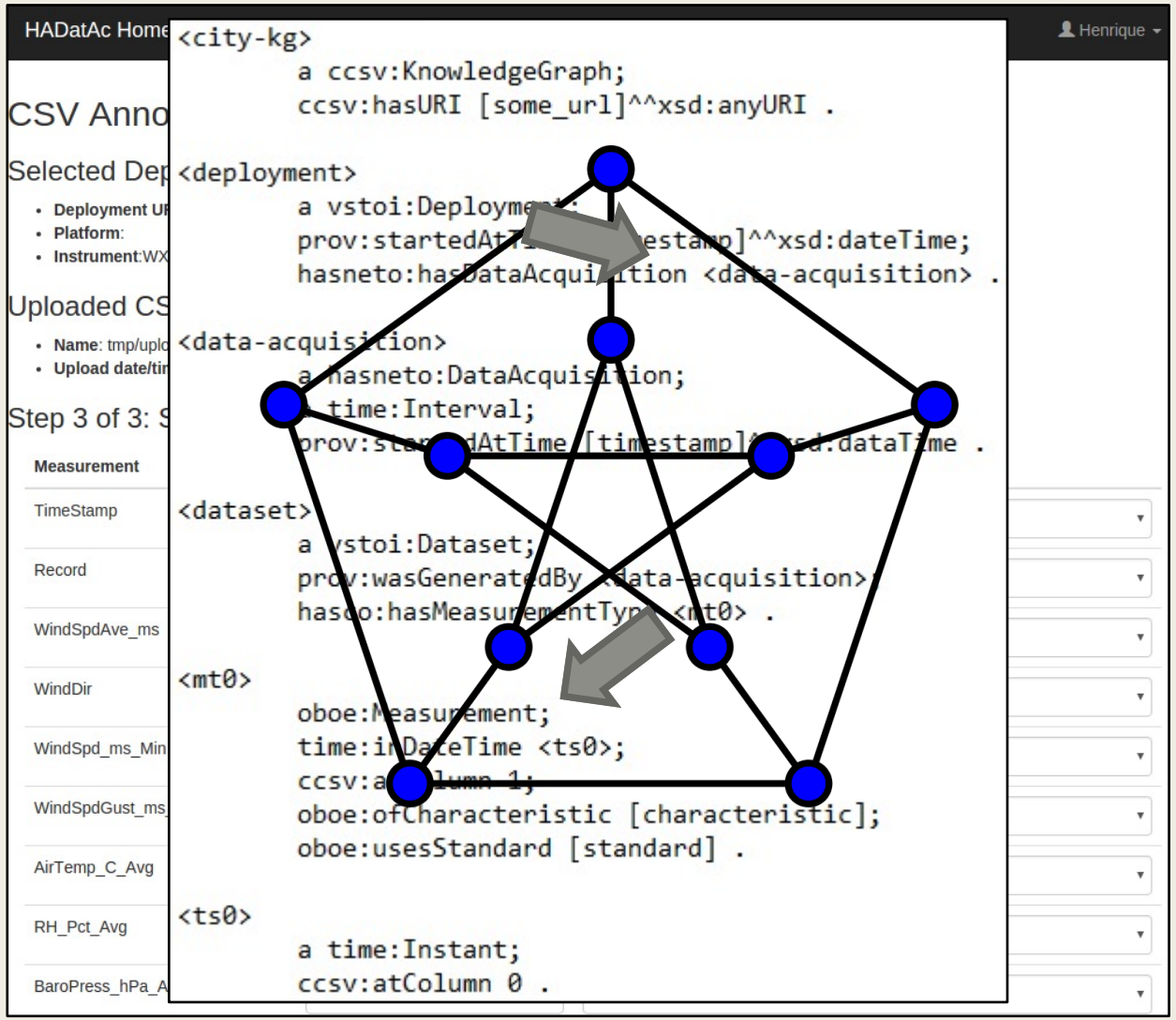


Data annotation → 1 → Data loading / KG building → 2 → KG browsing → 3 → KG serialization → 4 → KG serialization (suitable indicators discovery) → 5 → Dashboard building

CCSV-Annotator

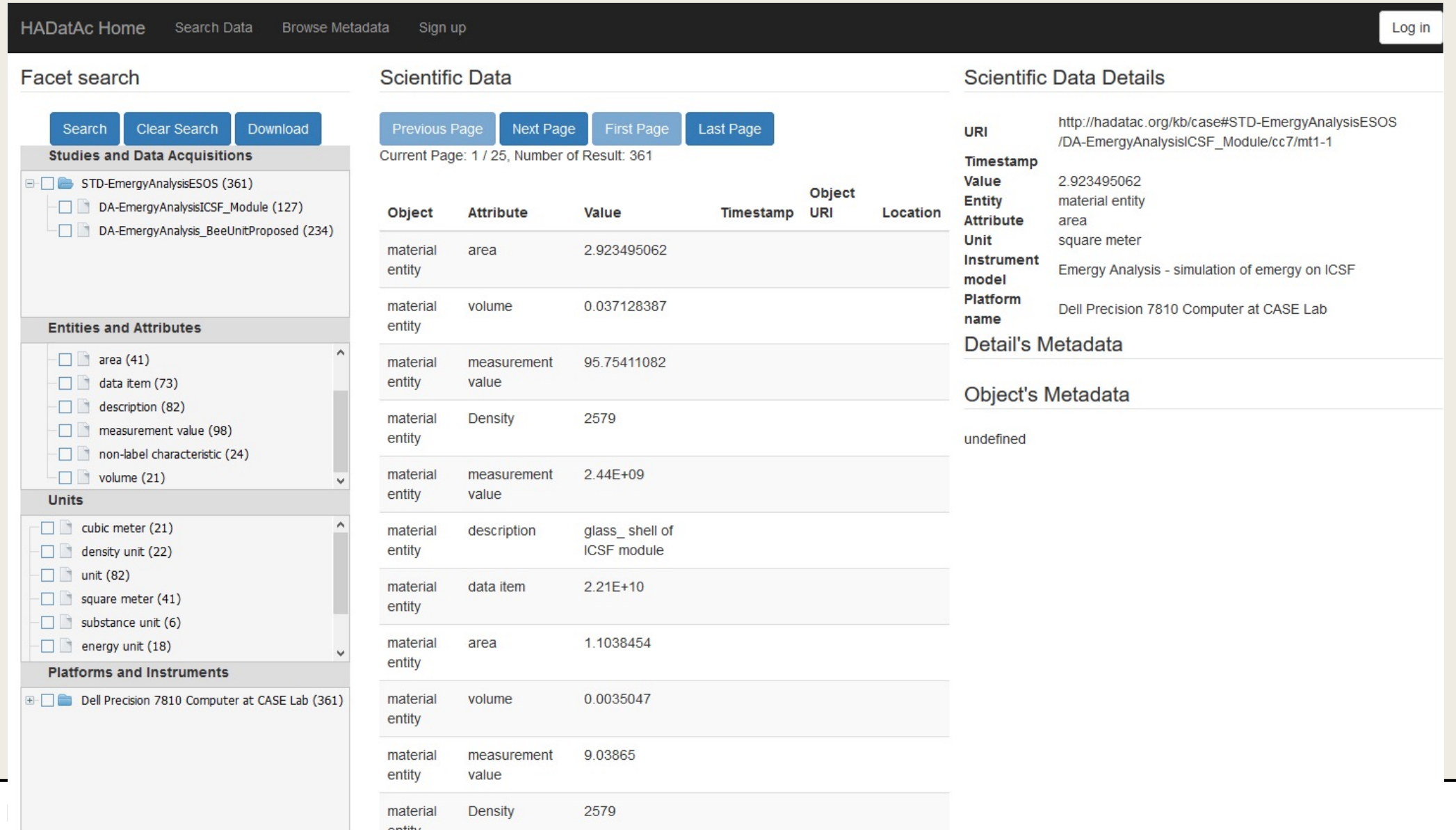
CCSV (Contextualized CSV)

CCSV-Loader



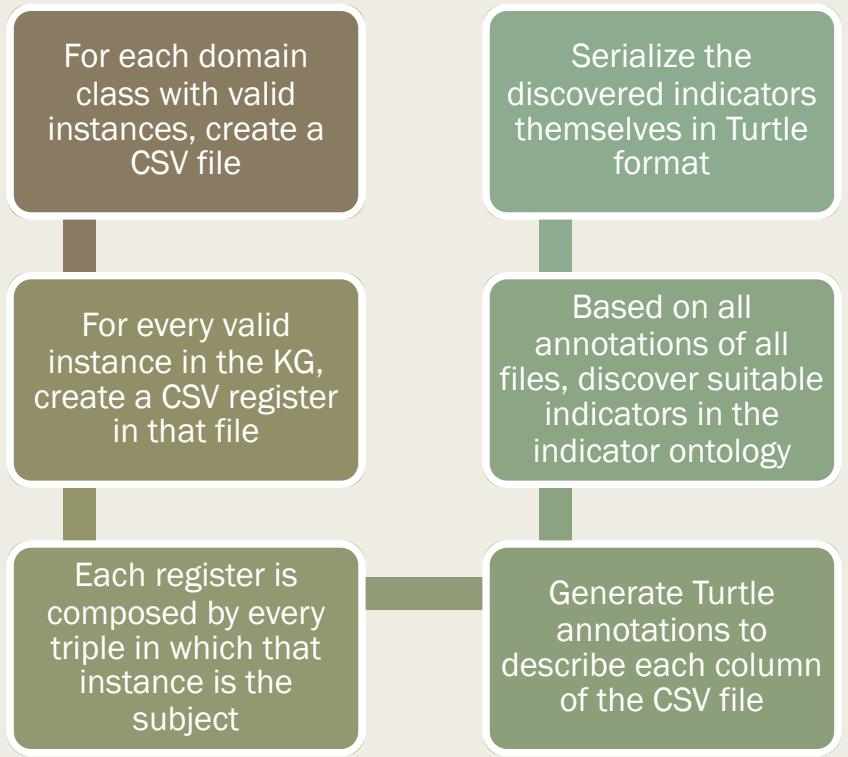
FUNDAÇÃO EDSON QUEIROZ
UNIVERSIDADE DE FORTALEZA
ENSINANDO E APRENDENDO

Data annotation → 1 → Data loading / KG building → 2 → KG browsing → 3 → KG serialization → 4 → KG serialization (suitable indicators discovery) → 5 → Dashboard building




- Most existing data tools (R, Python, Gephi, Business Intelligence softwares) are not ready to deal with RDF model serialization formats like RDF/XML, JSON-LD or Turtle
- Most of the times they expect tabular data
- To foster automatic visualization, a set of possible calculations (indicators) over the data should be attached to the serialized data

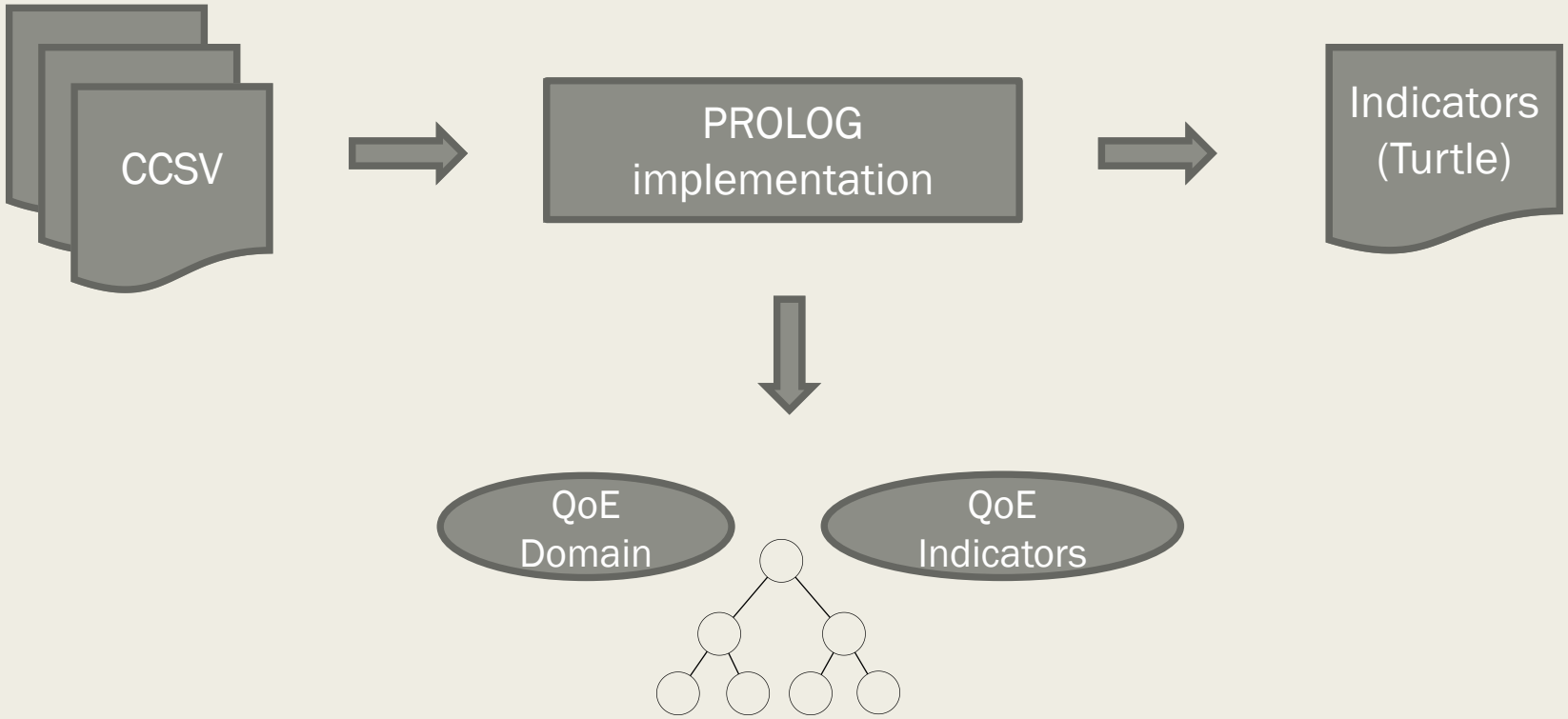

Data annotation
1
Data loading / KG building
2
KG browsing
3
KG serialization
4
KG serialization (suitable indicators discovery)
5
Dashboard building

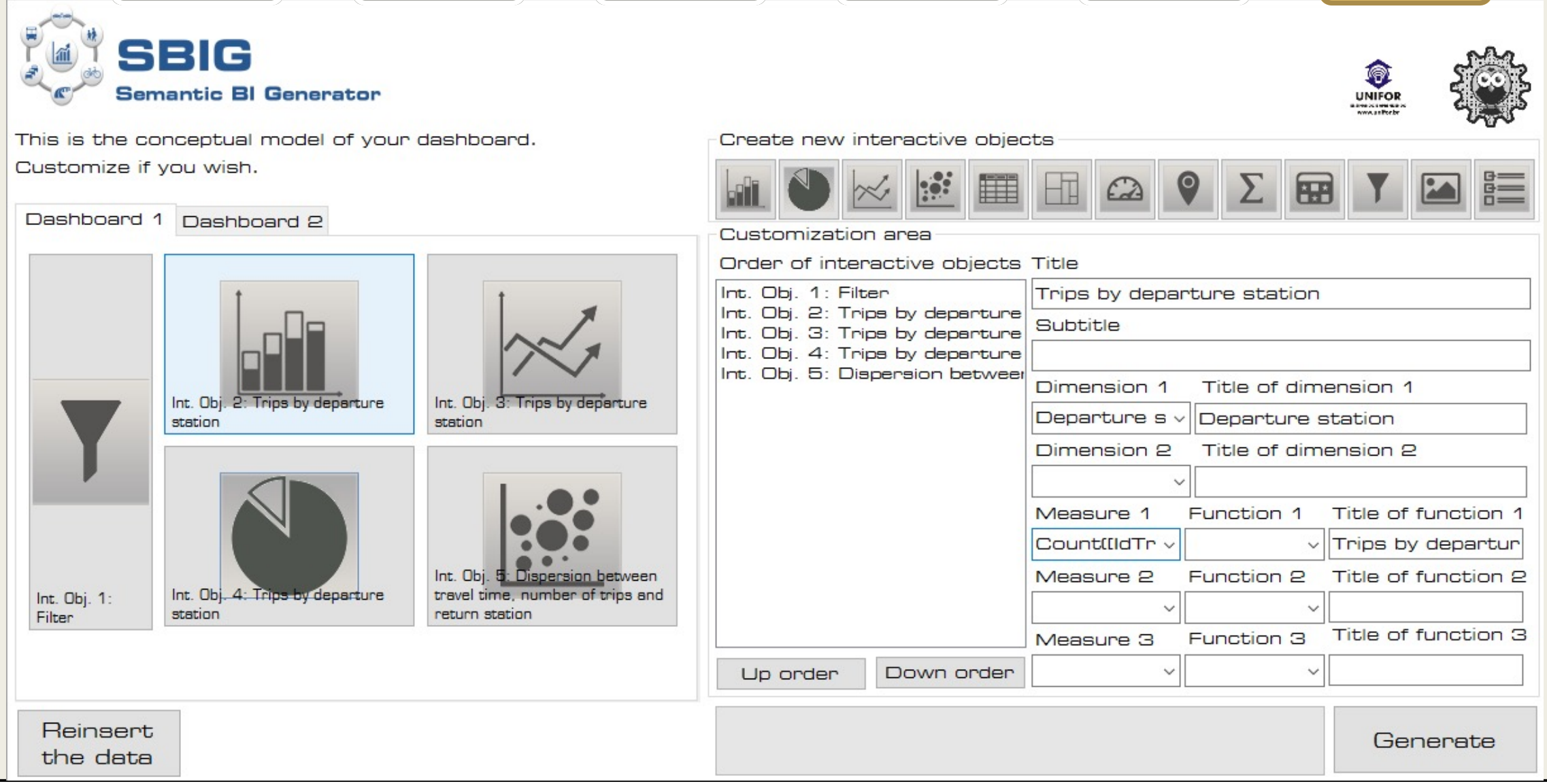

For each domain class with valid instances, create a CSV file
For every valid instance in the KG, create a CSV register in that file
Each register is composed by every triple in which that instance is the subject
Generate Turtle annotations to describe each column of the CSV file
Based on all annotations of all files, discover suitable indicators in the indicator ontology
Serialize the discovered indicators themselves in Turtle format

# Serialized KG entities in CCSV format

```
<trips> a vstoi:Dataset; ccsv:hasDataRecord <reg> .
<reg>
  a qoe-m:Bicycle-Share_Trip; dc:identifier <id> .
  qoe-m:has_Bicycle-Share_User <usr> ;
  qoe-m:has_source_Bicycle-share_Station <src> ;
  qoe-m:has_target_Bicycle-share_Station <trg> .
<src> a qoe-m:Bicycle-Share_Station; dc:identifier <src_id> .
<trg> a qoe-m:Bicycle-Share_Station; dc:identifier <trg_id> .
<usr> a qoe-m:Bicycle-Share_User; dc:identifier <usr_id> .
<id> ccsv:atColumn 0 .
<src_id> ccsv:atColumn 4 .
<trg_id> ccsv:atColumn 7 .
<usr_id> ccsv:atColumn 1 .
```

Data annotation 1 Data loading / KG building 2 KG browsing 3 KG serialization 4 KG serialization (suitable indicators discovery) 5 Dashboard building




SBIG
Semantic BI Generator
UNIFOR
This is the conceptual model of your dashboard.
Customize if you wish.
Dashboard 1
Dashboard 2
Int. Obj. 1: Filter
Int. Obj. 2: Trips by departure station
Int. Obj. 3: Trips by departure station
Int. Obj. 4: Trips by departure station
Int. Obj. 5: Dispersion between travel time, number of trips and return station
Create new interactive objects
Customization area
Order of interactive objects
Int. Obj. 1: Filter
Int. Obj. 2: Trips by departure
Int. Obj. 3: Trips by departure
Int. Obj. 4: Trips by departure
Int. Obj. 5: Dispersion betwee
Up order
Down order
Title
Trips by departure station
Subtitle
Dimension 1
Title of dimension 1
Departure s
Departure station
Dimension 2
Title of dimension 2
Measure 1
Function 1
Title of function 1
CountDIdTr
Trips by departur
Measure 2
Function 2
Title of function 2
Measure 3
Function 3
Title of function 3
Reinsert the data
Generate

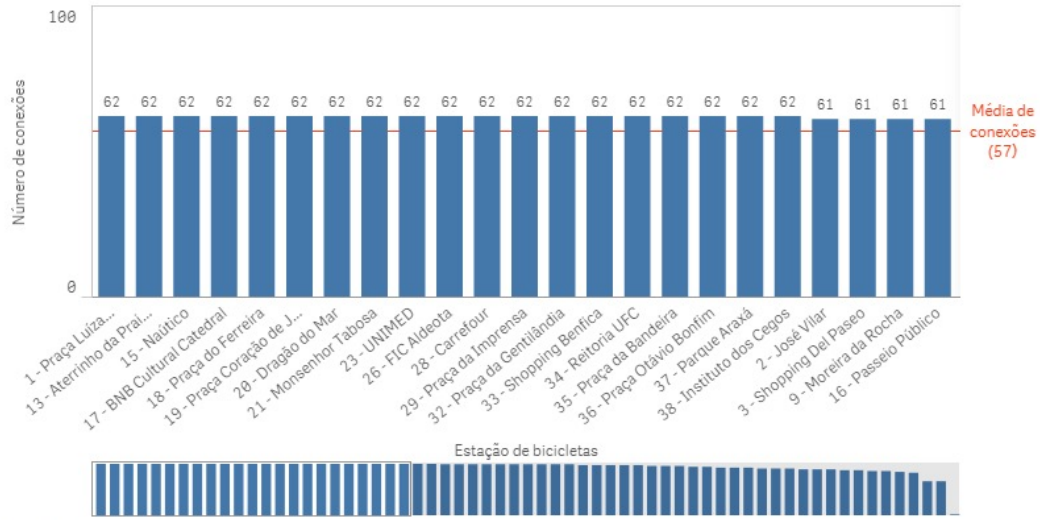
Nenhuma seleção aplicada

# Dashboard I

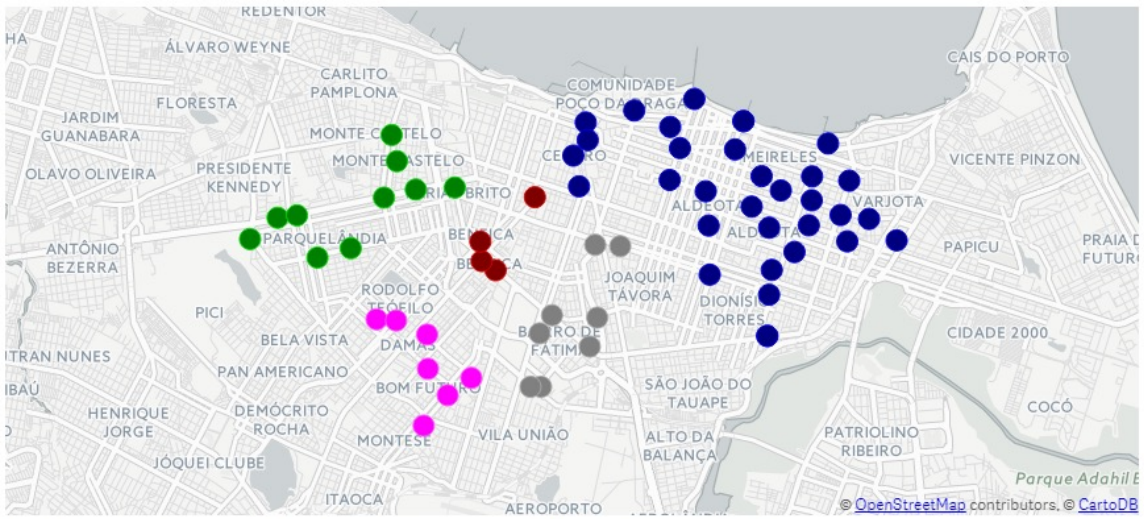
**Filtros**

- Rota
- Estação de retirada
- Estação de devolução
- Número de viagens

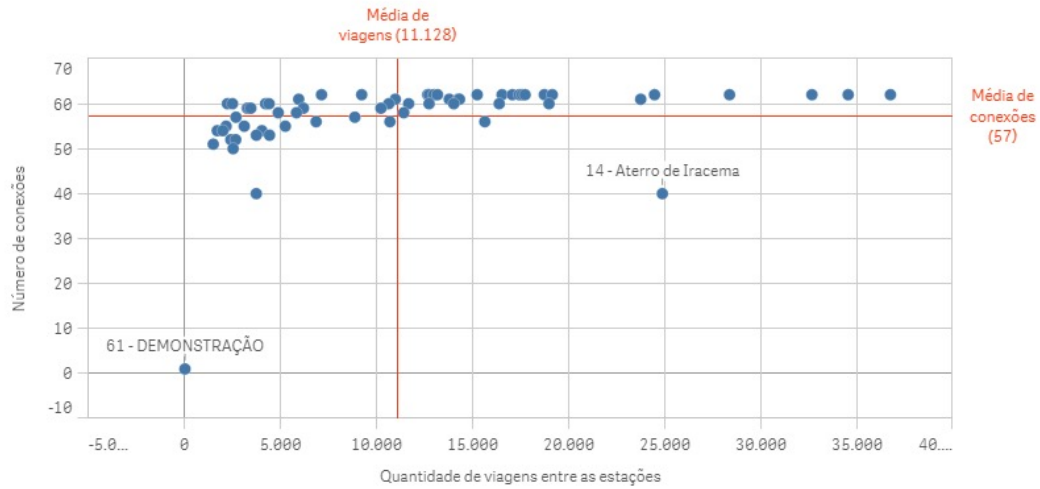
**Número de conexões por estação**




**Comunidades descobertas**

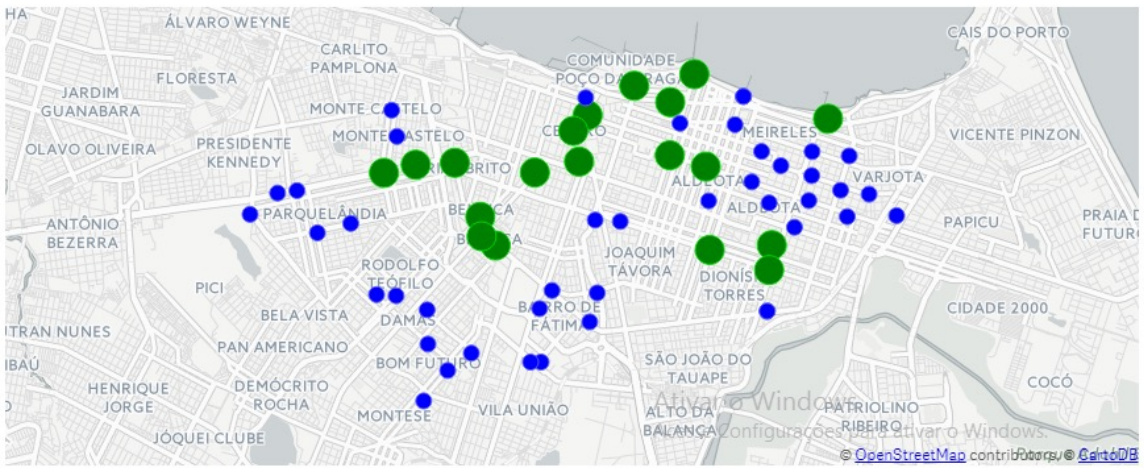



**Relação entre o número de viagens das estações e o número de conexões**




**Estações que podem ser consideradas o centro do sistema de bicicletas compartilhadas**

Verde = alta probabilidade, Vermelho = baixa probabilidade, Azul = média probabilidade




*O cálculo leva em consideração a conectividade das estações e não a sua posição geográfica.*

Ativar o Windows
Acesse Configurações para ativar o Windows.

Nenhuma seleção aplicada

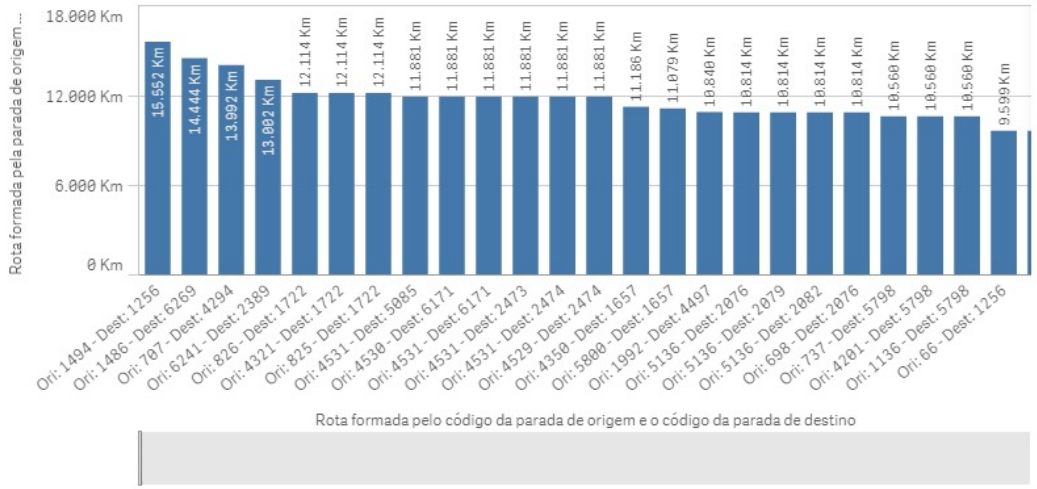
# Relatório de caminhos reais e ótimos (1)

**Filtros**

- Rota
- Paradas de origem
- Paradas de destino
- Número de viagens

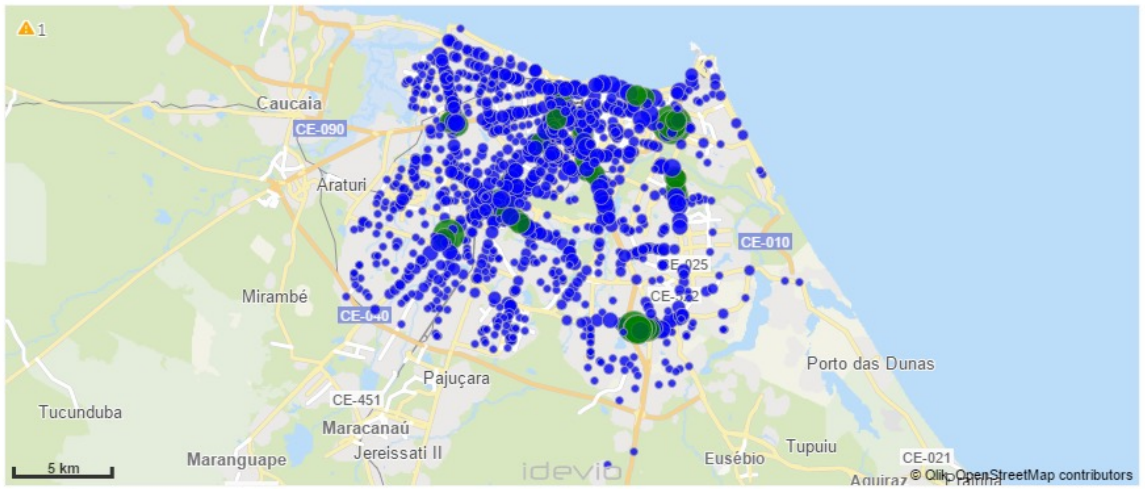
**Maior distância que um usuário pode realizar em uma única viagem**


**Rotas candidatas para criação de linhas expresas**



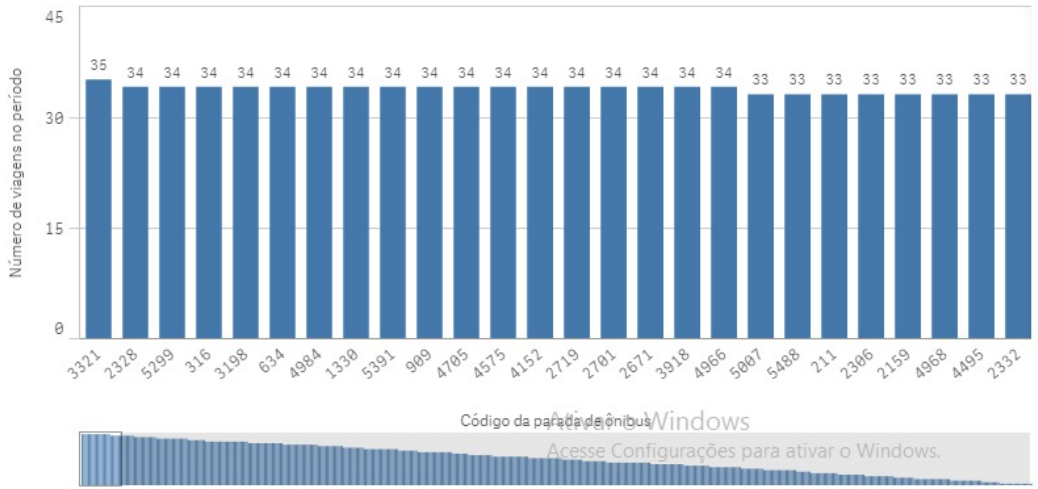

*A indicação das rotas leva em conta a distância e o volume de viagens*

**Pontos favoráveis para criação de terminais de ônibus**




*A indicação dos pontos toma como base o volume de oferta.*

**Sugestão de paradas de ônibus que poderiam ser desativadas**




*A sugestão leva em conta os menores volumes de oferta*

Ativar o Windows
Acesse Configurações para ativar o Windows.

# Conclusions and future work

- City Knowledge Graph description in support of automatic generation of dashboards
- Indicator and domain ontology
- SBIG: Semantic BI Generator application

- Extension of the approach to support more complex indicator values (for instance, network algorithms and their meanings for each network)
- HADatAc: Human-Aware Data Acquisition Framework

# Thank you for your attention

**From Data to City Indicators: A Knowledge Graph for Supporting Automatic Generation of Dashboards**

Henrique Santos, Victor Dantas, Vasco Furtado, Paulo Pinheiro and Deborah L. McGuinness

hos@edu.unifor.br

@hansidm

http://henriquesantos.org

FUNDAÇÃO EDSON QUEIROZ
UNIVERSIDADE DE FORTALEZA
ENSINANDO E APRENDENDO

TetherlessWorld