WeKG-MF: a Knowledge Graph of Observational Weather Data

Nadia Yacoubi Ayadi, Catherine Faron, Franck Michel, Fabien Gandon, Olivier Corby

WeKG-MF:
a Knowledge Graph of Observational Weather Data

Nadia Yacoubi Ayadi[0000-0002-6132-8718],
Catherine Faron[0000-0001-5959-5561], Franck Michel[0000-0001-9064-0463],
Fabien Gandon[0000-0003-0543-1232], and Olivier Corby[0000-0001-6610-0969]

University Côte d'Azur, Inria, CNRS, I3S (UMR 7271), France
{nadia.yacoubi-ayadi,fabien.gandon,olivier.corby}@inria.fr
{faron,fmichel}@i3s.unice.fr

Abstract. In this paper, we present the WeKG-MF Knowledge Graph constructed from open weather observations published by Météo-France institution. WeKG-MF relies on a semantic model that formalizes knowledge about meteorological observational data. The model is generic enough to be adopted and extended by meteorological data providers to publish and integrate their sources while complying with Linked Data principles. WeKG-MF offers access to a large number of meteorological variables described through spatial and temporal dimensions and thus has the potential to serve several scientific case studies from different domains including agriculture, agronomy, environment, climate change and natural disasters.

Keywords: Knowledge Graph · Semantic Modelling · Observational Data · Linked Data · Meteorology.

1 Introduction

Meteorological data are crucial for many application domains. They typically include measurements of several weather parameters such as wind direction and speed, air pressure, rainfall, humidity and temperature. However, these data are most commonly collected and stored separately in different files using a tabular data format that lacks explicit semantics, which impedes their integration and sharing to serve researchers from different domains such as agriculture, climate change studies or natural disaster monitoring. A typical approach in integrating and publishing such data is to formalize a knowledge graph relying on linked data and semantic Web standard models and practices. To deal with the complexity of the knowledge domain to be modelled, we adopted the SAMOD agile methodology [6]. The SAMOD process is initiated by a motivating scenario that leads to a set of competency questions that, in turn, provide requirements on the knowledge graph model. As output, we designed a semantic model in which meteorological variables are semantically defined and described at a fine grained level, including aspects of time, location, units of measurement, etc. The WeKG-MF knowledge graph, constructed from open weather observations published by Météo-France, is compliant with the proposed semantic model. The first release of the WeKG-MF includes weather observations from January 2019 till December 2021. The paper is structured as follows. Section 2 presents an overview of WeKG semantic model and highlights its design principles. Section 3 presents the RDF-based knowledge graph WeKG-MF constructed from weather data archives of Météo-France and illustrates how it serves use cases identified in the context of the D2KAB French research project[1].

2 Semantic Model for Weather Data

In order to propose a self-contained model for representing and publishing meteorological data, we extend the SOSA/SSN [4,3] ontologies with three new classes. First, `weo:MeterologicalObservation` is the core class of our model; it supports the description of a single, atomic observation that is related to a particular feature of interest, instance of the `weo:MeteorologicalFeature` class, and an observable property, instance of class `weo:WeatherProperty`. These three classes specialize classes from the SOSA/SSN ontologies [4,3] as reflected by their formal definitions. These definitions express that only one weather property and one meteorological feature is used for a given meteorological observation:

$$
\begin{aligned}
weo\!:\!MeteorologicalFeature \equiv\ & sosa\!:\!FeatureOfInterest \sqcap \\
& \forall\ ssn\!:\!hasProperty.weo\!:\!WeatherProperty \sqcap \\
& \geq 1\ ssn\!:\!hasProperty.weo\!:\!WeatherProperty
\end{aligned}
$$

$$
\begin{aligned}
weo\!:\!WeatherProperty \equiv\ & sosa\!:\!ObservableProperty \sqcap \\
& \forall ssn\!:\!isPropertyOf.weo\!:\!MeteorologicalFeature \sqcap \\
& \geq 1\ ssn\!:\!isPropertyOf.weo\!:\!MeteorologicalFeature
\end{aligned}
$$

$$
\begin{aligned}
weo\!:\!MeteorologicalObservation \equiv\ & sosa\!:\!Observation \sqcap \\
& \forall sosa\!:\!observedProperty.weo\!:\!WeatherProperty \sqcap \\
& = 1\ sosa\!:\!observedProperty \sqcap \\
& \forall sosa\!:\!hasFeatureOfInterest.weo\!:\!MeteorologicalFeature \sqcap \\
& = 1\ sosa\!:\!hasFeatureOfInterest
\end{aligned}
$$

We reused the Value Sets[2](VP) ontology design pattern and, because we can enumerate the values, this led us to define a SKOS vocabulary whose concepts are instances of `weo:WeatherProperty` or `weo:MeteorologicalFeature` and represent the possible values of observable properties and features of interest. Inline with Linked Data best practices, we aligned the weather properties of our SKOS vocabulary with terms from the NERC Climate and Forecast Standard Names vocabulary[3]. To avoid redundancies of measurements units among observations, we define for each SKOS weather property an applicable unit re-using QUDT Unit vocabulary[4]. Thus, observation results are modelled as literals and an observation is linked to its result by RDF property `sosa:hasSimpleResult`. Finally, since observable properties of our vocabulary are also defined as instances of the `qudt:QuantityKind` class, we aligned them with terms from the QUDT Quantity Kind vocabulary. Figure 1 presents an RDF graph of a meteorological observation relative to the wind feature of interest and reporting the average wind speed observable property.

[1] https://www.d2kab.org/
[2] https://www.w3.org/TR/swbp-specified-values/
[3] http://vocab.nerc.ac.uk/collection/P07/
[4] http://qudt.org/vocab/quantitykind/

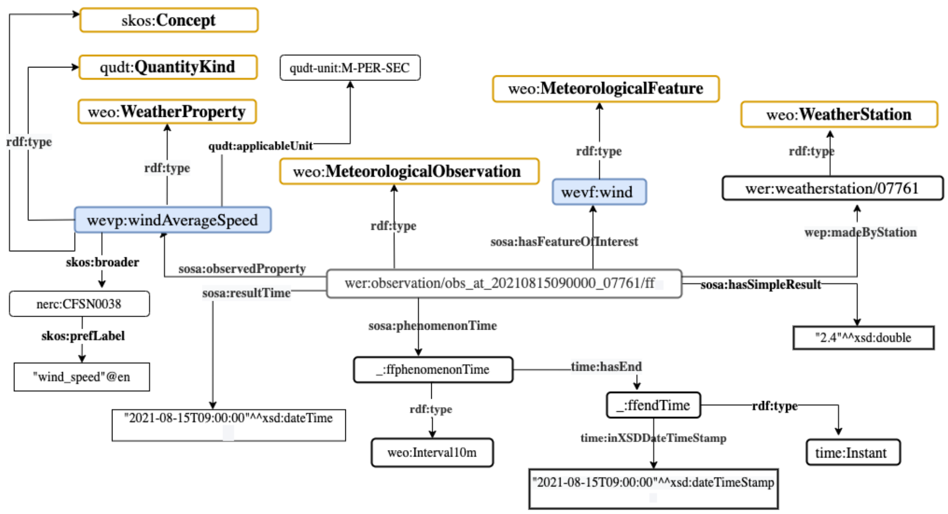

Fig. 1. Example of a meteorological observation reporting the Wind Speed Average property

Our semantic model deals with both temporal and spatial dimensions of meteorological observations. Along temporal dimension, the model captures the instant or interval at/during which the weather parameter is measured. Since duration of time intervals are described in the official documentation of the World Meteorological Organization (WMO)[5], we defined different time interval classes, based on the `time:Interval` class [2], by expressing OWL restrictions on their duration that may be declared in seconds, minutes or hours. The interest of doing this is that these time intervals are declared once in our semantic model and are reused for all observations, and thus avoid substantial redundancy. In Figure 1, the `wevp:windAverageSpeed` weather property is measured during a period of 10 minutes. This is denoted by `sosa:phenomenonTime` property whose value is an instance of `weo:Interval10m` class, while the end time of the interval is an instance of class `time:Instant`.

Meteorological observations are provided by weather stations which spatial information, such as longitude, latitude and altitude. We introduce the `weo:WeatherStation` class as a subclass of the `geosparql:Feature` class and `sosa:Platform`. According to GeoSPARQL vocabulary [1], each instance of `weo:WeatherStation` has a geometry with specific coordinates which enables us to query weather observations based on spatial information. The OWL version of our semantic model as well as the related SKOS vocabularies are available in our Github repository[6]. The prefixes of ontologies and vocabularies reused or introduced in this paper are listed in the repository's README[7].

[5] https://public.wmo.int/en/
[6] https://github.com/Wimmics/d2kab/tree/main/meteo/ontology
[7] https://github.com/Wimmics/d2kab/tree/main/meteo

3 WeKG-MF: Weather Knowledge Graph - Météo-France

We constructed the *WeKG-MF* knowledge graph according to the model presented in Section 2 from open weather observations published by Météo-France[8]. We first downloaded from Météo-France's portal the list of SYNOP weather stations in GeoJSON format, as well as the monthly observation reports generated by these stations as CSV files. Then, we implemented a reproducible pipeline to generate WeKG-MF in compliance with the proposed semantic model where the mapping is performed by the Morph-xR2RML tool [5]. The first version of WeKG-MF covers the period from January 2019 to the end of December 2021. Key statistics of the dataset are presented in Table 1.

Category	
Downloadable RDF dump	https://doi.org/10.5281/zenodo.5925413
Total Nr. of triples	62.306.102
Nr. of Observations	8.335.258
Nr. of weather stations	62
Nr. of weather properties	22
Nr. of meteorological features	6
Nr. of Observations per weather property	$\approx$ 416.762
Nr. of links to Wikidata	92

Table 1. key statistics of the WeKG-MF dataset

To illustrate how observations in the WeKG-MF can be queried and aggregated, we first developed a set of SPARQL queries available on the Github repository of the project[9]. Beyond this, we initially started this work to address the needs of the D2KAB French project whose primary objective is to create a framework to turn agronomy and biodiversity data into semantically described, interoperable, actionable, and open knowledge. A preliminary needs analysis pointed to competency questions that potential users may want to get answers.We only present some of them due to space constraints. Experts in agronomy investigate the correlations between the development rate of plants and weather parameters. They are especially interested in comparing aggregated values of a weather parameter for the same period of time in the same geographic location across years, e.g. the Growing Daily Degrees (GDD) calculated from the daily average air temperature minus a certain threshold called base temperature.

We developed, together with the set of SPARQL queries, a Jupyter Notebook that demonstrate how the results of queries can be used to generate visualizations from the WeKG-MF knowledge graph. As an example, Figure 2 presents two plots. The first one shows daily cumulative precipitations measured at the '*Bordeaux-Merignac*' station and the second one shows the evolution of daily average temperature collected from weather stations located in the French region of '*Nouvelle Aquitaine*'. Each plot shows a comparison of aggregated values calculated based on two weather parameters (precipitation and air temperatures) available in the WeKG-MF knowledge graph. Several use cases in agronomy can benefit from pre-calculated spatio-temporal slices of observations. For this, our model supports the

[8] https://www.meteofrance.com/

[9] https://github.com/Wimmics/d2kab/tree/main/meteo/sparql-examples

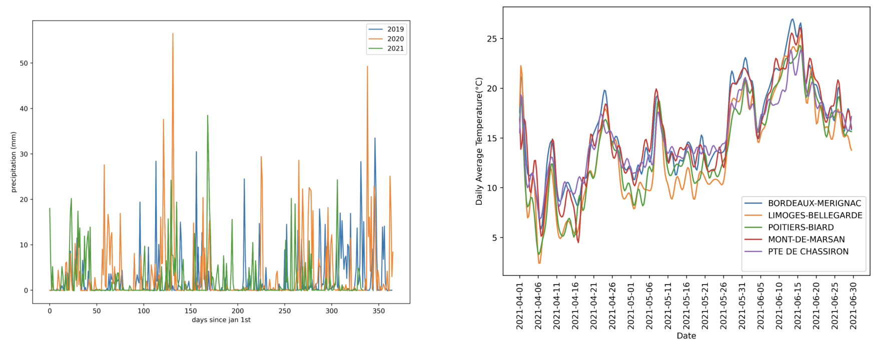

Fig. 2. Examples of Visualisation of Daily Precipitations and Average Temperature

definition of SPARQL CONSTRUCT queries with respect to a specific structure. Taking again the example of the GDD, we have defined SPARQL queries to construct a set of of slices of daily min., max. and avg. temperatures for each weather station and for each year.

4 Conclusion and Future Work

In this paper, we presented WeKG-MF, a knowledge graph of weather observational data generated from Météo-France's open data archives. It is based on a reusable, formal model exploiting and extending a network of well-known ontologies. In terms of sustainability, we released a fully automatic pipeline that enables anyone to generate and update the WeKG-MF graph over time with new data downloaded from Météo-France. In the short term, we will investigate identified use cases in the agronomy and agriculture domains in which our meteorological knowledge graph will be integrated with other knowledge sources in order to meet D2KAB project partners requirements.

References

1. Battle, R., Kolas, D.: Enabling the geospatial semantic web with parliament and geosparql. Semantic Web **3**(4), 355–370 (2012)
2. Cox, S., Little, C.: Time ontology in OWL (2020), https://www.w3.org/TR/owl-time/
3. Haller, A., Janowicz, K., Cox, S.J.D., Lefrançois, M., Taylor, K., Phuoc, D.L., Lieberman, J., García-Castro, R., Atkinson, R., Stadler, C.: The modular SSN ontology: A joint W3C and OGC standard specifying the semantics of sensors, observations, sampling, and actuation. Semantic Web **10**(1), 9–32 (2019)
4. Janowicz, K., Haller, A., Cox, S.J.D., Phuoc, D.L., Lefrançois, M.: SOSA: A lightweight ontology for sensors, observations, samples, and actuators. J. Web Semant. **56**, 1–10 (2019)
5. Michel, F., Djimenou, L., Faron-Zucker, C., Montagnat, J.: Translation of relational and non-relational databases into RDF with xR2RML. In: Proc. of the WEBIST 2015 Conference, Lisbon, Portugal. pp. 443–454. SciTePress (2015)
6. Peroni, S.: SAMOD: an agile methodology for the development of ontologies. In: Dragoni, M., Poveda-Villalón, M., Jimenez-Ruiz, E. (eds.) OWL: Experiences and Directions – Reasoner Evaluation. pp. 55–69. Springer (2016)