



HAL
open science

CoffeeWKG: A Weather Knowledge Graph for Coffee Regions in Colombia

Cristhian Figueroa, Nadia Yacoubi Ayadi, Nicolas Audoux, Catherine Faron

► **To cite this version:**

Cristhian Figueroa, Nadia Yacoubi Ayadi, Nicolas Audoux, Catherine Faron. CoffeeWKG: A Weather Knowledge Graph for Coffee Regions in Colombia. SmartFood 2023 - Workshop on Controlled Vocabularies and Data platforms for Smart Food Systems / co-located with the 42nd International Conference on Conceptual Modeling (ER2023), Nov 2023, Lisbon, Portugal. pp.329-342, 10.1007/978-3-031-47112-4_30 . hal-04714781

HAL Id: hal-04714781

<https://hal.science/hal-04714781v1>

Submitted on 30 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CoffeeWKG: A Weather Knowledge Graph for Coffee Regions in Colombia

Cristhian Figueroa¹[0000-0001-6697-886X], Nadia Yacoubi Ayadi²[0000-0002-6132-8718], Nicolas Audoux³[0009-0004-0379-9302], and Catherine Faron³[0000-0001-5959-5561]

¹ Universidad del Cauca, Department of Telematics, Popayán, 190003, Colombia
`cfigmart@unicauca.edu.co`

² Université de Lyon, CNRS, LIRIS UMR 5205, F-6922 Villeurbanne, France
`nadia.yacoubi-ayadi@univ-lyon1.fr`

³ Université Côte d’Azur, Inria, CNRS, I3S (UMR 7271), France
`nicolas.audoux@etu.univ-cotedazur.fr, faron@i3s.unice.fr`

Abstract. Coffee is one of the major crops produced in Colombia which is the third largest producer of coffee after Brazil and Vietnam. Together, these three countries produce more than 50% of the world’s total coffee. One of the main challenges facing coffee producers in Colombia is to determine the effects of climate variability and climate change on their production. This paper presents CoffeeWKG, an RDF knowledge graph focused on weather conditions in the coffee-growing regions of Colombia over 15 years (2006-2020), to facilitate the understanding of climate impacts on coffee crops. CoffeeWKG enables the integration of heterogeneous sensor data collected from different weather stations and the definition of semantic metadata on agro-climatic parameters. This knowledge graph enables coffee growers and experts to explore and query historical weather conditions to establish a correlation between weather data and information on coffee crops, thus revealing the complex interaction between climate and production dynamics. This research is essential to improving the resilience of agriculture and optimizing resources in the face of changing climatic challenges.

Keywords: Coffee crop · Weather Knowledge Graph · Linked Data · Meteorology

1 Introduction

Coffee plays a critical role in Colombia’s economy and cultural heritage, serving as the most important agricultural export crop. With a cultivated area of about 884 thousand hectares and 540 thousand families depending economically on coffee production, it significantly contributed to the country’s economic growth [20,10]. Colombian coffee is well-marketed in the specialty coffee industry due to its mild and sweet flavor obtained from arabica beans grown in distinctive climatic conditions (precipitation, solar radiation, temperature, and humidity)

and due to topological and edaphic factors (soil depth, acidity/alkalinity, fertility) in the cultivable areas [10].

However, a significant challenge facing coffee cultivation in Colombia is climate variability, which has become more problematic due to climate change. This variability adversely affects coffee production, resulting in economic losses for farmers. According to the Colombian Federation of Coffee Growers (FNC), in the last 12 months (June 2022 – June 2023), coffee production fell by 13%, from 12.3 million bags to 10.7 million, due to unfavorable climatic variables that affected crucial stages like flowering and grain filling in the coffee-growing regions of the country [14].

In this context, data-driven approaches are increasingly necessary to mitigate the adverse effects of climate change and fostering sustainable economic growth within the coffee farming industry. Access to accurate and comprehensive data, especially meteorological data, becomes a powerful tool for improving coffee production. In this paper, we propose to use the Semantic web models and technologies and adopt the Linked data principles to publish and integrate weather data in coffee-growing regions. We conceived and built the CoffeeWKG knowledge graph for weather conditions in Colombia’s central coffee production regions. This knowledge graph aims to help agronomists and farmers understand long-term weather trends and adapt coffee crops to shifting growing conditions, mitigating potential risks and losses.

The rest of this paper is organized as follows. Section 2 presents previous works using semantic web technologies and knowledge graphs for weather and climate conditions in agriculture. Section 3 illustrates the motivation scenario and competency questions that guided the construction of the proposed knowledge graph CoffeeWKG. Section 4 presents the data model in CoffeeWKG. Section 5 describes the data sources and the generation process of CoffeeWKG. Section 6 presents the validation of CoffeeWKG through the implementation of the competency questions as SPARQL queries. Section 7 shows conclusions and future works.

2 State of the Art

The adoption of semantic web models has brought about a ground-breaking approach towards the implementation of intelligent agricultural practices. Knowledge graphs and ontologies integrated with weather and agricultural data offer a novel perspective on understanding, analyzing, and optimizing the intricate interplay between climate, agricultural practices, and crop outcomes [7]. This section presents an overview of the state-of-the-art advancements in applying semantic web technologies to enhance crop management in the general context of Smart Food systems.

2.1 Ontologies and Vocabularies for Agriculture

Semantic web technologies have proven to be instrumental in capturing, sharing, and integrating heterogeneous agricultural data. Ontologies play a central role in

semantic representations, acting as formal vocabularies that define domain concepts and their relationships. In coffee crop management, ontologies can capture diverse aspects such as weather conditions, soil properties, disease profiles, and farming practices, enabling comprehensive knowledge representation. Currently, there are many studies describing semantic web technologies for agriculture.

AgroVoc [18] is one of the key controlled vocabularies covering areas of interest to the Food and Agriculture Organization of the United Nations (FAO). It provides concepts and terms in multiple languages to describe agricultural, forestry, fisheries, food, and related domains.

Crop Ontology [11] is a community-based project providing a central place for creating crop-related ontologies. It provides a standardized and controlled vocabulary for describing crop plants' anatomy, morphology, and phenology. Additionally, it helps in data integration and sharing across agricultural research communities.

RustOnt [17] is an ontology for modeling favorable climatic conditions to prevent coffee rust. This ontology gathers relevant concepts and instances of meteorological variables used by coffee rust control systems or models.

B BCH [16] is an ontological framework for the semantic description of plant development stages and for transforming specific scales into RDF vocabularies.

2.2 Knowledge Graphs for Climate Conditions

The AEMET meteorological dataset [2] is the first LOD dataset that made available Spanish meteorological data through a SPARQL endpoint. AEMET dataset is based on a network of ontologies, where each ontology describes a sub-domain involved in the modeling of meteorological measurements. These sub-domains include meteorological parameters such as Measurements, Sensors, Time, and Location.

In a recent study [22], a knowledge graph was created for meteorological observational data in France, referred to as WeKG-MF. This modular ontology-based knowledge graph provides a comprehensive data model that can be customized by weather data providers. WeKG-MF has also been designed to serve different use cases identified in the context of the D2KAB French project.

LinkClimate is a knowledge graph presented in [21] that uses a similar network of ontologies as [22], excluding the GeoSPARQL ontology. The researchers created a knowledge graph that incorporates information from daily climate summaries provided by National Oceanic and Atmospheric Administration (NOAA), and established connections between their graph and external geographic sources in the LOD cloud such as OpenStreetMap and Wikidata.

In our research work, we rely on the study presented in [22] and examine how the WeKG-MF data model can be used to represent weather conditions in Colombia's coffee growing areas. Our first step is to create a list of competency questions that address the needs of coffee producers.

3 Motivating Scenario and Competency Questions

Several research projects have been undertaken to demonstrate the influence of weather conditions on crop growth, encompassing various cultivars and plant species in different spatial and temporal contexts. These studies have shown that farmers can refine their farming practices by taking advantage of historical and real-time weather data. It can enable them to plan planting and harvesting times more effectively, adjust irrigation and fertilisation strategies in response to changing weather conditions, and even predict future harvests more accurately. However, it is essential to provide experts and farmers with access to semantically enriched data presented in an interoperable format, so that it can be queried and reused to meet different requirements and use cases, and integrated with other data sources. Colombian coffee (*Coffea arabica L.*) is grown using arabica beans, known for their light and flowery taste. This crop is farmed in distinctive climatic conditions (precipitation, solar radiation, temperature, and humidity) provided mainly in the intertropical zone between latitude 20° N and 24° S in altitudinal ranges from 700 to 2000 masl⁴ [10].

According to [1], the optimum mean temperature range for *Coffea arabica* per year is about 18–21°C or up to 26°C. At temperatures exceeding this range, the maturation and ripening of fruits accelerate, frequently resulting in compromised coffee quality. The optimal rainfall range for this crop is approximately 1200-1800 mm per year, about 120mm per month [6]. Excessive precipitation may damage the flowering of coffee. Moreover, the recommended sun exposure is around 1800 hours annually, considering that cultivating coffee crops under the shade of trees is advisable to shield them from excessive solar radiation, particularly during dry seasons. This practice also contributes to organic matter proliferation in the soil and fosters a conducive habitat for beneficial microbial life, bolstering plant nutrition. Consequently, we have elicited a set of competency questions [13,19] that reflect, amongst all, the specific needs of coffee growing in terms of climatic conditions. Here are some of them:

- *CQ1. What are the meteorological properties measured at stations near coffee-growing areas?*

Coffee growers and experts can analyze the main climate variables such as air temperature, relative humidity, solar exposure, and cumulative precipitation. By doing so, they can gain insights into the properties measured at the stations near their crops. This can help them identify climate-related risks such as droughts, excessive rainfall, or extreme temperatures and take proactive measures to protect their crops. They can also obtain data about past planting times, which can enable them to choose the best times for planting and harvesting coffee. Furthermore, monitoring favorable weather conditions for plant pests and diseases can allow for early detection and preventive measures to protect the crop, among other benefits.

- *CQ2. What are the closest weather stations to a spatial location of a given coffee farm?*

⁴ masl is the abbreviation of meters above sea level

Coffee growers and experts need to locate weather stations close to their crops' location for easy access to weather data. Consequently, a semantic model should offer a spatial module to capture the geographical coordinates of the stations.

- *CQ3: In which months are the highest temperatures during a specific year?*
High temperatures are correlated with drought events that produce a water deficit in coffee crops. According to [10], in coffee-growing regions of Colombia, the soil is vulnerable to water deficit, which can affect the fruit load by 25% when it lasts more than 40 days. Coffee growers may benefit of this knowledge to choose the best times for planting and harvesting coffee avoiding extreme temperatures and drought events.
- *CQ4: In which months do the highest rainfalls occur during a specific year?*
Excessive rainfalls affect coffee crop productivity, hinder disease control and weeds, increase the lixiviation of nutrients, and elevate the costs of cultivation practices [15,1]. This competency question illustrates the need of farmers and experts to avoid cultivating coffee in months more prone to rainfall.

4 The CoffeeWKG Semantic Model

To publish weather conditions in Colombia's coffee crop regions, we rely and reuse the WeKG semantic model presented in [22]. Accordingly, the CoffeeWKG model comprises semantic representations of weather stations as well as meteorological observations recorded in different coffee-growing areas in Colombia. We first present an overview of the WeKG semantic model and then we detail how weather stations and meteorological observations are represented.

4.1 Overview of the WeKG Semantic Model

WeKG relies on a self-contained semantic model presented in [22] and available on github ⁵. It is a modular ontology that reuses and extends standard ontologies, among which the GeoSPARQL ontology for spatial features and their relations [3], the Time ontology [4] for temporal entities, the Sensor, Observation, Sample, and Actuator (SOSA) [9] and Semantic Sensor Network (SSN) ontologies [8] for sensors and observations and the RDF Data Cube ontology [5] for aggregation and multidimensionality features. The WeKG model has several interesting conceptual modeling features, especially the conceptual representation of meteorological observations and the definition of observable properties and features of interest as a SKOS vocabulary.

The WeKG model distinguishes two conceptual levels of meteorological observations, atomic and aggregated ones. Atomic and fine-grained weather observations that can be recorded at different times of the day are modeled by reusing and extending classes from the SOSA/SSN ontologies [8,9]. The WeKG model

⁵ <https://github.com/cfigmart/coffeeWKG/blob/main/meteo/WeKG-modular-ontology/ontology.ttl>

introduces three additional classes, namely `weo:meteorologicalObservation`, `weo:WeatherProperty` and `weo:MeteorologicalFeature`, to represent respectively weather observations performed at a specific location and time providing a measurement of a given observable property (e.g., relative air humidity) characterizing some feature of interest (i.e., air). On the other hand, the observations may correspond to aggregated values (for example, the maximum daily air temperature) calculated in most cases on the basis of atomic observations or directly measured by specific instruments, when available at weather stations. In the WeKG model, these types of observations are modeled as multi-measure observations using the RDF Data Cube Vocabulary (DCV) [5], a W3C recommendation to describe multi-dimensional data structures.

Furthermore, the definition of observable properties and features of interest as well as their alignment with existing controlled vocabularies in SOSA/SSN ontologies and W3C RDF Data Cube vocabulary is delegated to the community of interest. The WeKG model includes a formalization of a subset of the World Meteorological Organization (WMO) vocabulary⁶ as a comprehensive SKOS vocabulary whose concepts are named individuals of classes and represent the possible observable properties and features of interest in the meteorological domain. They are aligned with concepts from the Climate and Forecast (CF) Metadata Conventions vocabulary⁷.

4.2 Modeling Weather Stations with Spatial Features

The WeKG model introduces the `weo:WeatherStation` class to represent any type of weather station. This class is a subclass of the `geosparql:Feature` class to capture the spatial location of weather stations. Therefore, each instance of `weo:WeatherStation` has a geometry embodied as a point with specific coordinates. Following the GeoSPARQL vocabulary, geo-coordinates of a weather station are defined as a Well-Known Text (WKT) literal (e.g., `POINT(8.792667 41.918)`). The adoption of GeoSPARQL is motivated by the ability to efficiently query spatial data based on a set of spatial functions. It enables to express spatial queries involving meteorological data, e.g., retrieving the cumulative precipitations recorded at the closest station to a given location. The model also reuses latitude, longitude, and altitude datatype properties from the WGS84 vocabulary since WKT literals do not integrate information about the altitude of a station.

4.3 Modeling Meteorological Observations

The CoffeeWKG knowledge graph is based on the available weather data provided by National Center for Coffee Research (Cenicafé) yearbooks which includes daily aggregate values of weather parameters as described further in section 5.1. Thus, we decided to reuse the DCV ontology that introduces the class

⁶ <https://public.wmo.int/en/resources/language-resources/meteoterm/>

⁷ <http://cfconventions.org/>

`qb:DataSet` to identify a collection of observations typed as `qb:Observation`. This class corresponds to a set of measurements (values) that share some attributes (e.g., a date) and can be aggregated according to one or more dimensions (e.g., a temporal period and/or a location). The `qb:Observation` class is reused in order to represent a set of aggregated agro-meteorological parameters calculated based on atomic observations [23].

The WeKG model also reuses the `qb:DataStructureDefinition` (DSD) class to define a set of dimensions, attributes, and measures included in the dataset along with qualifying information such as the ordering of dimensions and whether attributes are required or optional. Measures, dimensions, and attributes in DCV are represented by RDF properties, instances of the abstract `qb:ComponentProperty` class. We defined a DSD named `:annualTimesSeries` available in ⁸. According to this DSD, each observation in our dataset includes 6 daily measures : the minimum, maximum and average air temperatures, the cumulative precipitations, the relative air humidity and the sunshine duration. One or more attributes can be attached to the observation instance and thus will apply to the whole measures included, which is the case of `:date` in the example (Figure 1). It is also possible to define subsets of observations as an instance of the `qb:Slice` class, within a `qb:Dataset` in order to refer to all observations that share the same spatial and temporal dimensions. More specifically, `:annualTimesSeries` allows to define spatio-temporal slices of observations by fixing the spatial and temporal dimensions: the spatial dimension may refer to the weather station, while the temporal dimension may correspond to a calendar interval.

5 CoffeeWKG: A Weather Knowledge Graph for Coffee Regions in Colombia

In this section we present the source datasets used to generate the CoffeeWKG knowledge graph and the generation process itself.

5.1 Weather Data Sources for Coffee Regions in Colombia

The official Colombian institute responsible for providing information related to weather is the Institute of Hydrology, Meteorology, and Environmental Studies (IDEAM). It deployed weather stations distributed throughout the Colombian territory. However, the FNC also owns around 232 meteorological stations, which, unlike IDEAM stations, are more oriented to collecting meteorological data related to coffee production. These stations cover the entire Colombian coffee regions in a range between 1,000 and 2,000 meters of altitude, located on the slopes of the mountain ranges.

The FNC’s meteorological stations are classified as major and minor. Major climatic stations are equipped with instruments that measure air temperature,

⁸ <https://github.com/cfgmart/coffeeWKG/blob/main/meteo/dataset-metadata/DSD-CoffeeWKG.ttl>

relative air humidity, rainfall, and the hours of sunlight, and secondary stations measure a subset of the parameters of the main stations. In this study, we only selected the major climatic stations because these contained most of the variables required for coffee crop studies.

The observations recorded at these stations are centralized by Cenicafé and published as coffee meteorological yearbooks in PDF format. These yearbooks have been available online since 2016 and gather historical agro-climatological information obtained in the stations starting from 2006. These stations are part of the FNC’s climatological network, which is strategically located near regions with the most significant influence on coffee crops in Colombia. This network allows coffee growers and experts to permanently analyze the main climate variables such as temperature (max and min), relative humidity, solar brightness, and precipitation.

5.2 Generation of CoffeeWKG

The process for the knowledge graph generation started with downloading the coffee meteorological yearbooks in PDF format from the Cenicafé website⁹ from 2006 to 2020. Next, we manually extracted data tables from these pdf files using the Tabula tool¹⁰. It generated CSV files that were processed to obtain two datasets, one for weather stations and the other one for weather observations. The weather stations CSV file contains data about the stations, such as their code, name, municipality, latitude, longitude, and altitude. The weather observations CSV file contains measurement related to air temperatures (min, max, and mean), relative air humidity, precipitations (night, day, and total), and solar brightness.

To generate the CoffeeWKG knowledge graph from these CSV data, we updated the pipeline for WeKG-MF available on Github¹¹. We first defined mapping rules from the available CSV data to RDF triples following the WeKG data model presented in Section 4. Then we used these mapping rules to automatically generate CoffeeWKG using the Morph-xR2RML tool¹²[12].

CoffeeWKG is available under an open license at Zenodo¹³. The current version of CoffeeWKG covers a period from 2006 to 2020 with data obtained from Cenicafé yearbooks. The main statistics about CoffeeWKG are shown in Table 1.

6 Querying CoffeeWKG

To validate CoffeeWKG, we implemented the competency questions discussed in Section 3 in the SPARQL language. To query CoffeeWKG with them, we in-

⁹ Cenicafé yearbooks download page: https://www.cenicafe.org/es/index.php/nuestras_publicaciones/anuarios_meteorologicos

¹⁰ <https://tabula.technology/>

¹¹ WeKG-MF GitHub repository: <https://github.com/Wimmics/weather-kg>

¹² <https://github.com/frmichel/morph-xr2rml/>

¹³ CoffeeWKG download page: <https://zenodo.org/record/8237867>

Table 1. Statistics for the CoffeeWKG knowledge graph

Category	Number of resources
Triples	2.625.361
Weather Stations	62
Observations	786.513
Agro-climatic properties	6

stalled the Apache Jena Fuseki server¹⁴, which allows us to upload our knowledge graph and provides a SPARQL endpoint to execute queries.

6.1 Querying Meteorological Properties

The first competency question (CQ1) is about knowing the meteorological properties available in weather stations located in coffee regions in Colombia. The SPARQL query for this purpose is shown in listing 1.1.

Listing 1.1. SPARQL query to retrieve the meteorological properties measured by weather stations available in CoffeeWKG (CQ1).

```
SELECT ?measure ?obsproperty
WHERE {
  ?observation qb:observation
               [a qb:Observation; ?measure ?value ]
  ?measure qb:concept ?obsproperty .
}
```

This query helps farmers and experts make informed decisions about their crops by providing information on the capabilities and limitations of nearby weather stations.

6.2 Querying Weather Stations Near a Specific Coffee Farm

The second competency question looks for weather stations near a specific coffee farm, indicated by its latitude and longitude coordinates. For example, the SPARQL query shown in Listing 1.2 finds the closest weather station near a coffee farm at latitude 4° 57' 1" and longitude - 74° 24' 2".

Listing 1.2. SPARQL query to retrieve the closest station to a specific spatial point (CQ2).

```
SELECT ?stationName ?lat ?long ?coordinates ?distance
WHERE {
  ?station rdfs:label ?stationName ;
           geosparql:hasGeometry [geosparql:asWKT ?coordinates];
           geo:lat ?lat; geo:long ?long .
BIND("POINT(-74.40055555555556 4.950277777777778)
```

¹⁴ Apache Jena Fuseki Server: <https://jena.apache.org/documentation/fuseki2>

```

    ""<http://www.opengis.net/ont/geosparql#wktLiteral>
    as ?Currentposition)
BIND (geof:distance(?coordinates,?Currentposition ,
    uom:metre) as ?distance)
}
ORDER BY ?distance
LIMIT 1

```

Understanding the precise locations of meteorological stations in proximity to the crops is instrumental in guiding experts toward decisions on establishing new meteorological stations or devising sophisticated meteorological interpolation systems for data acquisition in these specific zones.

6.3 Querying Highest Temperatures by Month

This competency question identifies the highest monthly temperatures recorded to help farmers identify possible drought events causing a water deficit in coffee crops based on similar past events that look like the current one. Listing 1.3 calculates the number of days per month when the temperature exceeds a threshold (in the sample query $25^{\circ}C$) as well as the maximum temperature and the average temperature measured at a specific weather station (in the sample query the station is named "Alberto Santos"). It is possible to filter the results for a specific year by adding a `FILTER` clause.

Listing 1.3. SPARQL query to calculate highest temperatures (CQ3).

```

SELECT ?year ?month
  (SUM(IF(?temp_max>25,1,0)) AS ?nbExtrHeatDays)
  (MAX(?temp_max) AS ?max_T_Month)
  (AVG(?temp_avg ) AS ?month_T_AVG)
WHERE {
  ?station a weo:WeatherStation ;rdfs:label ?stationName .
  ?s a qb:Slice ;
    wes-dimension:station ?station ;
    wes-dimension:year ?year ;
    qb:observation [
      a qb:Observation ;
      wes-attribute:observationDate ?date ;
      wes-measure:maxDailyTemperature ?temp_max ;
      wes-measure:avgDailyTemperature ?temp_avg ].
  BIND (SUBSTR(STR(?date),6,2) AS ?month)
}
GROUP BY ?stationName ?year ?month
ORDER BY DESC(?year) ?month

```

6.4 Querying Highest Rainfalls by Month

This competency question identifies the highest monthly rainfalls to warn farmers about cultivating coffee in months more prone to rainfall. Listing 1.4 shows

monthly accumulated rainfall for each year at a specific station identified by its name (in the sample query, the station with a name equal to "Alberto Santos"). It is also possible to filter the results for a specific year by adding a `FILTER` clause.

Listing 1.4. SPARQL query to calculate cumulative precipitations (CQ4).

```

SELECT ?year ?month (SUM(?rain ) as ?monthlyRain)
WHERE {
  ?station a weo:WeatherStation ;
           rdfs:label ?stationName .
  FILTER(?stationName = "Alberto Santos")
  ?s a qb:Slice ;
     wes-dimension:station ?station ;
     wes-dimension:year ?year ;
     qb:observation [
       a qb:Observation ;
       wes-attribute:observationDate ?date ;
       wes-measure:rainfall24h ?rain ] .
  BIND (SUBSTR(STR(?date),6,2) AS ?month)
}
GROUP BY ?year ?stationName ?month
ORDER BY ?year ?month

```

7 Conclusions

Climate variability is a significant constraint in coffee production, which plays a crucial role in Colombia's economy. Therefore, studying and analyzing meteorological data in coffee regions helps farmers and experts fine-tune their agricultural practices based on historical and real-time weather patterns to schedule planting and harvesting periods more effectively, adjust irrigation and fertilization strategies, and even predict future harvests more accurately. In this paper we presented the CoffeeWKG knowledge graph for weather conditions of coffee regions in Colombia constructed with data extracted from FNC yearbooks and processed with an adaptation of the WeKG-MF [22] model initially developed for data from the French weather data provider Météo-France. The CoffeeWKG knowledge graph can be queried using the SPARQL query language to ask for historical weather conditions in coffee regions and correlate this information with coffee crop data to understand how these conditions may favor or hamper coffee production or diseases.

In future work, we intend to complement the meteorological data initially obtained from Cenicafé with satellite data and IDEAM stations located near coffee regions to cover some areas not included in the FNC yearbooks. Furthermore, we are working on integrating other types of data related to coffee production, pest and disease advent, and other factors that may be affected by climate conditions to enrich our knowledge graph. This will enable to meet more complex requirements about weather conditions, coffee production, or diseases. In particular,

we plan to link CoffeeWKG with the RustOnt ontology [17] that models data, expressions, and samples related to coffee rust. In the same way, we also plan to link CoffeeWKG with other relevant Linked Open Data (LOD) datasets and to add more weather indicators, e.g., the growing degree days (GDD)¹⁵ described in the Global Change Master Directory (GCMD), which could help farmers to predict plant and pest development rates.

References

1. Ahmed, S., Brinkley, S., Smith, E., Sela, A., Theisen, M., Thibodeau, C., Warne, T., Anderson, E., Dusen, N.V., Giuliano, P., Ionescu, K.E., Cash, S.B.: Climate change and coffee quality: Systematic review on the effects of environmental and management variation on secondary metabolites and sensory attributes of *Coffea arabica* and *Coffea canephora*. *Frontiers in Plant Science* **12** (2021). <https://doi.org/10.3389/fpls.2021.708013>
2. Atemezing, G., Corcho, O., Garijo, D., Mora, J., Poveda-Villalón, M., Rozas, P., Vila-Suero, D., Villazón-Terrazas, B.: Transforming meteorological data into linked data. *Semantic Web* **4**, 285–290 (2013). <https://doi.org/10.3233/SW-120089>
3. Battle, R., Kolas, D.: Enabling the geospatial semantic web with parliament and geosparql. *Semantic Web* **3**, 355–370 (2012). <https://doi.org/10.3233/SW-2012-0065>
4. Cox, S., Little, C.: Time ontology in owl. w3c candidate recommendation draft. Tech. rep., W3C (11 2022), <https://www.w3.org/TR/owl-time/>
5. Cyganiak, R., Reynolds, D.: The rdf data cube vocabulary. Tech. rep., W3C (1 2014), <https://www.w3.org/TR/2014/REC-vocab-data-cube-20140116/>
6. DaMatta, F.M., Ronchi, C.P., Maestri, M., Barros, R.S.: Ecophysiology of coffee growth and production. *Brazilian Journal of Plant Physiology* **19**, 485–510 (12 2007). <https://doi.org/10.1590/S1677-04202007000400014>
7. Davis, A.P., Gole, T.W., Baena, S., Moat, J.: The impact of climate change on indigenous arabica coffee (*Coffea arabica*): Predicting future trends and identifying priorities. *PLOS ONE* **7**(11), 1–13 (11 2012). <https://doi.org/10.1371/journal.pone.0047981>
8. Haller, A., Janowicz, K., Cox, S.J., 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**, 9–32 (12 2018). <https://doi.org/10.3233/SW-180320>
9. Janowicz, K., Haller, A., Cox, S.J., Phuoc, D.L., Lefrançois, M.: Sosa: A lightweight ontology for sensors, observations, samples, and actuators. *Journal of Web Semantics* **56**, 1–10 (5 2019). <https://doi.org/10.1016/j.websem.2018.06.003>
10. León-Burgos, A.F., Unigarro, C., Balaguera-López, H.E.: Can prolonged conditions of water deficit alter photosynthetic performance and water relations of coffee plants in central-west colombia? *South African Journal of Botany* **149**, 366–375 (9 2022). <https://doi.org/10.1016/j.sajb.2022.06.034>
11. Matteis, L., Chibon, P.Y., Espinosa, H., Skofic, M., Finkers, R., Bruskiwich, R., Hyman, G., Arnaud, E.: Crop ontology: Vocabulary for crop-related concepts. In: *CEUR Workshop Proceedings*. vol. 979 (2013)

¹⁵ The concept GDD is available at the URI of the GCMD vocabulary <https://gcmd.earthdata.nasa.gov/kms/concept/6d808909-ce04-4401-a883-aff4d723d025>

12. Michel, F., Djimenou, L., Faron-Zucker, C., Montagnat, J.: Translation of relational and non-relational databases into rdf with xr2rml. In: Proceedings of the 11th International Conference on Web Information Systems and Technologies. pp. 443–454. SCITEPRESS - Science and Technology Publications (2015). <https://doi.org/10.5220/0005448304430454>
13. Peroni, S.: A Simplified Agile Methodology for Ontology Development, vol. 10161 LNCS, pp. 55–69. Springer International Publishing (2017). https://doi.org/10.1007/978-3-319-54627-8_5
14. Prensa Federación Nacional de Caficultores de Colombia: En junio importaciones de café disminuyeron un 30 por ciento (7 2023), <https://federaciondefeferos.org/wp/listado-noticias/en-junio-importaciones-de-cafe-disminuyeron-un-30/>
15. Ronchi, C.P., Miranda, F.R.: Flowering percentage in arabica coffee crops depends on the water deficit level applied during the pre-flowering stage. *Revista Caatinga* **33** (2020). <https://doi.org/10.1590/1983-21252020v33n121rc>
16. Roussey, C., Delpuech, X., Amardeilh, F., Bernard, S., Jonquet, C.: Semantic description of plant phenological development stages, starting with grapevine. *Communications in Computer and Information Science* **1355 CCIS**, 257–268 (2021). https://doi.org/10.1007/978-3-030-71903-6_25
17. Suarez, C., Griol, D., Figueroa, C., Corrales, J.C., Corrales, D.C.: Rustont: An ontology to explain weather favorable conditions of the coffee rust. *Sensors* **22**, 9598 (12 2022). <https://doi.org/10.3390/s22249598>
18. Subirats-Coll, I., Kolshus, K., Turbati, A., Stellato, A., Mietzsch, E., Martini, D., Zeng, M.: Agrovoc: The linked data concept hub for food and agriculture. *Computers and Electronics in Agriculture* **196** (5 2022). <https://doi.org/10.1016/j.compag.2020.105965>
19. Uschold, M., Gruninger, M.: Ontologies: principles, methods and applications. *The Knowledge Engineering Review* **11**, 93–136 (6 1996). <https://doi.org/10.1017/S0269888900007797>
20. Vélez-Vallejo, R.: Informe del gerente - 90 congreso nacional de cafeteros. Tech. rep., Federación Nacional de Cafeteros de Colombia (11 2022), <https://federaciondefeferos.org/app/uploads/2022/12/Informe-del-Gerente-D.pdf>
21. Wu, J., Orlandi, F., O’Sullivan, D., Dev, S.: Linkclimate: An interoperable knowledge graph platform for climate data. *Computers and Geosciences* **169**, 105215 (2022). <https://doi.org/10.1016/j.cageo.2022.105215>
22. Yacoubi Ayadi, N., Faron, C., Michel, F., Gandon, F., Corby, O.: A Model for Meteorological Knowledge Graphs: Application to Météo-France Data. In: ICWE 2022- 22nd International Conference on Web Engineering. 22nd International Conference on Web Engineering, ICWE 2022, Bari, Italy (Jul 2022). https://doi.org/10.1007/978-3-031-09917-5_19
23. Yacoubi Ayadi, N., Faron, C., Michel, F., Gandon, F., Corby, O.: Computing and visualizing agro-meteorological parameters based on an observational weather knowledge graph. In: Ding, Y., Tang, J., Sequeda, J.F., Aroyo, L., Castillo, C., Houben, G. (eds.) *Companion Proceedings of the ACM Web Conference 2023, WWW 2023, Austin, TX, USA, 30 April 2023 - 4 May 2023*. pp. 242–245. ACM (2023). <https://doi.org/10.1145/3543873.3587357>

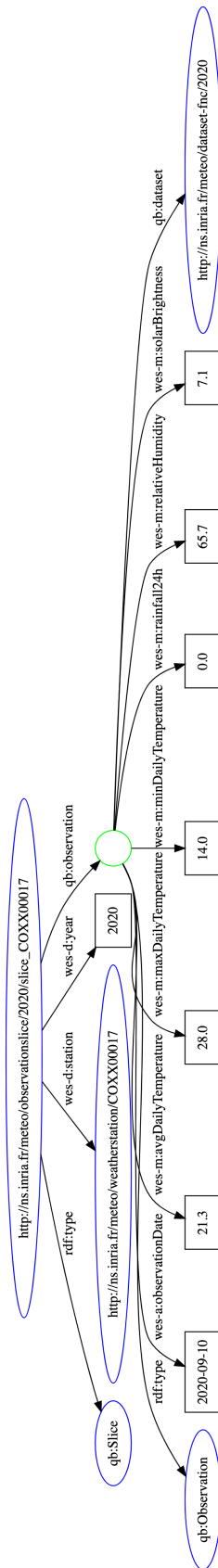


Fig. 1. RDF graph representing an instance of class `qb:Observation` in CoffeeWKG