An Interoperable Open Data Portal for Climate Analysis

Jiantao Wu\textsuperscript{1,2} Huan Chen\textsuperscript{2} Fabrizio Orlandi\textsuperscript{1,3} Yee Hui Lee\textsuperscript{4} Declan O’Sullivan\textsuperscript{1,3} Soumyabrata Dev\textsuperscript{1,3}

\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
\textsuperscript{4}School of Electrical and Electronic Engineering, Nanyang Technological University (NTU), Singapore

November 8, 2021, AP-S/URSI
Outline

1 Motivation & Objective

2 Methodology
   - Graph Data Schema—“Climate Analysis (CA)” Ontology
   - Interlink data sources
   - Triple Store—The Linked Data Publishing Tool
   - Data Visualization
   - All-in-one open data portal

3 Future Work
Motivation & Objective

Methodology
- Graph Data Schema—“Climate Analysis (CA)” Ontology
- Interlink data sources
- Triple Store—The Linked Data Publishing Tool
- Data Visualization
- All-in-one open data portal

Future Work
Motivation—the problems we’re addressing

- Vast amounts of climatic data are needed for data-driven technologies in understanding climate issues such as global warming, energy shortage...
- Despite that many climate data are already available across the web or other medium, those data are separate and are often in different formats\(^1\), e.g. CSV, JSON, RDB...
- A neglect of data relationships\(^2\), for example, CO\(_2\) emission is relevant to air temperature.

---

\(^1\) F. Orlandi, A. Meehan, M. Hossari, S. Dev, D. O’Sullivan and T. Alskauf, “Interlinking Heterogeneous Data for Smart Energy Systems,” in *Proc. International Conference on Smart Energy Systems and Technologies (SEST)*, 2019.

\(^2\) S. Manandhar, S. Dev, Y. H. Lee, Y. S. Meng, and S. Winkler, “A Data-Driven Approach for Accurate Rainfall Prediction,” in *IEEE Transactions on Geoscience and Remote Sensing*, 2019.
Objective of the study

- Facilitating the integration of various climate data sources in UK and Ireland
- Creating a knowledge graph data portal to encourage the use of climate knowledge graph data, which is more interoperable and easier to be extended widely on the web.
Outline

1. Motivation & Objective

2. Methodology
   - Graph Data Schema—“Climate Analysis (CA)” Ontology
   - Interlink data sources
   - Triple Store—The Linked Data Publishing Tool
   - Data Visualization
   - All-in-one open data portal

3. Future Work
Table of Contents

1 Motivation & Objective

2 Methodology
   - Graph Data Schema—“Climate Analysis (CA)” Ontology
   - Interlink data sources
   - Triple Store—The Linked Data Publishing Tool
   - Data Visualization
   - All-in-one open data portal

3 Future Work
All begin with the definitions

CA ontology\(^3\) defines Classes, Properties, Relations, etc. for climate observations using RDF data structure (i.e. the triples).

Here are some examples:

- \(<\text{ca} : \text{TemperatureObservation}>\text{ca} : \text{TemperatureObservation}> <\text{rdf} : \text{type}> <\text{rdfs} : \text{Class}>\);  
  \(<\text{ca} : \text{TemperatureObservation}>\text{ca} : \text{TemperatureObservation}> <\text{rdfs} : \text{subClassOf}> <\text{sosa} : \text{Observation}>\)  
- \(<\text{ca} : \text{isLocatedIn}>\text{ca} : \text{isLocatedIn}> <\text{rdf} : \text{type}> <\text{rdf} : \text{Property}>\)  
- \(\ldots\)  

\(^3\) J. Wu, F. Orlandi, D. O’Sullivan and S. Dev, “An Ontology Model for Climatic Data Analysis,” in Proc. IEEE IGARSS conference, 2021.
Table of Contents

1 Motivation & Objective

2 Methodology
   - Graph Data Schema—“Climate Analysis (CA)” Ontology
   - Interlink data sources
   - Triple Store—The Linked Data Publishing Tool
   - Data Visualization
   - All-in-one open data portal

3 Future Work
Linked data is data which contains links to other datasets.

- We build the climate knowledge graph model on NOAA climate data, and reuse vocabularies from other ontologies.
- The URIs are the identifiers of the data, which are resolvable to discover more facts.
- The reuse of other ontologies provide the context without loss of any semantic information.

---

4. C. Bizer, T. Heath, and T. Berners-Lee, “Linked data: The story so far,” in *Semantic services, interoperability and web applications: emerging concepts*, 2019.
Table of Contents

1 Motivation & Objective

2 Methodology
   - Graph Data Schema—“Climate Analysis (CA)” Ontology
   - Interlink data sources
   - Triple Store—The Linked Data Publishing Tool
   - Data Visualization
   - All-in-one open data portal

3 Future Work
Apache Jena Fuseki

- Our knowledge graph is stored using Apache Fuseki.
- Fuseki expose our data as open data on the web. The data is retrievable over HTTP via URIs.
- The query language to the triplestore is the standard SPARQL.

**Figure:** A screenshot of our configured fuseki server
Table of Contents

1 Motivation & Objective

2 Methodology
   - Graph Data Schema—“Climate Analysis (CA)” Ontology
   - Interlink data sources
   - Triple Store—The Linked Data Publishing Tool
   - Data Visualization
   - All-in-one open data portal

3 Future Work
**URI dereferencing visualized with LodView**

- Users can make queries to the fuseki server and click on the URI-encoded results to obtain detailed information about a specific entity.
- If an item in the profile is URI-encoded, it can be retrieved for more facts regarding the URI.

---

**Figure:** A screenshot of dereferenced Ireland's climate station

---

5 [https://github.com/LodLive/LodView](https://github.com/LodLive/LodView)
Exploring the knowledge graph visually with LodLive

- Graphical representation is accumulative when users explore the knowledge graph.
- More information can be seen when nodes are expanded.
- The structure of the knowledge graph is clear.

Figure: The graphical representation of dereferencing navigation.

6https://github.com/LodLive/LodLive

6 Jiantao WU (UCD Computer Science)
Table of Contents

1 Motivation & Objective

2 Methodology
   - Graph Data Schema—“Climate Analysis (CA)” Ontology
   - Interlink data sources
   - Triple Store—The Linked Data Publishing Tool
   - Data Visualization
   - All-in-one open data portal

3 Future Work
Methodology

All-in-one open data portal

Link-Climate project

- Link-Climate open data portal provide the entrance to all the aforementioned technologies.
- Instructions on our CA ontology and the manual on how to manipulate the knowledge graph are integrated in the portal.

Figure: A screenshot of the open data portal
Outline

1 Motivation & Objective

2 Methodology
- Graph Data Schema—“Climate Analysis (CA)” Ontology
- Interlink data sources
- Triple Store—The Linked Data Publishing Tool
- Data Visualization
- All-in-one open data portal

3 Future Work
Future Work

- A continuous extension to the types of the classes and properties in CA ontology, catering for different datasets
- After having a rich set of vocabularies, define some climate domain-specific rules to bring the CA ontology an inference ability (i.e. exploring the knowledge graph applications)
- Leveraging the power of Linked Data to a certain degree by integrating CA ontology with other published ontologies in order to enable users to have some knowledge on how climate influences other domains.
Thank You

Please feel free to reach me at email: jiantao.wu@ucdconnect.ie