ONETT

Systematic Knowledge Graph Generation for National Access Point

David Chaves-Fraga, Ontology Engineering Group
Universidad Politécnica de Madrid, Spain
Adolfo Antón, OEG-UPM
Jhon Toledo, OEG-UPM
Oscar Corcho, OEG-UPM

dchaves@fi.upm.es
@dchavesf

09/09/2019
Sem4Tra@SEMANTiCS2019

National Access Points

*"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**."* [1]

[1] Supplementing Directive 2010/40/EU of the European Parliament and of the Council with regard to the provision of EU-wide multimodal travel information services

Are Semantic Web technologies a good option to integrate and generate transport data at scale for the National Access Points?

Requirements:

- Ontology based on Transmodel
- Maintainable Knowledge Graph (KG) generation from other data models
- Efficient and robust KG Generation engines

Why Transmodel?

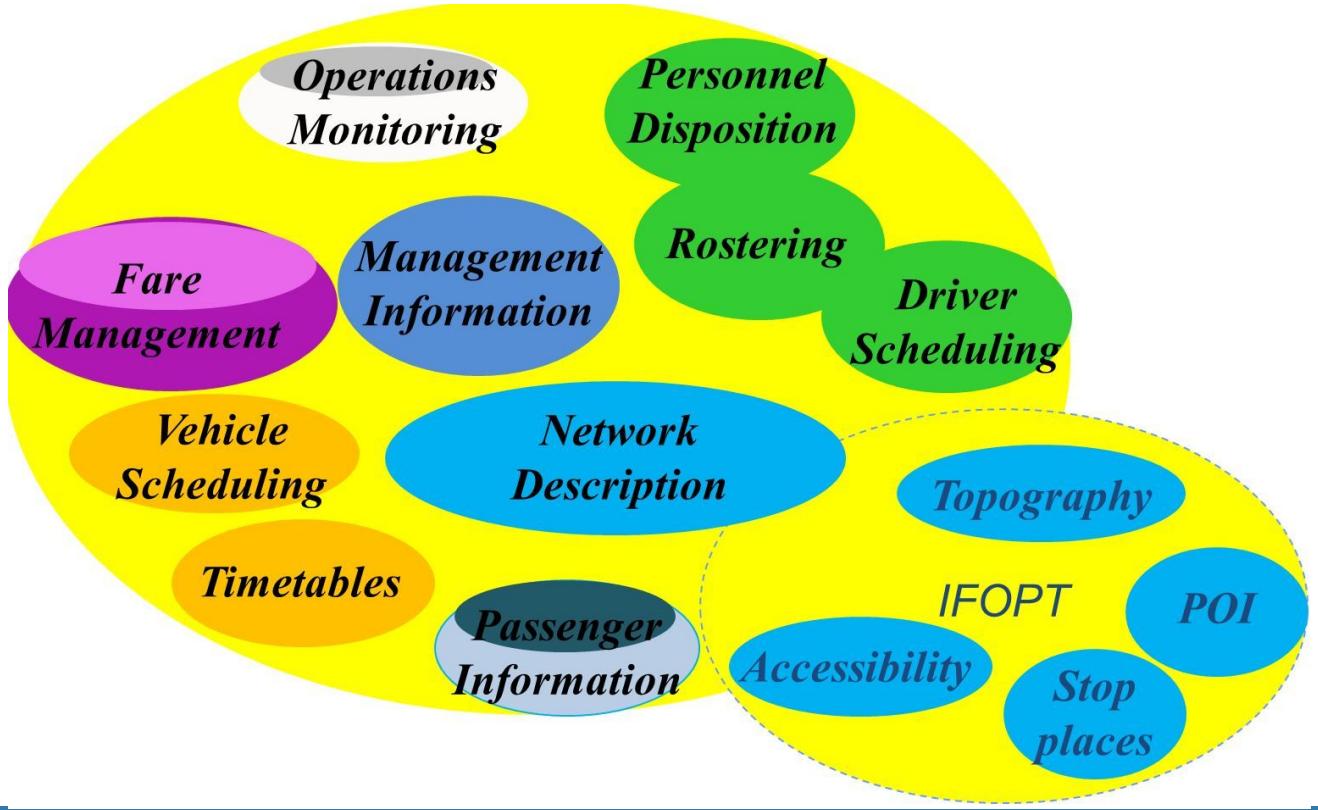
Transmodel is the short name for the European Standard “Public Transport Reference Data Model”:

- European? European ITS Directive 2010/40/EU to provide EU-wide multimodal travel information services available across borders.
- Standard? CEN is the European Committee for Normalization.
- Data Model: NAP should use the CEN data exchange standard NeTEx CEN/TS 16614

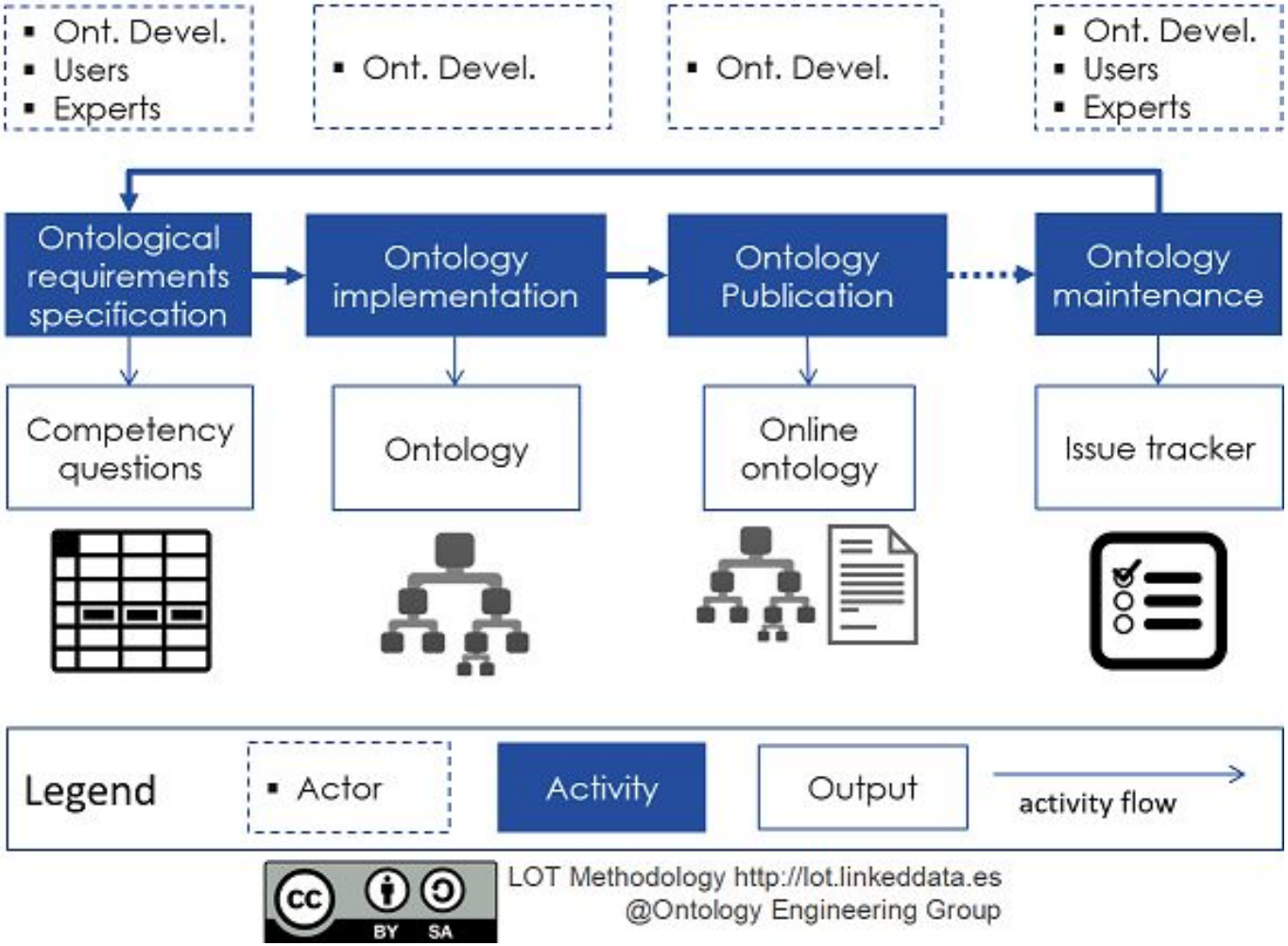
Transmodel is organised in 4 main sections:

- Common Concepts
- Network description
- Timing information
- Vehicle scheduling

But also:



Transmodel Ontology powered by the Linked Open Terms methodology for ontology development



Pitfails...

- The movement of people through many countries, many roads, different policies..
- Huge domain with plenty of technical papers.
- Complex projects for complex reality.
- Different terms in GTFS and Transmodel for similar concepts.
- Some controversial terms in basic concepts as Trip, Journey, Passenger, Vehicle, Service, Organisation.
- Vocabulary split in modules for faster development.

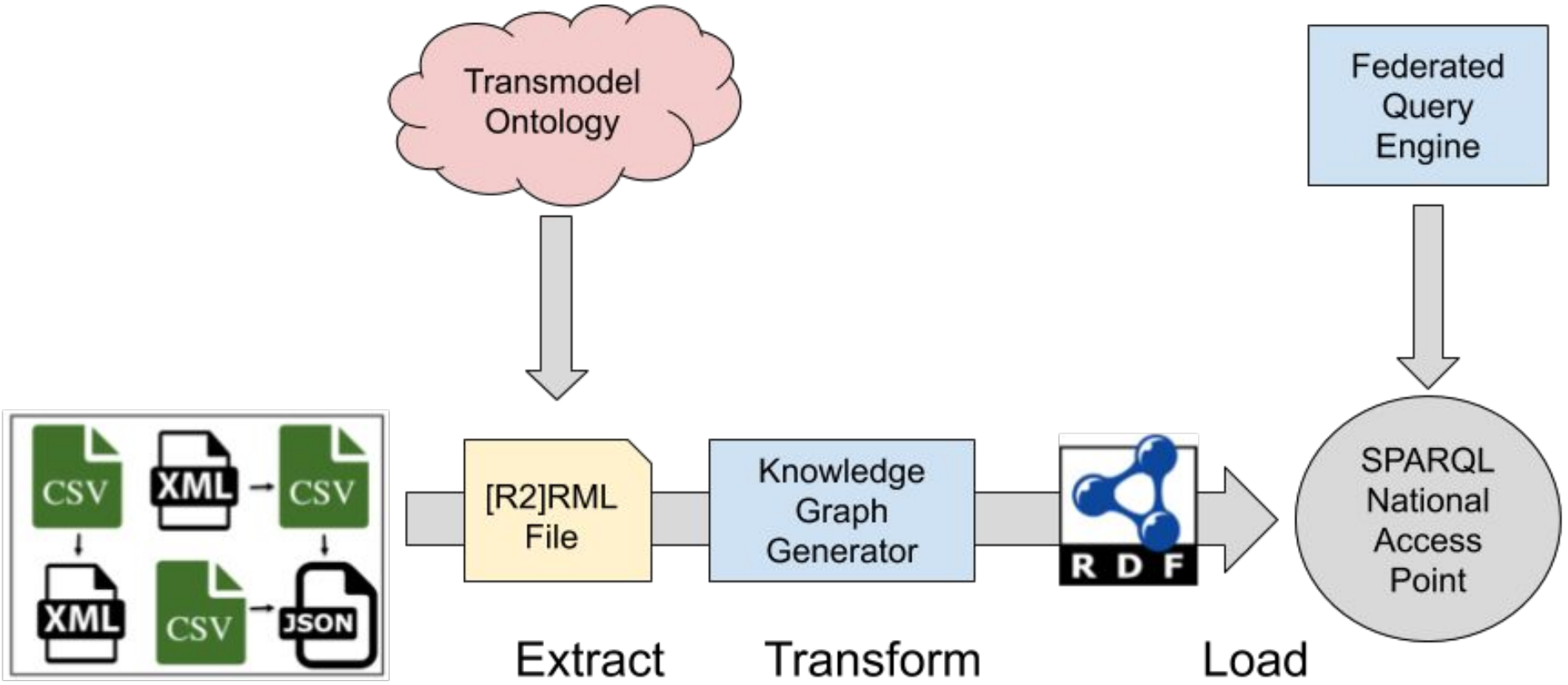
Are Opportunities:

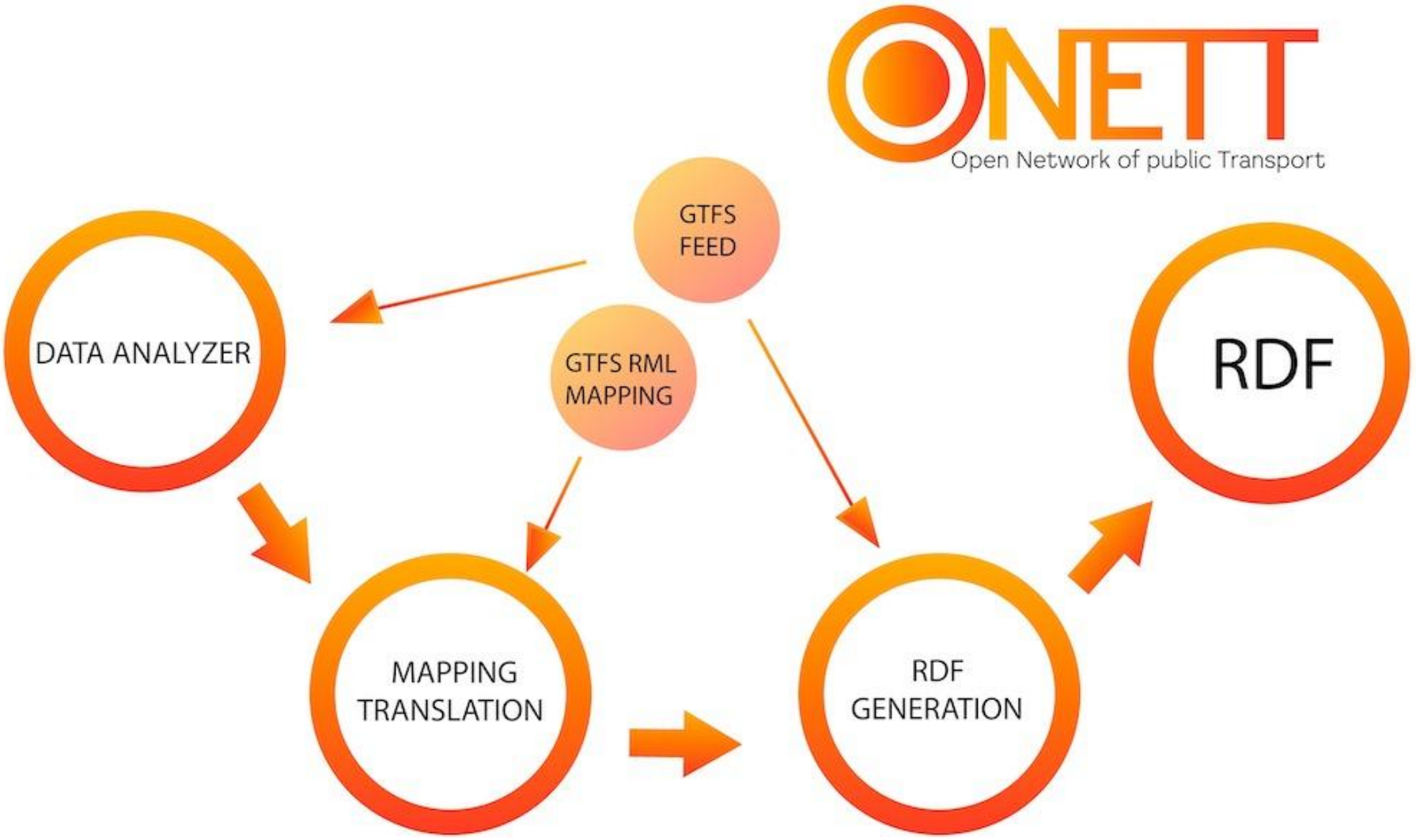
- The transport system is always testing its performance.
- Solid foundations for better implementations.
- Simplicity comes from knowledge based projects.

Check the state of art:

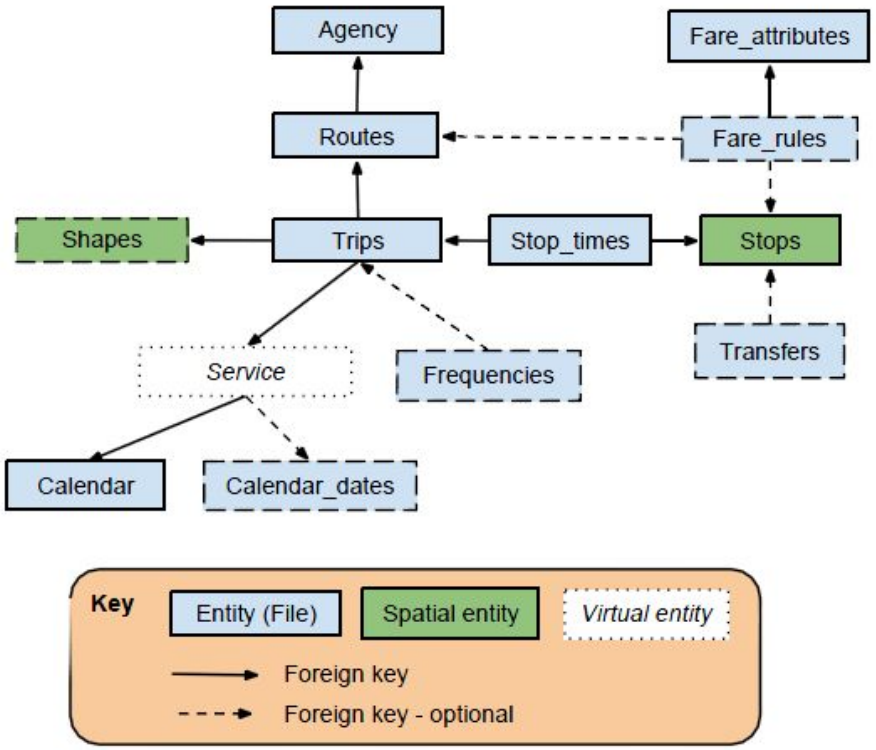
- https://github.com/oeg-upm/transmodel-ontology
- 2 modules already published: authorities and facilities
- 3 modules in process: commons, journeys, fares

KG Generation on the Transport Domain



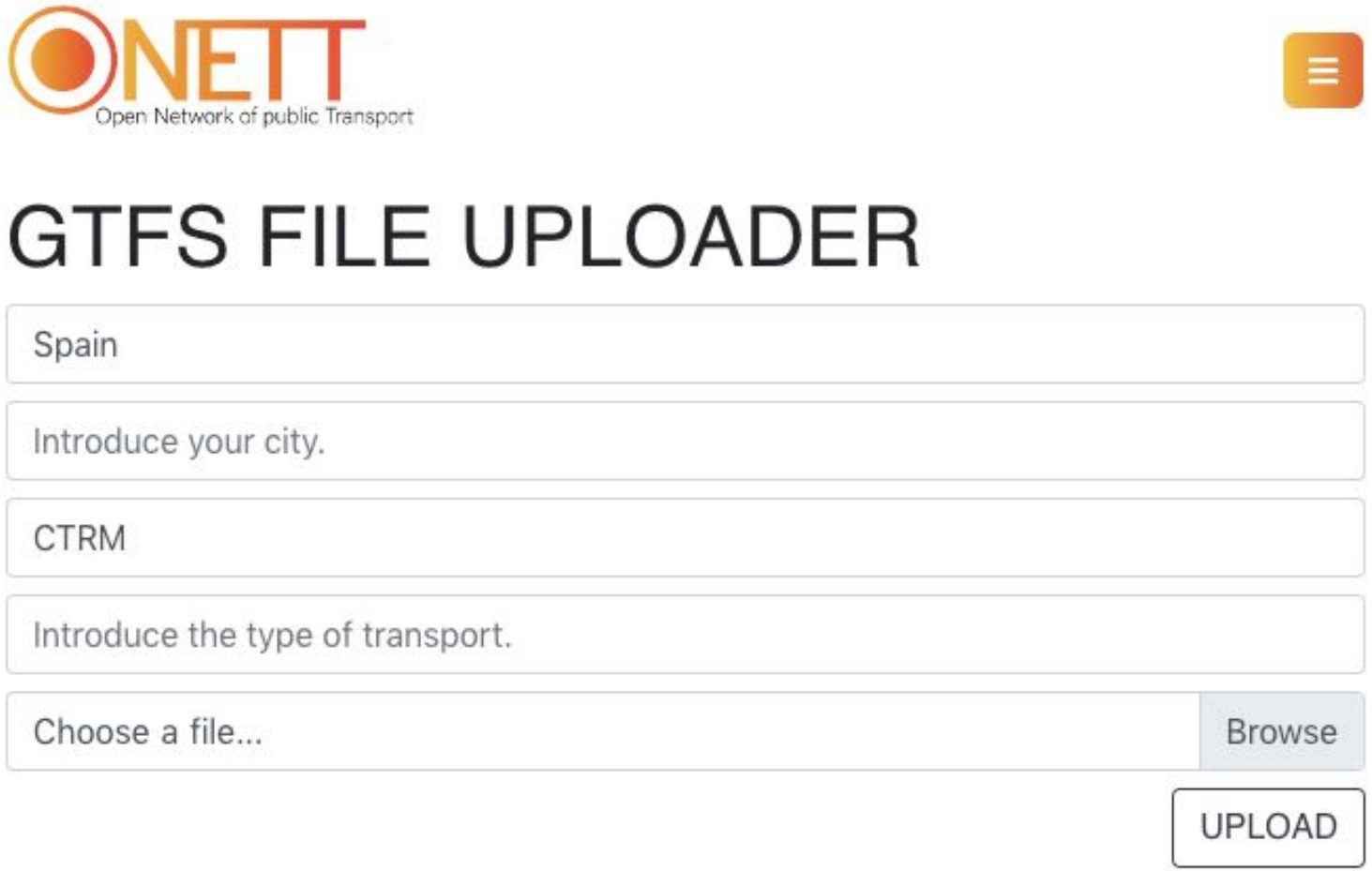

ONETT
Open Network of public Transport
GTFS FEED
DATA ANALYZER
GTFS RML MAPPING
RDF
MAPPING TRANSLATION
RDF GENERATION

- Open and easy model for publishing transport data
- Complex model:
  - Multiple joins among CSV files (performance)
  - Optional files and fields (completeness)



RML: The RDF Mapping Language

- Support for multiple data sources (CSV, JSON, XML, etc.)
- Extension of the W3C recommendation R2RML
- De-facto standard for KG Generation from heterogeneous data sources
- Emergence of its use
  - See: http://rml.io/implementation-report/
- User-friendly serialisation YARRRML
  - See: http://rml.io/yarrrml/


ONETT
Open Network of public Transport
GTFS FILE UPLOADER
Spain
Introduce your city.
CTRM
Introduce the type of transport.
Choose a file...
Browse
UPLOAD

Data Analyzer & Mapping Translation

Original GTFS
YARRRML mapping

```
prefixes:
  trm: https://w3id.org/transmodel/terms#
  trmo:https://w3id.org/transmodel/resource/
  dct: http://purl.org/dc/terms/
  foaf: http://xmlns.com/foaf/0.1/
  schema: http://schema.org/

mappings:
fare_rules:
    sources: ['../gtfs/fare_rules.csv~csv']
    s: trmo:fare_rules/$(fare_id)
    po:
      - [a, trm:FareProduct]
      - [trm:GroupOfLinesRef , $(route_id)]
      - [trm:AuthorityRef , $(route_id)]
      - [trm:StartTariffZoneRef , $(origin_id)]
      - [trm:EndTariffZoneRef , $(destination_id)]
      - [trm:TariffZoneRef , $(contains_id)]
```

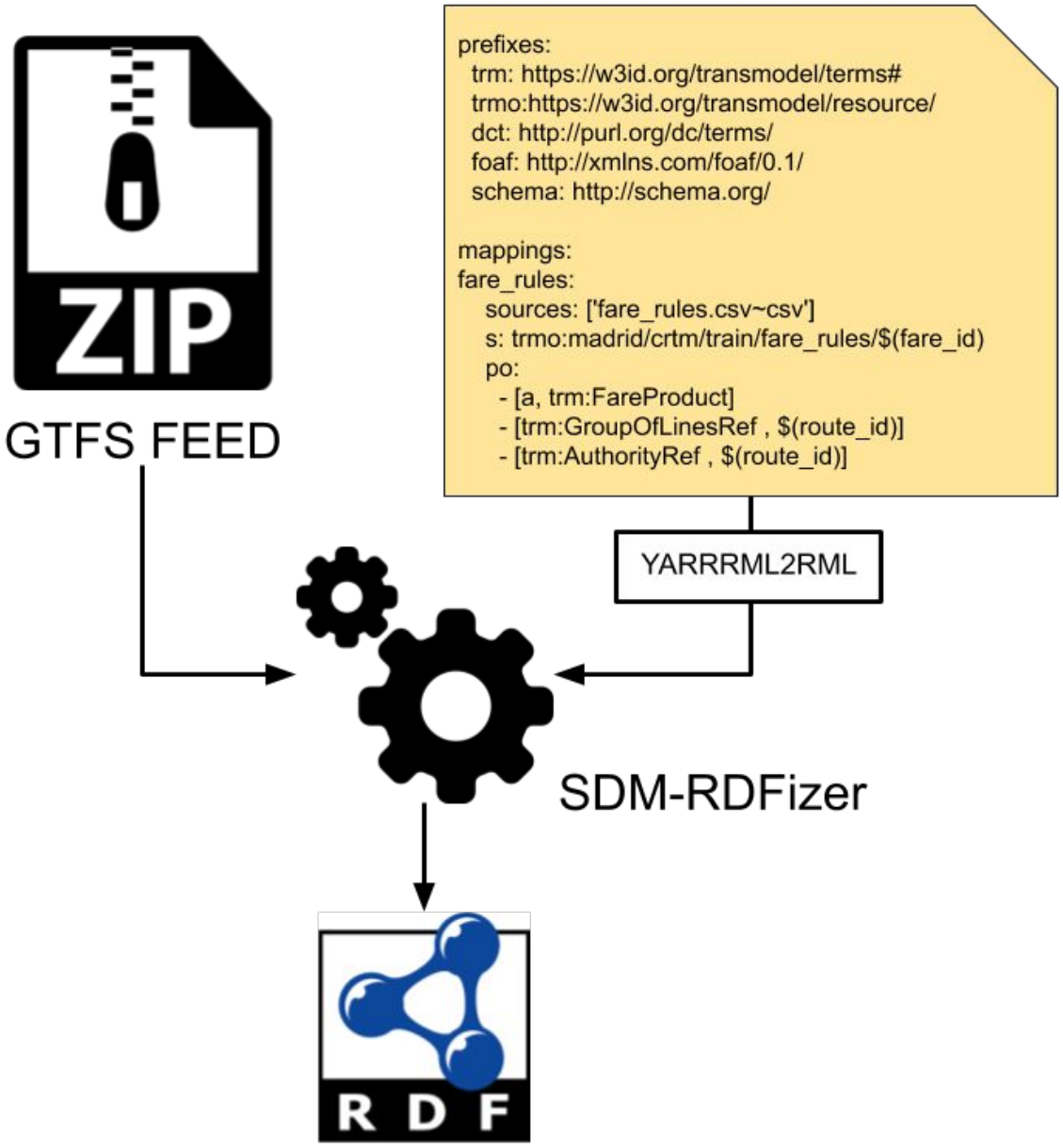
Translated GTFS
YARRRML mapping

```
prefixes:
  trm: https://w3id.org/transmodel/terms#
  trmo:https://w3id.org/transmodel/resource/
  dct: http://purl.org/dc/terms/
  foaf: http://xmlns.com/foaf/0.1/
  schema: http://schema.org/

mappings:
fare_rules:
    sources: ['../gtfs/fare_rules.csv~csv']
    s: trmo:madrid/crtm/train/fare_rules/$(fare_id)
    po:
      - [a, trm:FareProduct]
      - [trm:GroupOfLinesRef , $(route_id)]
      - [trm:AuthorityRef , $(route_id)]
      - [trm:StartTariffZoneRef , $(origin_id)]
      - [trm:EndTariffZoneRef , $(destination_id)]
      - [trm:TariffZoneRef , $(contains_id)]
```

Data Analysis + Mapping Translation

RDF Generation



Online resources

Website: https://osoc-es.github.io/onett/

Application: https://snap.summerofcode.es

Paper (HTML): https://osoc-es.github.io/onett-paper/output/

Code: https://github.com/osoc-es/?q=onett

SNAP project: https://www.snap-project.eu/

SDM-RDFizer: https://github.com/SDM-TIB/SDM-RDFizer

Open Summer of Code Spain 2019

Conclusions

- Declarative solution > maintainability
- On the fly generation of mappings (1st approach)
- Adaptability over the heterogeneity of the GTFS model
- Efficient KG generation
- Use case for the KGC W3C community group

Future work:

- Transformation to NeTEx (using mappings)
- Fare recommendation system
- Quantitative evaluation of ONETT
- Integration in commercial product powered by SNAP

ONETT

Systematic Knowledge Graph Generation for NAP

David Chaves-Fraga, Ontology Engineering Group
Universidad Politécnica de Madrid, Spain
Adolfo Antón, OEG-UPM
Jhon Toledo, OEG-UPM
Oscar Corcho, OEG-UPM

dchaves@fi.upm.es
@dchavesf

09/09/2019
Sem4Tra@SEMANTiCS2019