Generating linked RDF data from heterogeneous streaming and archival data sources: Populating the datAcron ontology

Georgios M. Santipantakis, George A. Vouros, Apostolos Glenis, Christos Doulkeridis, Akrivi Vlachou

University of Piraeus, Greece

{gsant, georgev, cdoulk}@unipi.gr, apostglen46@gmail.com, avlachou@aueb.gr

SEMANTiCS
Amsterdam 2017

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS

UNIVERSITY OF PIRAEUS
RESEARCH CENTER

Introduction

This work makes the following contributions:

- Proposes a framework for generating RDF from heterogeneous sources using graph templates as a generic way to map data to RDF, supporting easy tuning of the generation process and verification of the RDF data generated; also enabling cross-data source link discovery as RDF data is generated;

- Provides experimental results demonstrating the computational efficiency of the proposed method using real-world data sources

Requirements

Requirements the proposed framework was called to fulfill:

- support for a wide variety of data formats, in a unique workflow,

- easily connecting to new data sources, and extending to new data formats,

- easily validating the RDF output,

- capable to consume streams of data,

- fast processing data to RDF triples,

- as generic as possible, with no dependency to any vocabulary,

- support communication between concurrent RDFization workflows, for the construction of triples "on demand",

- (re)using user defined functions, for data conversion and linking
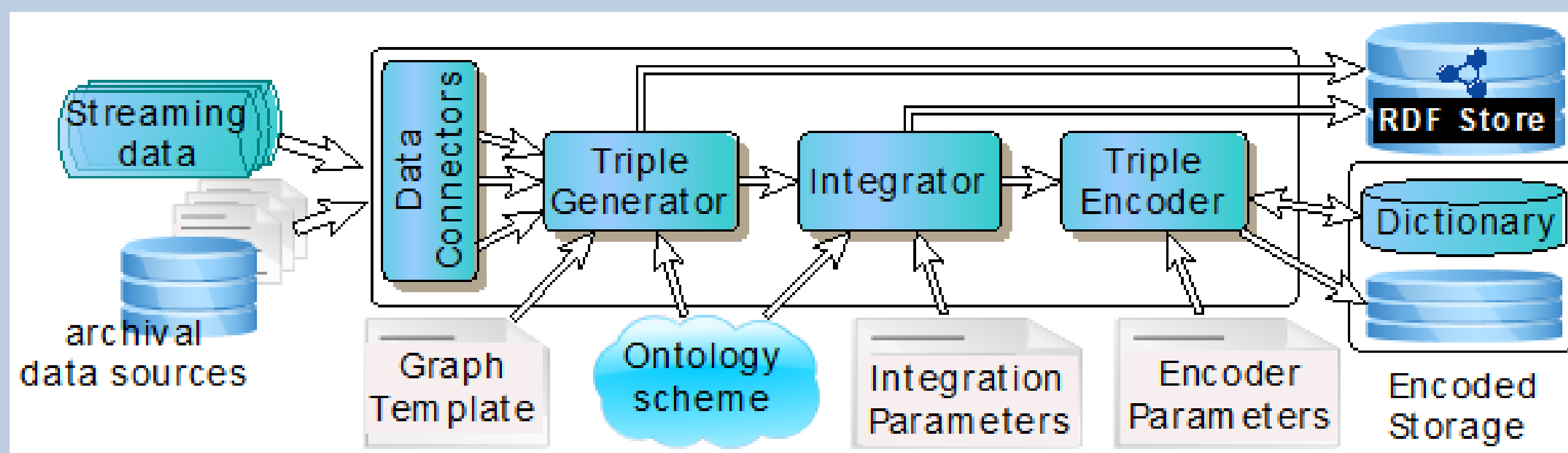
Results on RDF generation
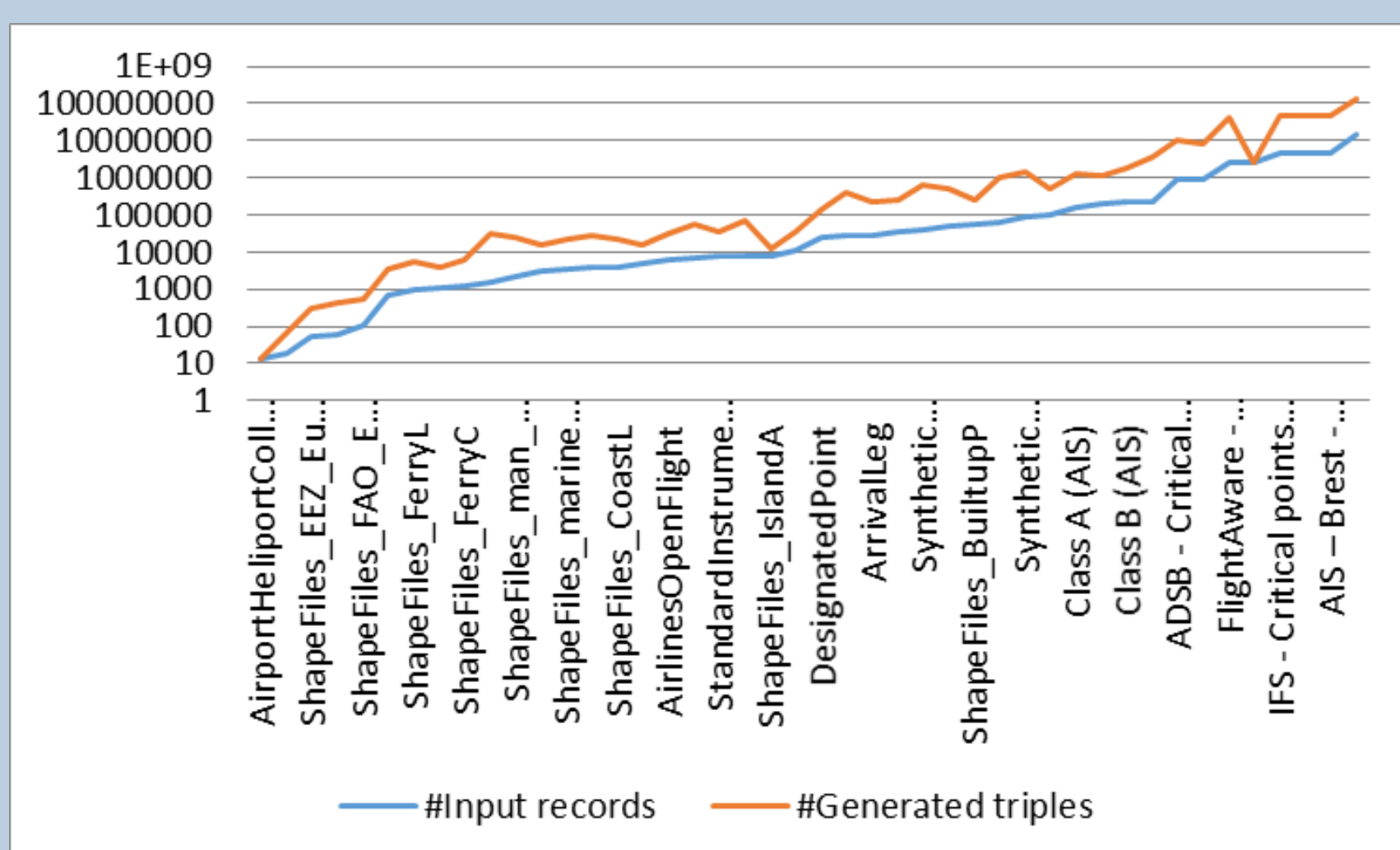
Number of input records (blue line) and generated triples (red line) per data source, ordered by number of input records in the data source
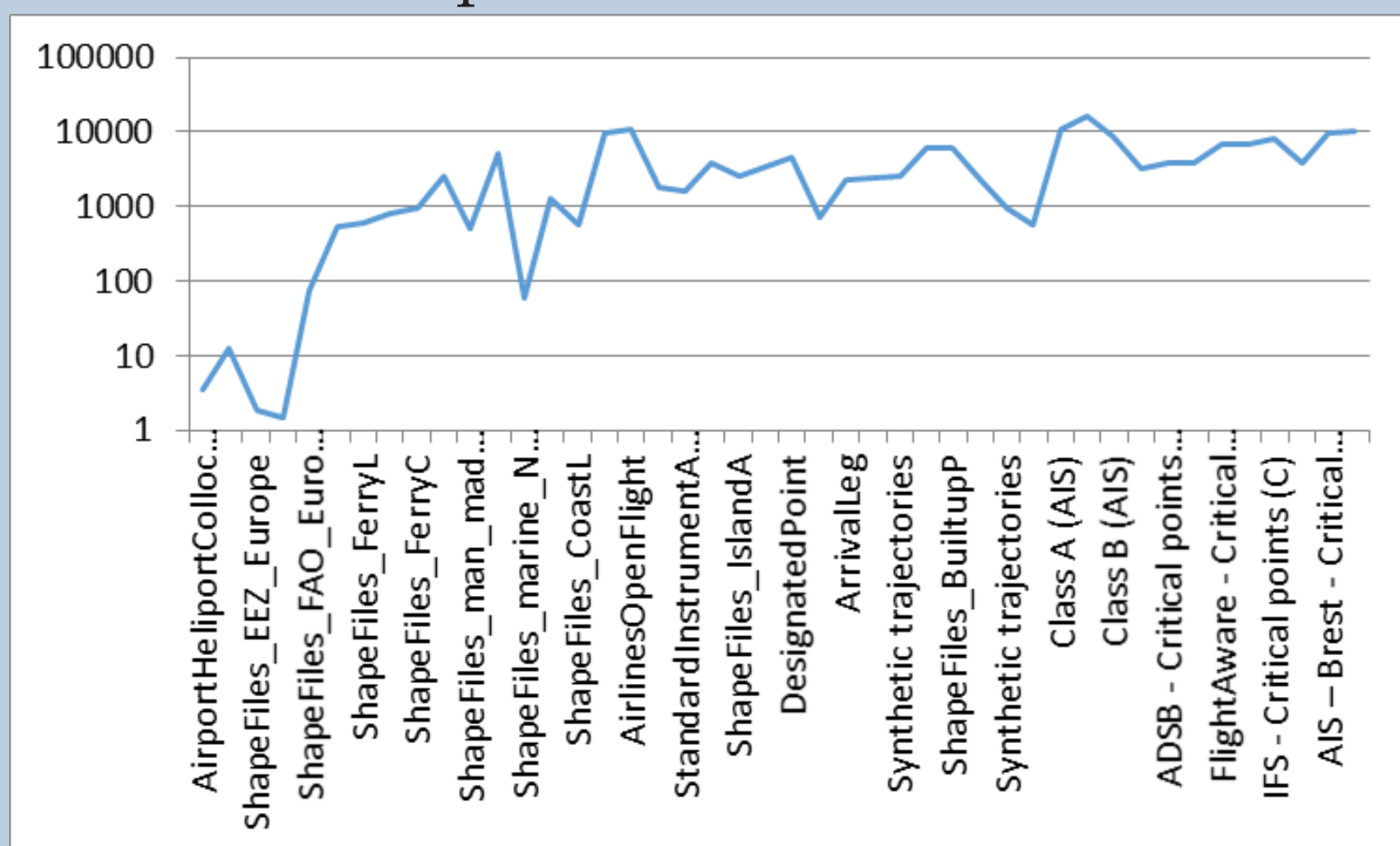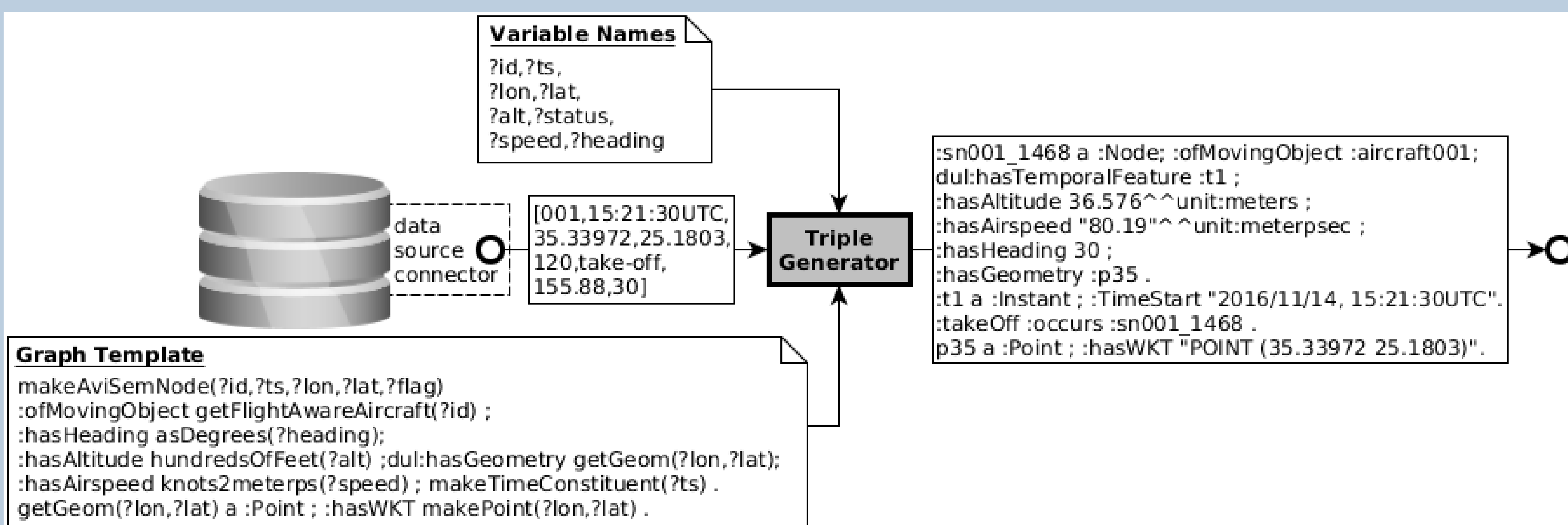

Number of generated triples per second for each data source, ordered by number of input records in the data source
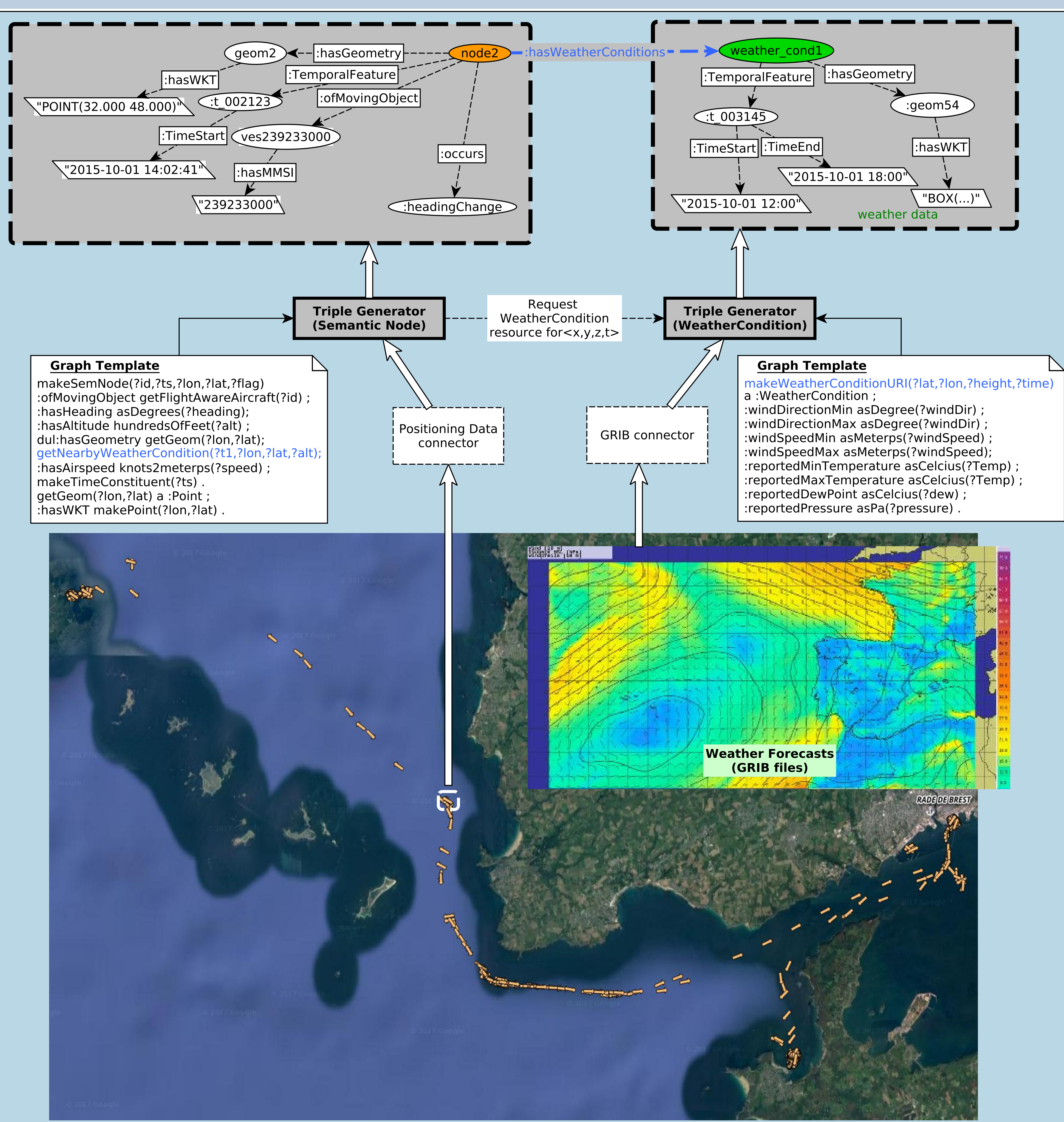
The overall RDF generation

Triple generator

Example

Acknowledgements