Multi-Level Visual Tours of Weather Linked Data

Nadia Yacoubi, Damien Graux and Catherine Faron

Inria, Université Côte d'Azur, CNRS, I3S, France

Abstract
The recent trend of adopting linked-data principles to integrate and publish semantically described open data using W3C standards has led to a large amount of available resources. In particular, meteorological sensor data have been uplifted into public weather-focused RDF graphs, such as WeKG-MF which offers access to a large set of meteorological variables described through spatial and temporal dimensions. Nevertheless, these resources include huge numbers of raw observations that are tedious to explore by lay users. In this article, we aim at providing them with visual exploratory "tours", benefiting from RDF data cubes to present high-level aggregated views together with on-demand fine-grained details through a unified Web interface.

Keywords
Weather data, Spatio Temporal data, Visualisation, RDF Knowledge Graphs, SPARQL endpoints

1. Introduction

The recent trend of adopting linked-data principles to integrate and publish semantically described open data using W3C standards has led to a large amount of available resources. In particular, meteorological sensor data have been uplifted into public weather-focused RDF graphs, such as WeKG-MF graph which offers access to a huge number of sensor observations related to different weather variables, described through spatio-temporal dimensions. Hence, supporting lay users to browse, analyze, consume and reuse sensor data transformed and integrated into LOD datasets is challenging. In this article, we present the first release of a Web interface that enables users to visualize weather observational data at different levels of spatio-temporal granularity. We show how the WeKG-MF principles and the adoption of RDF data cubes can provide users with visual multi-level "tours". Our main objective is to provide users with interactive exploration means to navigate the WeKG-MF, leveraging RDF data cubes to present high-level aggregated views as well as fine details on demand through a unified Web interface.

2. The WeKG Spatio-Temporal Model

In this section, we present the WeKG spatio-temporal model and use-cases identified in the context of the D2KAB research project which highlighted the need to build a knowledge graph from historical records published as open data by the French weather service provider Météo-France.

VOILA! 2022: 7th International Workshop on Visualization and Interaction for Ontologies and Linked Data, October 23, 2022, Virtual Workshop, Hangzhou, China, Co-located with ISWC 2022.
nadia.yacoubi-ayadi@inria.fr (N. Yacoubi); damien.graux@inria.fr (D. Graux); catherine.faron@inria.fr (C. Faron)
https://dgraux.github.io/ (D. Graux); https://www.i3s.unice.fr/~faron/ (C. Faron)
0000-0002-6132-8718 (N. Yacoubi); 0000-0003-3392-3162 (D. Graux); 0000-0001-5959-5561 (C. Faron)

SPARQL endpoint → https://weakg.i3s.unice.fr/sparql			
Number of RDF Triples	123.413.015	Weather stations	62
Total Observations	16.433.031	Weather properties	22
Observations per weather property	≈ 416.762	Meteorological features	6
		Links to Wikidata	92

Table 1
Key Statistics of the WeKG-MF Knowledge Graph.

2.1. From weather Observations archives to an RDF Knowledge graph

In our previous work [1], we have presented a self-contained semantic model that re-uses and extends standard ontologies, among which the GeoSPARQL ontology for spatial features and their relations [2], the Time ontology [3] for temporal entities, the Sensor, Observation, Sample, and Actuator (SOSA) [4] and Semantic Sensor Network (SSN) ontologies [5] for sensors and observations, and the RDF Data Cube ontology [6] for aggregation and multidimensionality features. The WeKG model captures the semantics of atomic and fine-grained weather observations by reusing and extending SOSA classes as well as spatio-temporal time series of aggregate values using the RDF vocabulary of the data cube. The proposed 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.

Then, we built the WeKG-MF knowledge graph [7], based on this model, considering the open weather observations published by Météo-France[1]. The SPARQL WeKG-MF endpoint allows users to retrieve weather observations recorded every 3 hours by different sensors hosted by weather stations and related to different parameters (air temperature, humidity, wind speed, precipitation, atmospheric pressure, etc.). The Table 1 summarises some key statistics of WeKG-MF. WeKG-MF includes meteorological data from the period 2016-2021 and is continuously evolving to include new, newer and older data. The knowledge graph is intended to serve different use case scenarios in several domains, including agriculture, biodiversity and climate studies.

2.2. Use Case Scenarios for WeKG-MF

WeKG-MF was initially created to answer expert's needs in the context of the D2KAB French project[2]. A preliminary analysis revealed several competency questions that express the needs of experts to retrieve weather observations at different levels of granularity. For instance, an expert may be interested by the exact time of a day at which the minimum/maximum temperature was recorded and in this case, he is querying a fine-grained temporal entity represented by a XSD:DATETIMESTAMP literal in WeKG-MF. In several other situations, experts are more interested in aggregate values of some weather parameters, such as the daily total precipitation, the number of days with precipitation greater than 1 mm over a time period, or the monthly mean values of maximum, minimum and mean temperatures.

To address these needs, we reused the RDF Data Cube Vocabulary (DCV) [6] to create multidimensional RDF slices, that are pre-calculated by fixing temporal and spatial dimensions and by applying aggregation functions such as min/max/avg/sum on fine-grained observation. Thus,

[1] https://www.meteofrance.com/
[2] http://www.d2kab.org

```
SELECT distinct ?groupDate (SUM(?vp) as ?sum_precipitation) WHERE {
    ?obs a weo:MeteorologicalObservation;               sosa:hasSimpleResult ?vp;
    sosa:observedProperty wevp:precipitationAmount;   sosa:resultTime ?date;
    wep:madeByStation <http://ns.inria.fr/meteo/weatherstation/07434> .
    BIND (day(?date) as ?day)  BIND (month(?date) as ?month)  BIND (year(?date) as ?year)
    BIND (if (datatype(?year/4)=xsd:integer && ((?year/100)*100 != 0 ||
            (?year/400)*400 = 0) , 1, 0) as ?bissexYear)
    BIND ( if (?day = 1, if (?month in (1, 2, 4, 6, 8, 9, 11), 31,
               if (?month in (5, 7, 10, 12), 30,
                  if (?bissexYear = 1, 29, 28))), ?day - 1) AS ?previousDay)
    BIND (if (?day = 1, if (?month=1, 12, ?month - 1), ?month) as ?previousMonth)
    BIND (if (?day = 1 && ?month=1, ?year - 1, ?year) as ?previousYear)
    BIND ( xsd:date(if(hours(?date)<=6, concat(?previousYear, "-",
            if (?previousMonth<10, concat("0", ?previousMonth), ?previousMonth), "-",
            if (?previousDay<10, concat("0", ?previousDay), ?previousDay)), concat(?year, "-",
            if (?month<10, concat("0", ?month), ?month), "-",
            if (?day<10, concat("0", ?day), ?day)))) AS ?groupDate)
} GROUP BY ?groupDate ORDER BY ?groupDate
```

Figure 1: SPARQL Query for Daily values of Total Precipitation according to WMO documentation.

```
SELECT ?label_station ?date ?avg_temp WHERE {
    { # Query weather stations located in "Nouvelle Aquitaine" region.
      SELECT ?statURI ?label_station WHERE {
          ?statURI a weo:WeatherStation; rdfs:label ?label_station .
                   dct:spatial [ wdt:P131 [rdfs:label ?label ; wdt:P2585 '75']] . }
    }
    # Query slices for each statURI.
    VALUES ?year { "2021"^^xsd:gYear "2020"^^xsd:gYear "2019"^^xsd:gYear }
    ?slice a qb:Slice ; wes-dimension:station ?statURI ; wes-dimension:year ?year ;
       qb:observation [ a qb:Observation ; wes-attribute:observationDate ?date ;
              wes-measure:avgDailyTemperature ?avg_temp ] .                        }
```

Figure 2: Query to retrieve avg. daily temp. timeseries computed from the observation in WeKG-MF recorded by weather stations located in “Nouvelle Aquitaine" French region.

a set of observations which applies to a spatio-temporal dimensions (e.g. a region, a weather station, a year, a time interval) is represented by the DCV class QB:SLICE such as the attributes and measures attached to these observations are previously semantically described in a DCV QB:DATASTRUCTUREDEFINITION class. This class enables to represent the slice’s metadata along with the specification of dimensions, attributes, and measures. An example of a DSD definition of annual times series of min/max/avg air temperatures is available[3].

We have experimented different strategies to generate the RDF materialized slices according to a given DSD. A first strategy consists in relying on a unique SPARQL query of the CONSTRUCT form, enabling to create homogeneous RDF slices that include only aggregate values of one unique weather parameter (e.g., air temperature). A second strategy consists in combining several SPARQL queries of the SELECT form, whose results sets are integrated into the same slice. As

[3]https://github.com/Wimmics/weather-kg/blob/main/meteo/dataset-metadata/DSD-Definition.ttl

an example, Figure 1 illustrates the SPARQL query to calculate the total precipitation following the WMO documentation [8] which indicates that it is calculated for the day R_{day-j} as the accumulated precipitation of a specific day j from 6 am till 6 am of the following day j. Hence, six hours of the following UTC day shall be considered together with the current UTC day.

3. SPARQL-based Visual Tours

Aiming at simplifying the exploration of large RDF observational data available in WeKG-MF, we developed a Web application demo accessible at https://nadiaya2019.github.io/DemoKGViz/ using the D3 JavaScript library. The webpage provides different visualisations offering lay-users visual “tours" at different levels of granularity. Thanks to the WeKG spatio-temporal model and the incorporation of pre-calculated RDF slices, data retrieved from our SPARQL endpoint can be visualized with no additional transformations involved, while most approaches for Linked Data visualisation include pre-processing steps that can be time-consuming (see Section 4).

3.1. Retrieving Salient Information

In order to retrieve the WeKG-MF graph, we rely on three categories of SPARQL query patterns that could be easily adapted.

- An initial pattern allows to retrieve the Météo-France weather stations (and their geo-spatial coordinates) grouped by French regions.
- The second pattern follows up, retrieves materialised RDF slices and collects values of at least one aggregate parameter pre-calculated for each station over a period of time.
- The third pattern enables users to extract fine-grained, atomic observations based on results provided before. It provides detailed data giving information about aggregated value provenances.

3.2. Visual Features

From a graphical point-of-view, we have developed several features to enrich the WeKG-MF users' experience. As illustrated in Figure 3, we provide an interactive map allowing users to interact with the SPARQL endpoint by clicking on a French region (Figure 3.a). This action leads to the execution of an initial SPARQL query like the one depicted in Figure 2 that retrieves weather stations and their corresponding timeseries of a specific parameter, e.g., air temperature during 3 years (2019-2021). Markers indicating geo-spatial locations of weather stations are added to the map and the timeseries are represented through two interactive line charts, which x-axis represents time and y-axis represents the daily average air temperatures, while the color encodes the different weather stations for the selected region. The first chart (Fig.3.b – top-right) supports a brushing interaction allowing the user to select a time period to further explore the timeseries in the second chart (Fig.3.c – middle-right), which x-axis is updated according to the time selection. The brush selection is represented by a gray rectangle that can be resized at any time to expand/reduce the time span and by consequence update the view of the middle chart. Moreover, the brush selection can be handled through a click-and-drag movement to modify the time period while keeping the same time span. This chart (Fig.3.c – middle-right) supports interaction through a hovering technique, which displays a tooltip with detailed information

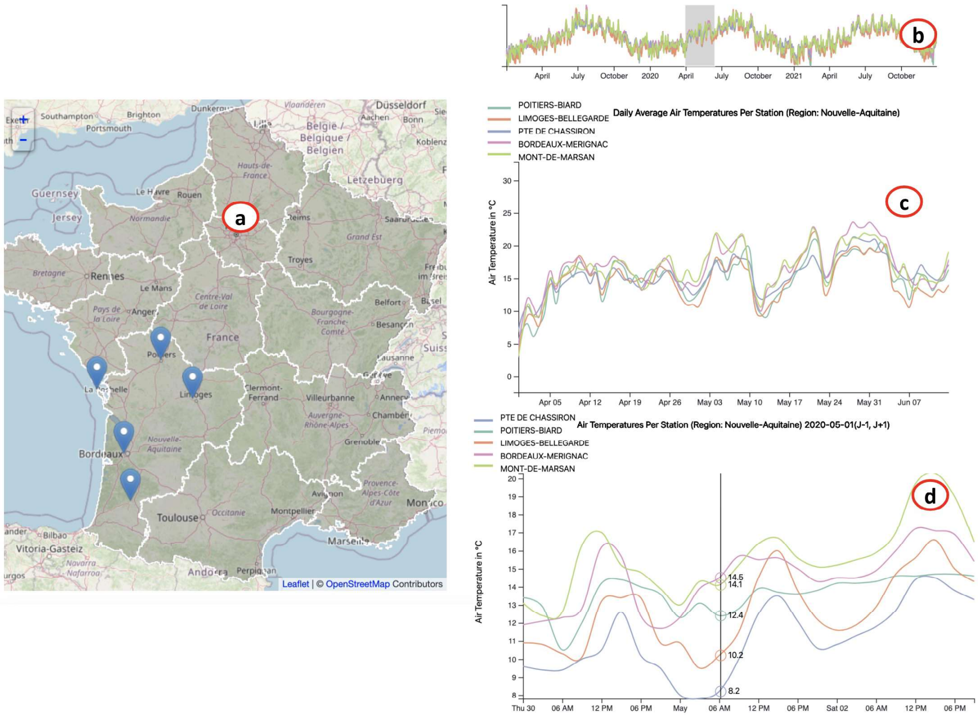

Figure 3: WeKG-MF exploration and navigation interface.

on the temperatures of a specific weather station and at the same time a third chart appears (Fig.3.d – bottom-right). Indeed, this chart offers a fine-grained view on WeKG-KG observations by displaying for a specific date $day - j$ un-aggregated atomic observations including those of the previous and following days. It supports also interaction through a hovering technique which displays a vertical line that the user could move to visualize values at the specific time of the day. Therefore, through this view, a user may easily explore the WeKG-MF knowledge graph from high-level aggregated timeseries to the elements from where the timeseries were calculated.

4. Related Work

Several research projects have focused on providing visualisation and exploration tools for LD datasets. Indeed, exploring, browsing and querying these datasets through space and time is very relevant for users but not straightforward for developers in order to transform RDF data into meaningful visualizations that suit users' needs. For an extensive review on LD exploration and visualization tools, we refer interested readers to [9]. While most existing approaches focus on how to shift pipelines to import/map/transform RDF data into data suitable for visualisations [10, 11], few of them highlight the importance of RDF modeling to easily support the generation of meaningful visualisations. Indeed, research works such as CubeViz [12] or OpenCube [13] aim to provide users with data cubes visualization and interactive analysis tools. However, to the best of our knowledge, multi-visualisation interfaces that combine high-level views on aggregated data using the RDF data cube vocabulary [6] and fine-grained views of un-aggregated values do not exist.

5. Conclusion and Future Works

We presented the first release of a Web application that offers interactive multi-level tours based on high-level aggregated views together with on-demand fine-grained data, and this through a unified multi-visualisations interface. In near future, we aim to work on a user evaluation study of our system to provide advanced analysis functionalities enabling experts to compare climatic conditions across geospatial and temporal dimensions. Moreover, we plan to enrich the interface to track data quality issues such as missing values across timeseries of weather parameters.

Acknowledgements. This study was carried out within the D2KAB project "From Data to Knowledge in Agronomy and Biodiversity", financed by the French National Research Agency (ANR-18-CE23-0017). We are also grateful to Aline Menin and Olivier Corby for their valuable help.

References

[1] N. Yacoubi Ayadi, C. Faron, F. Michel, F. Gandon, O. Corby, A model for meteorological knowledge graphs: Application to Météo-France observational data, in: ICWE, 2022.
[2] R. Battle, D. Kolas, Enabling the geospatial semantic web with parliament and geosparql, Semantic Web 3 (2012) 355–370.
[3] S. Cox, C. Little, Time ontology in OWL, 2020. URL: https://www.w3.org/TR/owl-time/.
[4] K. Janowicz, A. Haller, S. J. D. Cox, D. L. Phuoc, M. Lefrançois, SOSA: A lightweight ontology for sensors, observations, samples, and actuators, J. Web Semant. 56 (2019) 1–10.
[5] A. Haller, K. Janowicz, S. J. D. Cox, M. Lefrançois, K. Taylor, D. L. Phuoc, J. Lieberman, R. García-Castro, R. Atkinson, C. Stadler, The modular SSN ontology: A joint W3C and OGC standard specifying the semantics of sensors, observations, sampling, and actuation, Semantic Web 10 (2019) 9–32.
[6] D. Reynolds, R. Cyganiak, The RDF Data Cube Vocabulary, W3C Recommendation, W3C, 2014. URL: https://www.w3.org/TR/2014/REC-vocab-data-cube-20140116/.
[7] N. Yacoubi Ayadi, C. Faron, F. Michel, F. Gandon, O. Corby, WeKG-MF: A knowledge graph of observational weather data, in: ESWC 2022 Satellite Events, 2022.
[8] World Meteorological Organization., Handbook on climat and climat temp reporting, 2009. URL: https://library.wmo.int/doc_num.php?explnum_id=9253.
[9] L. Po, N. Bikakis, F. Desimoni, Linked Data Visualization Tools, 2020, pp. 47–72.
[10] A. Menin, C. Faron, O. Corby, C. Freitas, F. Gandon, M. Winckler, From linked data querying to visual search: Towards a visualization pipeline for LOD exploration, in: Proceedings of the 17th International Conference on Web Information Systems and Technologies, 2021.
[11] H. Raissya, F. Darari, F. Ekaputra, VizKG: A framework for visualizing SPARQL query results over knowledge graphs, volume 3023, CEUR-WS, 2021, pp. 95–102.
[12] M. Martin, K. Abicht, C. Stadler, S. Auer, A.-C. Ngonga Ngomo, T. Soru, CubeViz – exploration and visualization of statistical linked data, in: WWW, 2015.
[13] E. Kalampokis, A. Nikolov, P. Haase, R. Cyganiak, A. Stasiewicz, A. Karamanou, M. Zotou, D. Zeginis, E. Tambouris, K. Tarabanis, Exploiting linked data cubes with opencube toolkit, in: ISWC Poster & Demonstration, 2014.