TransportDCAT-AP: A common metadata model

Federated SPARQL queries over Public Transport domain

Civic Lab - 18/07/2017

David Chaves
dchaves@fi.upm.es

@dchavesf

dachafra

CC BY NC SA

- Introduction
- The importance of metadata in open data (portals)
- EU regulation on the provision of EU-wide multimodal travel information services
- TransportDCAT-AP: Metadata profile for public transport domain
- PhD thesis: Federated queries over heterogeneous resources

Who am I?

PhD Student at Ontology Engineering Group

PhD Student and researcher at OEG

- PhD thesis - Early stage (1/4 years)

  "Federated query processing over heterogeneous native and virtual RDF datasets"

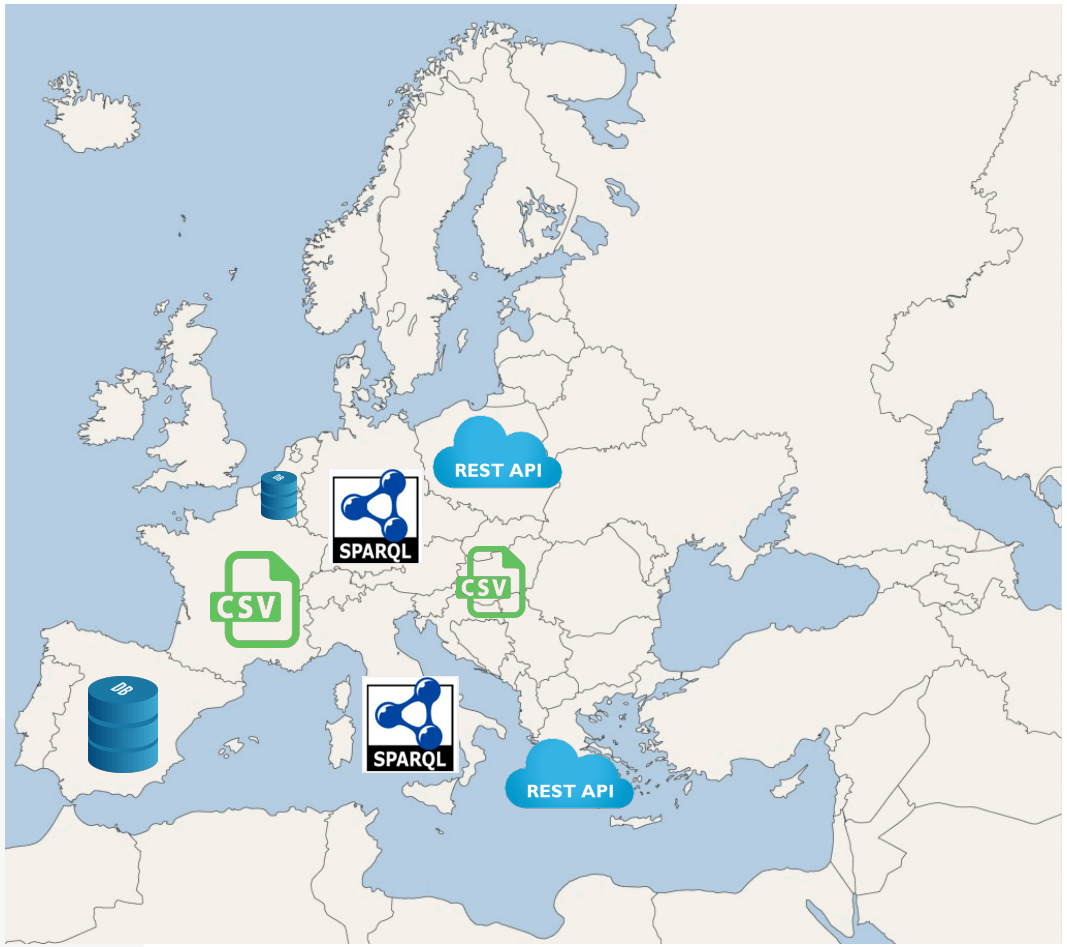
- EU Projects focused on **Semantic Interoperability in Public Transport**

The importance of metadata in Open Data

- There is a clear lack of metadata in most of the open data portals
- Open data is considered a key asset for the development of many types of open applications
- Availability of such metadata will make it possible for agents to discover Open Data throughout Europe
- Metadata provides an easy way to get relevant information from datasets

EU regulation on the provision of EU-wide multimodal travel information services

*"In order to facilitate the **easy exchange and re-use of these data** for the provision of comprehensive travel information services, corresponding metadata and information on the quality of the data will be accessible to users through **a national or common access point**."*

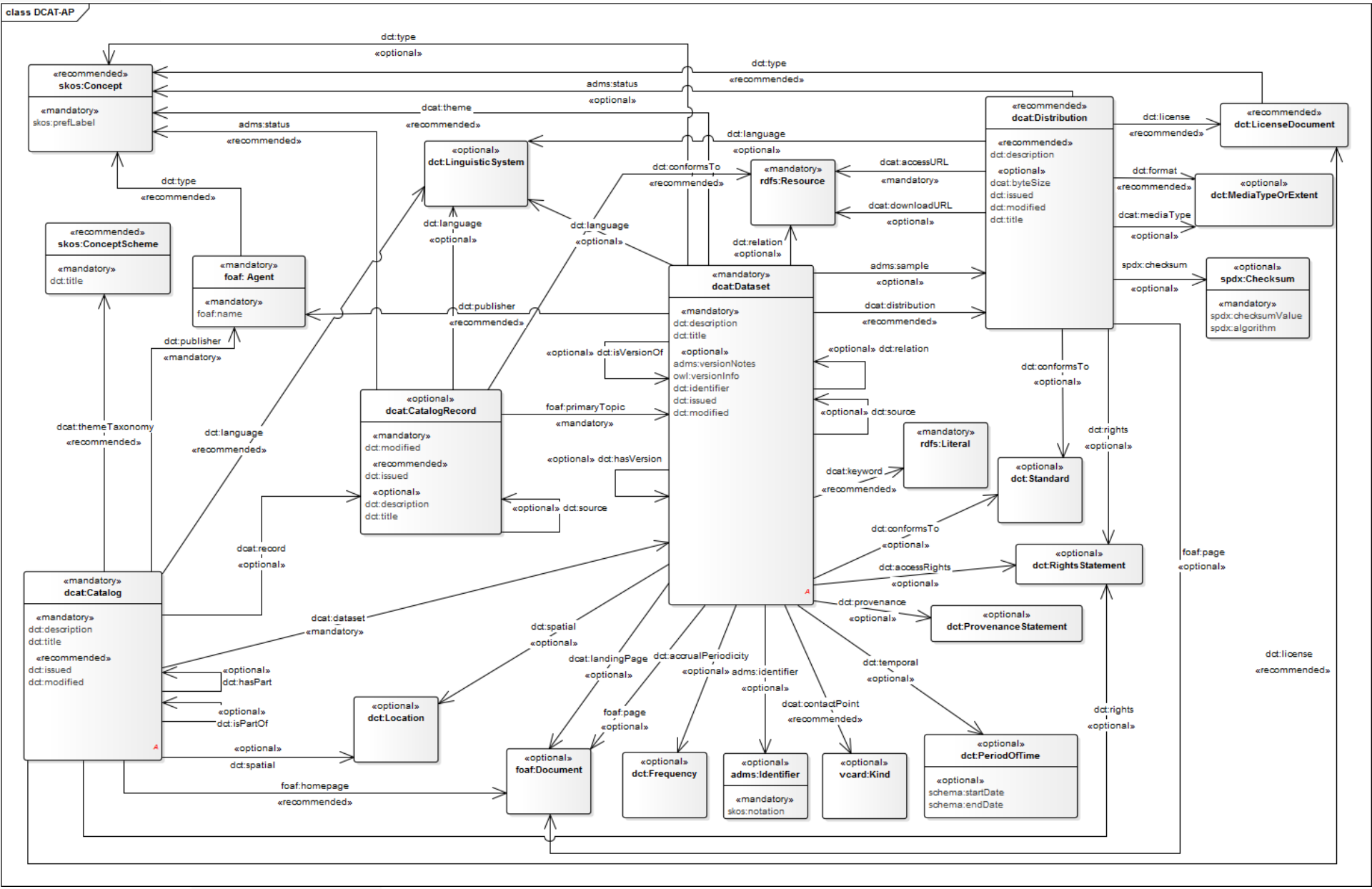


TransportDCAT-AP

Metadata profile for public transport domain

- OASIS: Open Applications for Semantically Interoperability Services
- Partners: UPM, CRTM, OKF Belgium, Stad Gent, iMinds and VICTOR
  - Toolkit for non-experts focused on Public Transport/Services

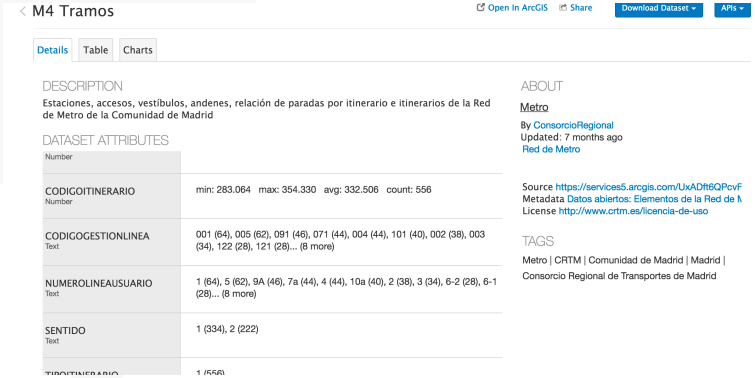
- Development of new metadata profile: TransportDCAT-AP
  - Based on EU metadata standard (DCAT-AP)
  - Included controller vocabularies
  - CRTM metadata is following TransportDCAT-AP
  - 10 organizations were involved in its validation
  - Undergrad student will develop a WebApp


class DCAT-AP
«recommended» skos:Concept
«mandatory» skos:prefLabel
«recommended» skos:ConceptScheme
«mandatory» dct:title
«mandatory» foaf: Agent
«mandatory» foaf:name
«optional» dct:LinguisticSystem
«mandatory» rdfs:Resource
«recommended» dcat:Distribution
«recommended» dct:description
«optional» dcat:byteSize dct:issued dct:modified dct:title
«recommended» dct:LicenseDocument
«optional» dct:MediaTypeOrExtent
«optional» spdx:Checksum
«mandatory» spdx:checksumValue spdx:algorithm
«mandatory» dcat:Dataset
«mandatory» dct:description dct:title
«optional» adms:versionNotes owl:versionInfo dct:identifier dct:issued dct:modified
«optional» dcat:CatalogRecord
«mandatory» dct:modified
«recommended» dct:issued
«optional» dct:description dct:title
«mandatory» rdfs:Literal
«optional» dct:Standard
«optional» dct:RightsStatement
«optional» dct:ProvenanceStatement
«mandatory» dcat:Catalog
«mandatory» dct:description dct:title
«recommended» dct:issued dct:modified
«optional» dct:Location
«optional» foaf:Document
«optional» dct:Frequency
«optional» adms:Identifier
«mandatory» skos:notation
«optional» vcard:Kind
«optional» dct:PeriodOfTime
«optional» schema:startDate schema:endDate
dct:type «optional»
dct:type «recommended»
adms:status «optional»
dcat:theme «recommended»
adms:status «recommended»
dct:language «optional»
dct:type «recommended»
dct:conformsTo «recommended»
dcat:accessURL «mandatory»
dcat:downloadURL «optional»
dct:license «recommended»
dct:format «recommended»
dcat:mediaType «optional»
spdx:checksum «optional»
dct:language «optional»
dct:language «optional»
dct:relation «optional»
adms:sample «optional»
dct:distribution «recommended»
dct:publisher «recommended»
dct:publisher «mandatory»
«optional» dct:isVersionOf
«optional» dct:relation
«optional» dct:source
dct:conformsTo «optional»
dcat:themeTaxonomy «recommended»
dct:language «recommended»
foaf:primaryTopic «mandatory»
«optional» dct:hasVersion
«optional» dct:source
dcat:keyword «recommended»
dct:rights «optional»
dct:conformsTo «optional»
dcat:record «optional»
dct:accessRights «optional»
foaf:page «optional»
dct:provenance «optional»
dcat:dataset «mandatory»
dct:spatial «optional»
dcat:landingPage «optional»
dct:accrualPeriodicity «optional»
adms:identifier «optional»
dct:temporal «optional»
dct:license «recommended»
«optional» dct:hasPart
«optional» dct:isPartOf
«optional» dct:spatial
foaf:page «optional»
dcat:contactPoint «recommended»
dct:rights «optional»
foaf:homepage «recommended»

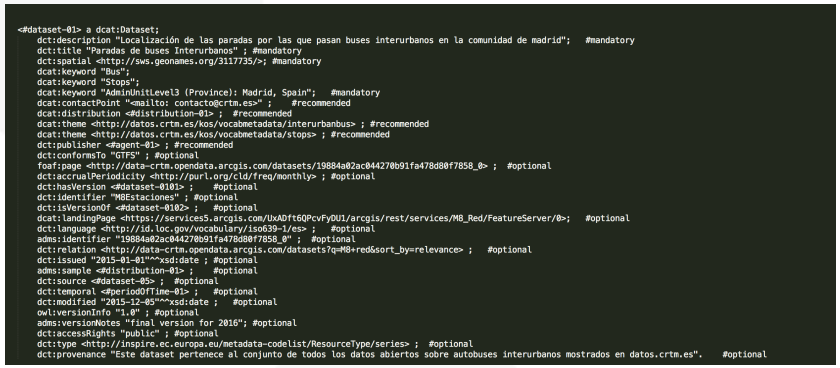
- The properties **dct:spatial** and **dcat:keywords** from dcat:Catalog and dcat:Dataset changes from **Optional** to **Mandatory**.
- The property dcat:keywords modify its range following:
  - dcat:keyword A, where A is the type of transport (Tram, Metro, Rail, Bus, Ferry, Cable car, Suspended cable car, Funicular or Other)
  - dcat:keyword B, where B is the type of dataset (Calendar, Fare, Frequencies, Routes, Shapes, Stops, Transfers, Trips or Other)
  - dcat:keyword C, where C is an INSPIRE feature for address (AdminUnitLevel3 (Province): Madrid)

dcat:Dataset dcat:keyword "Bus";
dcat:Dataset dcat:keyword "Stops";
dcat:Dataset dcat:keyword "AdminUnitLevel3 (Province): Madrid, Spain";

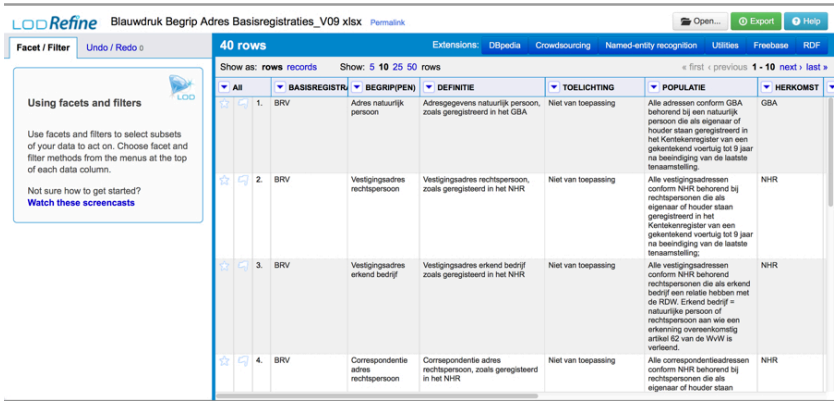
CRTM Transformation



	dcat:accessURL	dct:description	dct:format	dct:license	dcat:downloadURL	dct:language
<urbanbus-stops/csv>	http://data-crtm.opendata.arcgis.com/datasets/46044e95c2f340e6a9e07908842bbbef2_0	CSV file with the localizations of the urban buses stops	CSV	http://www.crtm.es/licencia-de-uso	http://data-crtm.opendata.arcgis.com/datasets/46044e95c2f340e6a9e07908842bbbef2_0.csv	http://id.loc.gov/vocabulary/iso639-1/es
<urbanbus-stops/kml>	http://data-crtm.opendata.arcgis.com/datasets/46044e95c2f340e6a9e07908842bbbef2_0	KML file with the localizations of the urban buses stops	KML	http://www.crtm.es/licencia-de-uso	http://data-crtm.opendata.arcgis.com/datasets/46044e95c2f340e6a9e07908842bbbef2_0.kml	http://id.loc.gov/vocabulary/iso639-1/es
<urbanbus-stops/shp>	http://data-crtm.opendata.arcgis.com/datasets/46044e95c2f340e6a9e07908842bbbef2_0	SHP files with the localizations of the urban buses stops	SHP	http://www.crtm.es/licencia-de-uso	http://data-crtm.opendata.arcgis.com/datasets/46044e95c2f340e6a9e07908842bbbef2_0.zip	http://id.loc.gov/vocabulary/iso639-1/es
<urbanbus-lines/csv>	http://data-crtm.opendata.arcgis.com/datasets/46044e95c2f340e6a9e07908842bbbef2_2	CSV file about the different lines of the urban buses in the City of Madrid	CSV	http://www.crtm.es/licencia-de-uso	http://data-crtm.opendata.arcgis.com/datasets/46044e95c2f340e6a9e07908842bbbef2_2.csv	http://id.loc.gov/vocabulary/iso639-1/es
<urbanbus-lines/kml>	http://data-crtm.opendata.arcgis.com/datasets/46044e95c2f340e6a9e07908842bbbef2_2	KML file about the different lines of the urban buses in the City of Madrid	KML	http://www.crtm.es/licencia-de-uso	http://data-crtm.opendata.arcgis.com/datasets/46044e95c2f340e6a9e07908842bbbef2_2.kml	http://id.loc.gov/vocabulary/iso639-1/es
<urbanbus-lines/shp>	http://data-crtm.opendata.arcgis.com/datasets/46044e95c2f340e6a9e07908842bbbef2_2	SHP files about the different lines of the urban buses in the City of Madrid	SHP	http://www.crtm.es/licencia-de-uso	http://data-crtm.opendata.arcgis.com/datasets/46044e95c2f340e6a9e07908842bbbef2_2.zip	http://id.loc.gov/vocabulary/iso639-1/es
<interurbanbus-stops/csv>	http://data-crtm.opendata.arcgis.com/datasets/19884a02ac044270b91fa478d80f7858_0	CSV file with the localizations of the interurban buses stops	CSV	http://www.crtm.es/licencia-de-uso	http://data-crtm.opendata.arcgis.com/datasets/19884a02ac044270b91fa478d80f7858_0.csv	http://id.loc.gov/vocabulary/iso639-1/es



```
<#dataset-01> a dcat:Dataset;
    dct:description "Localización de las paradas por las que pasan buses interurbanos en la comunidad de madrid";   #mandatory
    dct:title "Paradas de buses Interurbanos" ; #mandatory
    dct:spatial <http://sws.geonames.org/3117735/>; #mandatory
    dcat:keyword "Bus";
    dcat:keyword "Stops";
    dcat:keyword "AdminUnitLevel3 (Province); Madrid, Spain";   #mandatory
    dcat:contactPoint "<mailto: contacto@crtm.es>" ;    #recommended
    dcat:distribution <#distribution-01> ;  #recommended
    dcat:theme <http://datos.crtm.es/kos/vocabmetadata/interurbanbus> ; #recommended
    dcat:theme <http://datos.crtm.es/kos/vocabmetadata/stops> ; #recommended
    dct:publisher <#agent-01> ; #recommended
    dct:conformsTo "GTFS" ; #optional
    foaf:page <http://data-crtm.opendata.arcgis.com/datasets/19884a02ac044270b91fa478d80f7858_0> ;  #optional
    dct:accrualPeriodicity <http://purl.org/cld/freq/monthly> ; #optional
    dct:hasVersion <#dataset-0101> ;    #optional
    dct:identifier "M8Estaciones" ; #optional
    dct:isVersionOf <#dataset-0102> ;   #optional
    dcat:landingPage <https://services5.arcgis.com/UxADft6QPcvFyDU1/arcgis/rest/services/M8_Red/FeatureServer/0>;   #optional
    dct:language <http://id.loc.gov/vocabulary/iso639-1/es> ;   #optional
    adms:identifier "19884a02ac044270b91fa478d80f7858_0" ;  #optional
    dct:relation <http://data-crtm.opendata.arcgis.com/datasets?q=M8+red&sort_by=relevance> ;   #optional
    dct:issued "2015-01-01"^^xsd:date ; #optional
    adms:sample <#distribution-01> ;    #optional
    dct:source <#dataset-05> ;  #optional
    dct:temporal <#periodOfTime-01> ;   #optional
    dct:modified "2015-12-05"^^xsd:date ;   #optional
    owl:versionInfo "1.0" ; #optional
    adms:versionNotes "final version for 2016"; #optional
    dct:accessRights "public" ; #optional
    dct:type <http://inspire.ec.europa.eu/metadata-codelist/ResourceType/series> ;  #optional
    dct:provenance "Este dataset pertenece al conjunto de todos los datos abiertos sobre autobuses interurbanos mostrados en datos.crtm.es".    #optional
```



LocaliData

AYUNTAMIENTO
cáceres

red.es

CONSORCIO
TRANSPORTES
MADRID

¡MADRID!
EMT

ALCOBENDAS
Un modelo de ciudad

SNCB

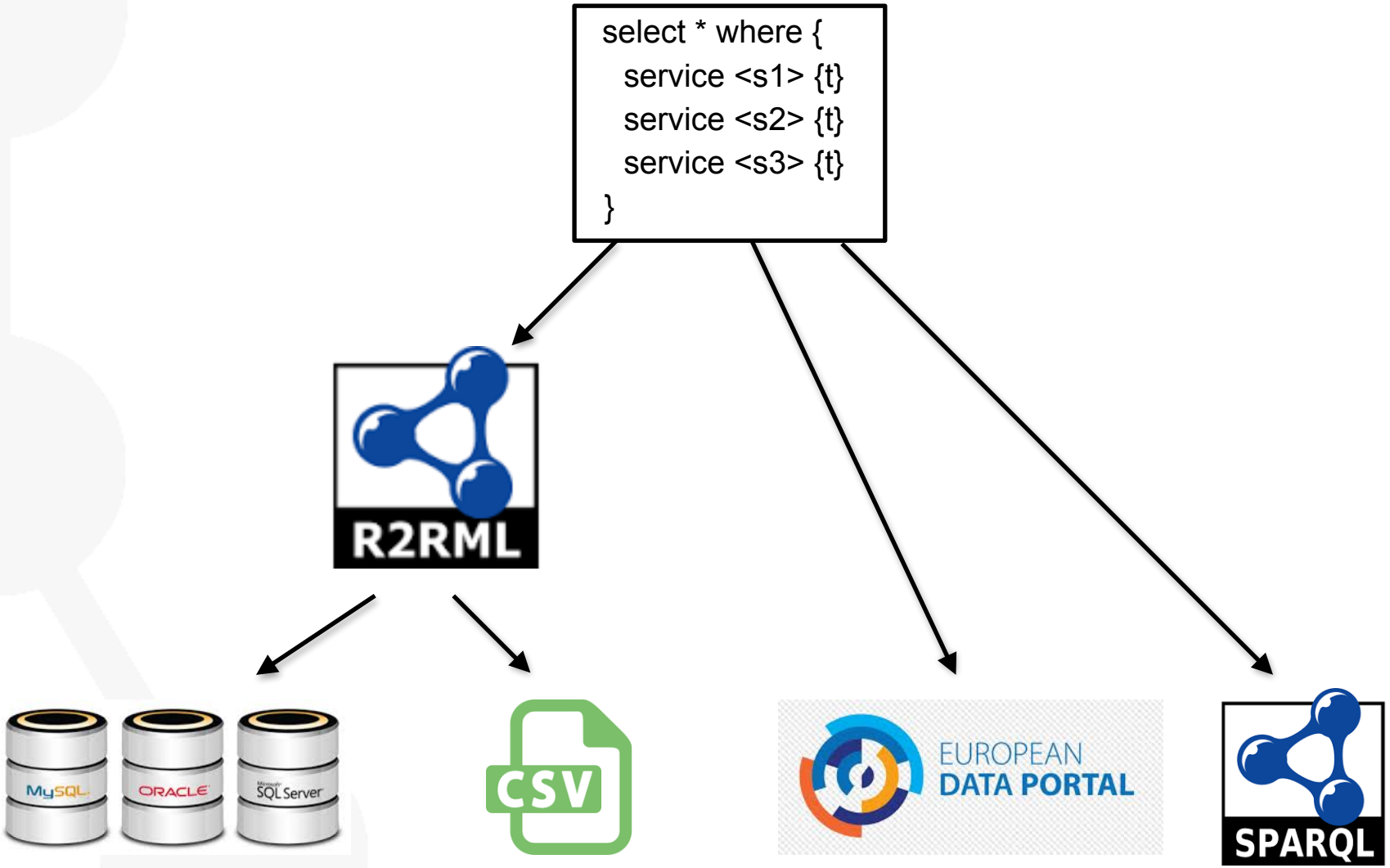
TEC

DE
LIJN

STIB
MIVB
.brussels

Federated query processing over heterogeneous native and virtual RDF datasets

"Federated query processing over heterogeneous native and virtual RDF datasets"



- Two different types:

A

```
select * where {
    gp
}
```

B

```
select * where {
    service <s1> {t}
    service <s2> {t}
    service <s3> {t}
}
```

- Disadvantages of A:
  - Closed world
  - Focused on the source selection problem
  - Without real application
  - Doesn't deal with the problem of availability

- Metadata profile for federated queries over heterogeneous resources
- Extend a federated benchmarking adding SERVICE and GRAPH
- Optimizations of the queries (Future Work)
- Real world environment for evaluation

Ontology Engineering Group

THANK YOU!

QUESTIONS?

David Chaves
dchaves@fi.upm.es

@dchavesf

dachafra

CC BY NC SA