A workflow to Convert Live atmospheric Sensor Data into Linked Data

Jiantao Wu\textsuperscript{1,2}, Fabrizio Orlandi\textsuperscript{1,3}, Declan O’Sullivan\textsuperscript{1,3}, Soumyabrata Dev\textsuperscript{1,2}

July 19, \textbf{IGARSS 2022}

\textsuperscript{1}The ADAPT SFI Research Centre, Ireland

\textsuperscript{2}School of Computer Science, University College Dublin

\textsuperscript{3}School of Computer Science and Statistics, Trinity College Dublin, Ireland
1. Motivation & Objective

2. Methodology

3. Future Work
Motivation

Most of live atmospheric data coming from REST APIs is organized with structured data formats such as JSON, XML, and HTML tables. The heterogeneous data formats has resulted minimal interoperability amongst the live datasets\footnote{Andrea Mauri et al. “Where are the rdf streams?: Deploying rdf streams on the web of data with triplewave”. In: 2015}.

- **Non-semantified streams**
  - Bulked data sources
  - Various formats, such as JSON, XML, and HTML tables.
  - Larger client-side preprocessing efforts for data heterogeneity

- **RDF streams**
  - Semantic interoperability, \textit{e.g.} interoperable schema
  - Linked Data techniques enabled, \textit{e.g.} different streams are findable to each other
  - Larger server-side preprocessing efforts for data heterogeneity
**Objective**

To enhance the live atmospheric data towards higher interoperability enabling semantic analysis of live atmospheric data streams, such as investigating the influence of human activities on atmosphere.

1. **We have**
   1. implemented the novel RDF stream data model to create a workflow allowing the conversion of the API-based live atmospheric data to Linked Data, **and**
   2. first developed a Java utilities integrated in the workflow to automate the conversion

---

1 In the spirit of reproducible research, all the source code is available at [https://github.com/futaoo/RDFStream](https://github.com/futaoo/RDFStream).
1. Motivation & Objective

2. Methodology

3. Future Work
Workflow overview

**Figure:** From a JSON stream to a RDF stream

**Strength:**

- **RML mapping**
  - Semantically interoperable mapping assertions
  - Support for multiprocessing ([https://rml.io/](https://rml.io/))

- **Automatic semantics enrichment**
  - Semantic interface — using **Fno** ontology ([https://fno.io/](https://fno.io/)) — to Java functions designed for atmospheric data
PurpleAir live air quality data conversion

Original timestamp data

- isolated to other data
- less readable, e.g. non-unit
- compact and minimal for storage

Timestamped knowledge graph

- Semantics rich
- Flexible to extend, i.e. Linked Data
- Human readable
1. Motivation & Objective

2. Methodology

3. Future Work
We will work on:

- continuing the expansion of generic functions that cater to the processing of atmospheric sensor data,
- providing continuous query answering capabilities for consumers interested in event-level information obtained from RDF sensor streams, and
- providing users with the option to specify ontological rules that will make their data more intelligent.