Weather Station Data Publication at Irstea: an Implementation Report

Catherine Roussey[1], Stephan Bernard[1], Géraldine André[1], Oscar Corcho[2], Gil De Sousa[1], Daniel Boffety[1], and Jean-Pierre Chanet[1]

[1] Irstea, UR TSCF Technologies et systèmes d'information pour les agrosystèmes, 9 avenue Blaise Pascal - CS 20085 Aubiére, F-63178, France
{firstname.lastname}@irstea.fr
[2] Ontology Engineering Group, Universidad Politécnica de Madrid, Spain
ocorcho@fi.upm.es

Abstract. We describe how we are publishing RDF data about one of the meteorological stations that Irstea owns in its experimental farm at Montoldre. Our objective is to revisit previous work done by other researchers on the publication of meteorological Linked Data, and provide some recommendations and best practices based on our experience.

Keywords: SSN, meteorological data

1 Introduction

Several works in the state of the art address the publication of meteorological data as Linked Data, using the Semantic Sensor Network (SSN) ontology [8] proposed by the W3C Incubator Group on Semantic Sensor Networks. The work presented in this paper aims to confirm whether the steps taken by these previous works on publishing meteorological data can be easily reused for a similar purpose and to detect any potential difficulties in the usage of the SSN Ontology, rather than presenting new innovations on the application of the SSN Ontology in this domain.

We have chosen to publish data from the Vantage Pro 2 weather station in use at our experimental farm located in Montoldre (France)[3]. We have followed the usual steps in Linked Data publication, as discussed in [1], paying special attention to the reuse of existing general and domain ontologies applicable to this type of data publication.

This paper is organized as follows: first, we describe our weather station and its data; next we study some works on weather data publication based on SSN. Then we briefly present the network of ontologies that we have reused. In Section 5, we describe how we have populated the selected ontology network, and Section 6 describes the workflow followed in this process. Finally we conclude by presenting an analysis of our work and perspectives.

[3] http://www.irstea.fr/la-recherche/themes-de-recherche/motive/station-de-montoldre

2 Montoldre's Weather Station Description and Data Sources

Irstea has a research and experimentation site located at Montoldre, where different types of experiments are run. This site has its own weather station, a Vantage Pro 2[4] from Davis Instruments[5]. The station has the following components: a clock, two temperature sensors (indoor and outdoor), an atmospheric pressure sensor, two air humidity sensors (indoor and outdoor), a wind vane, an anemometer, a rain gauge to measure water precipitation and speed, and a solar radiation sensor.

The external sensors communicate wirelessly with a console located inside a building, which (in addition to the sensors it contains) allows calculating other variables (wind chill, dew point, etc.), as well as weather forecasts. By connecting the console to a computer it is possible to store the values measured and put them online on a Web server[6]. The storage of measures is automatically done in the station according to the user parameters (intervals of time, units, etc.). These data can be extracted to generate a CSV (comma-separated values) file containing the following information:

Date	Time	Temp Out	Hi Temp	Low Temp	Out Hum	Wind Speed	Wind Dir
01/01/13	0:30	6.2	6.7	6.0	73	1.6	WSW
01/01/13	1:00	7.1	7.1	6.1	68	3.2	WSW
01/01/13	1:30	7.5	7.5	7.0	67	3.2	WSW
01/01/13	2:00	7.8	7.8	7.5	67	4.8	WSW
01/01/13	2:30	7.7	7.8	7.7	70	4.8	WSW
01/01/13	3:00	7.2	7.7	7.2	75	3.2	SW

For the purpose of data publication, we have generated CSV files for the period between 2010 and 2013 with the following measures: outdoor temperature, external atmospheric pressure, relative humidity of outside air, wind direction, wind speed, precipitation quantity, water precipitation rate, and solar radiation.

3 State of the Art

The SSN ontology [8] can be used as the basis for the publication of weather station data. This ontology must be linked with other ontologies to form an ontology network, including the following ones:

- Ontologies to describe the different type of sensors.
- Ontologies to describe weather phenomena and their measurable properties.
- Ontologies to describe units of measurement.

[4] `http://www.davis-meteo.com/Vantage-Pro2.php`
[5] `http://davisnet.com/`
[6] `http://meteo.clermont.cemagref.fr/` - service accessible only inside Irstea

– Ontologies to describe geographical places and their location.
– Ontologies to describe temporal entities.

Table 1 lists some of the datasets that use SSN for publishing meteorological data as Linked Data. The first column indicates the name of the dataset. The column "Environmental ontology" presents the ontology used to describe the Feature Of Interests and their associated properties. The column "Spatial ontology" indicates the ontology used to describe the localisation information. The column "time ontology" indicates the ontology used to describe time information, if any. The last column shows more ontologies used in the dataset.

dataset name	environmental ontology	spatial ontology	time ontology	other ontologies
AEMET	AEMET	WGS84 Geobuddies	W3C Time	
Swiss Experiment	SWEET	WGS84		QUDT
ACORN-SAT		WGS84	UK Intervals	DUL Data Cube
SMEAR	SWEET	Geoname WGS84	DUL	Data Cube Situation Theory

Table 1. Projects where SSN has been used to publish meteorological data

AEMET (Agencia Estatal de Meteorología) is the Spanish public agency in charge of collecting and publishing weather data. The workflow used to publish this dataset as Linked Data is described in [9]. It is important to note here that this work was done in parallel to the development of the SSN Ontology, and for this reason some of the design decisions taken for the publication of these data sources are not totally compliant with the current SSN Ontology. This dataset uses the ontologies: AEMET[7], WGS84[8], Geobuddies[9] and W3C Time[10].

The Swiss Experiment Linked Data publication effort reported in [10] proposes the use of the SSN ontology to combine and publish several meteorological data streams provided by heterogeneous sensor networks. Thus this work combines several ontologies to build the common schema that will be used to query sensor data and metadata. SWEET[11], WGS84 and QUDT[12] are reused in this project.

The Australian Bureau of Meteorology published an homogenised daily temperature dataset called ACORN-SAT. In [12], the authors describe how this

[7] `http://aemet.linkeddata.es/models_en.html`
[8] `http://www.w3.org/2003/01/geo/`
[9] `http://mayor2.dia.fi.upm.es/oeg-upm/index.php/en/ontologies/83-geobuddies-ontologies`
[10] `http://www.w3.org/TR/owl-time`
[11] `http://sweet.jpl.nasa.gov/`
[12] `http://www.qudt.org/`

dataset was published on the Linked Data using two main ontologies: SSN and RDF Data Cube[13]. UK Intervals[14] and DUL[15] are also used.

The Finnish Station for Measuring Ecosystem Atmosphere Relations (SMEAR) is a large scale sensor network which measures environmental phenomena like weather or atmospheric gases. The works presented in [13] propose a software framework able to interpret sensor data at different level of details. The global architecture is composed of hierarchical layers : measurement, observation, derivation and situation. Each layer reads the data from previous one and increases data complexity in order to enhance its interpretation. The observation layer is based on the SSN ontology.

To conclude, table 1 shows that all meteorological datasets use different networks of ontologies. Thus SSN is generic and does not prevent to build heterogeneous schema for meteorological data publication.

4 An Ontology Network for Weather Station Data Publication

In this section we describe the ontologies that we have reused for the publication of our weather station data. We describe briefly these ontologies and the main parts that we have used from them, if any.

4.1 The W3C Semantic Sensor Network (SSN) Ontology

The *"Semantic Sensor Network"* (SSN) ontology [8] is a generic ontology created by the W3C Semantic Sensor Network Incubator Group. This ontology contains different modules. Given our interest on the publication of sensor data and on describing the sensors used to produce such data, we have focused on the classes and properties defined in the following modules: Skeleton, Data, Platform Site, Device and System. More specifically, the classes that we have reused from this ontology are:

– $ssn : Observation$ to describe the measurement context,
– $ssn : FeatureOfInterest$ to specify the observed phenomena,
– $ssn : SensingDevice$ to describe the sensors,
– $ssn : Platform$ to describe where the sensors are installed,
– $ssn : System$ to describe a system composed of several sensors such as, for example, our weather station.

We have also used the main properties associated to these classes: $ssn : observedProperty$, $ssn : observedBy$, $ssn : hasSubsystem$, $ssn : featureOfInterest$, etc.

[13] `http://purl.org/linked-data/cube`
[14] `http://reference.data.gov.uk/def/intervals`
[15] `http://www.loa-cnr.it/ontologies.DOL.owl`

4.2 The AWS Ontology for Meteorological Sensors

The *"Ontology for Meteorological Sensors"* [6] (AWS) extends the SSN ontology by extending its class $ssn : SensingDevice$. It is focused on the description of different models of sensors that can be used to measure weather phenomena, and hence is also of interest for our purposes.

More specifically, we have mostly reused the following classes:

- $aws : AtmosphericPressureSensor$ for the atmospheric pressure sensor,
- $aws : CapacitiveThinFilmPolymer$ specialisation of $aws : HumiditySensor$ for the hygrometer,
- $aws : Pyranometer$ specialisation of $aws : RadiationSensor$ for the solar radiation sensor,
- $aws : Thermistor$ specialisation of $aws : TemperatureSensor$ for the thermometer,
- $aws : TippingBucketRainGaugeTbrgWithoutCorrection$ specialisation of $PrecipitationSensor$ for the pluviometer; this sensor is able to produce two separate measurements: the quantity of rain and the speed of the precipitation,
- $aws : WindVane$ specialisation of $aws : WindSensor$ for the weather vane,
- $aws : CupAnemometer$ specialisation of $aws : WindSensor$ for the anemometer,

Furthermore, some of these classes contain restrictions indicating which type of $ssn : Property$ they are able to measure. For example, the class $aws : Pyranometer$ is defined by $Pyranometer \sqsubseteq \exists observes.EnergyFlux$. This has been also useful for our purpose. Note that sometimes the documentation of the weather station is not complete enough in order to select the appropriate sensor model in AWS hierarchy. This is the case of the atmospheric pressure sensor. We were not able to select one of those sensors. Thus we do not try to specialise this sensor. We have noticed that AWS proposes lots of sensor models. In our case AWS provides all the sensor descriptions we need for our purpose. Note that AWS does not import any ontology but it defines and reuses several prefixes like dim of the QU ontology.

4.3 Climate and Forecast (CF) features Ontology

The *"Climate and Forecast features"* ontology [2] (aka $cf - feature$) is a translation of the "Climate and Forecast (CF) standard names vocabulary"[16] maintained by the "Program for Climate Model Diagnosis and Intercomparison"[17]. This ontology was used to produce one of the use cases of the W3C SSN ontology.

The ontology $cf - feature$ proposes elements to describe climate measurements and weather phenomena. It consists of two modules. The module "cf-feature" is used to define environmental observed phenomena (rain, wind, etc.). It thus contains classes that specialise the $ssn : FeatureOfInterest$ class. We have used the following individuals and classes from this ontology:

[16] `http://cf-pcmdi.llnl.gov/documents/cf-standard-names/`
[17] `http://cf-pcmdi.llnl.gov/`

- $cf - feature : air$ instance of $cf - feature : Medium$,
- $cf - feature : ground\_level\_soil$ instance of $cf - feature : SurfaceMedium$,
- $cf - feature : rainfall$ instance of $cf - feature : Precipitation$,
- $cf - feature : wind$ instance of $cf - feature : Wind$.

The module "dim" is used to define measurable properties (speed, volume, etc.). It contains classes that should specialise the $ssn : Property$ class[18]. Note that this module uses the prefix dim of the QU ontology. All the individuals of this module are defined with the prefix $cf - property$. We have reused the following individuals:

- $cf - property : air\_temperature$ instance of $dim : Temperature$,
- $cf - property : air\_pressure$ instance of $dim : StressOrPressure$,
- $cf - property : relative\_humidity$ instance of $dim : Dimensionless$,
- $cf - property : wind\_from\_direction$ instance of $dim : Angle$,
- $cf - property : wind\_speed$ instance of $dim : VelocityOrSpeed$,
- $cf - property : rainfall\_amount$ instance of $dim : SurfaceDensity$,
- $cf - property : rainfall\_rate$ instance of $dim : VelocityOrSpeed$,
- $cf - property : downward\_heat\_flux\_at\_ground\_level_in_soil$ instance of $dim : EnergyFlux$.

Some of these individuals are linked to individuals of the module "cf-feature". For example, $cf - property : air\_pressure$ is linked to $cf - feature : air$ by the property $ssn : isPropertyOf$. However, this is not the case for all individuals (for example, $cf - property : relative\_humidity$ is not linked to any instance of $ssn : FeatureOfInterest$). Hence we had to provide some of these links, as shown in the Table 2.

In general, this ontology still needs to be better documented. We have difficulty to make the distinction between $cf - feature$ ontology and $cf - property$ one, these two ontologies having very similar schema. Thus for clarity purpose maybe only one ontology should be defined. Moreover, sometimes it is difficult to choose between two individuals. This is, for example, the case, for $cf - feature : air$, which is an instance of $cf - feature : Medium$, and $cf - feature : atmosphere\_air$, which is an instance of $cf - feature : LayerMedium$. As this ontology also imports many other ontologies, it is sometimes difficult to know the origin of some of the elements that it defines, in order to make decisions about their usage or reuse.

4.4 The Library for Quantity Kinds and Units (QU)

The *"Library for Quantity Kinds and Units"* (QU) ontology [4] has been also created by a W3C working group. The $cf - feature$ ontology, discussed in the previous section, imports directly the QU ontology. Indeed, the $cf - feature$ ontology reuses the modules "dim" and "unit" of the QU ontology. The AWS

[18] All these classes are defined as subClassOf $qu : QuantityKind$. They are not defined directly has subClassOf $ssn : Property$

ontology does not import QU but reuses its prefixes like *dim*. The module “dim” of the QU ontology defines classes that are useful to categorise physical quantities, such as $dim : Angle$, $dim : Distance$, $dim : Dimensionless$, etc. The module “unit” defines classes that are useful to categorise measurement units, and also provides instances of those classes, so as to identify units such as $unit : hectopascal$, $unit : percent$, etc. Therefore, we have reused this ontology[19] for the representation of our units of measurement, as we will discuss later.

4.5 The ISA Location Core Vocabulary (LOCN) and GeoSPARQL

Currently, several ontologies exist for the publication of spatial data. Each of which has a different origin and different purposes. This makes it sometimes difficult to determine which are the best options to follow when aiming at representing geospatial information, such as the one associated to a weather station.

The WGS84 vocabulary is the oldest and most commonly used vocabulary to indicate the spatial coordinates of any geographical feature. All the previous work on publishing meteorological dataset use it. We decide not to use it because all these datasets do not contain the location property that link a spatial thing to the point which is its geometrical representation. We think that a clear distinction should be made between the spatial object and one of its possible geometrical representations.

We decide to use the *“GeoSPARQL”* vocabulary [5]. GeoSPARQL is the result of a standardization process at the Open Geospatial Consortium (OGC), which was initially focused on querying geographical data. It has also proposed the model to use for describing geometries of spatial objects (through the object property *hasGeometry*, and the use of GML or WKT strings). GeoSPARQL extends the WGS84 vocabulary and proposes different types of geometries like: point, polygon, multipolygon, etc. It also allows defining topological relationships between geometric elements.

The *“ISA Core Location”* vocabulary [3] (LOCN) was released in November 2013, and has recently been given a W3C-owned namespace, although it was initially generated outside the consortium. This lightweight ontology is focused on the description of places and their address, providing a set of three classes and several properties for their description. Notably, aspects related to the geometry description of places are still in an *“unstable”* state, and hence they may change in the future. This is the ontology that we have used for specifying the address of the experimental farm.

4.6 The W3C Time Ontology

The W3C Time ontology [7][11] enables the description of time instants (instances of the class $time : Instant$) and intervals (instances of the class $time : Interval$). Hence it may be useful when we need to describe the timestamps or the time intervals associated to the measurements made by the weather station.

[19] prefixed qu-rec20 `http://purl.oclc.org/NET/ssnx/qu/qu-rec20`

We have used the classes $time : Interval$ and $time : Instant$, and the associated properties $time : month$, $time : hour$, etc.

5 Design and Population of the Ontology Network for Weather Station Data

Based on the ontology network described in the previous section, we are now able to create a dataset containing all the individuals describing measurements of our weather station. Now we explain the decisions taken in order to create resource URIs (Section 5.1) and we provide examples of how we publish different types of data according to our ontology network (Section 5.2).

5.1 Resource URIs for our Weather Station Data

URIs have been designed with several principles in mind, such as simplicity, stability and manageability. We have followed common guidelines for their effective uses, following in many cases the recommendations already applied in [9]. This section presents the main URI design decisions and conventions used, and Table 3 provides a summary of the main types of URIs that we generate.

The base URI is `http://ontology.irstea.fr/weather/`, prefixed as *irstea*. Hence all individuals follow the URI scheme `http://ontology.irstea.fr/weather/resource`. For example, the URI to identify the experimental research site of Montoldre is: `http://ontology.irstea.fr/weather/resource/location#irsteaClermontMontoldre`.

Specially relevant is the template used for generating identifiers for the observations (that is, instances of the class $ssn : Observation$). In this case, we had initially considered moving from the URI template that had been used for the AEMET Linked Data generation (where cool but rather long URIs were generated as a consequence of including in the observation the identifiers of the timestamp, sensor and property that was measured) into more simple URIs generated as an MD5 hash code from the string proposed in Table 3. For instance, the MD5 hash code for that observation would be something like *eea6c7338102cb8866c8ad563bb85faf*. After discussing with our domain experts, they did not find it so troublesome to have long URIs with a descriptive identifier, since they considered that this is a rather normal way to name files in many file systems for many scientific domains. Hence we kept them as cool URIs.

5.2 Excerpts of our Ontology-based Weather Station Data

In the following subsections we provide examples of how we generate some of our RDF data according to the selected ontologies. We think that these excerpts of the global ontology instances will be useful for others trying to generate and publish RDF data from a similar domain. It is also a good example of the use of SSN with other ontologies.

Describing the Weather Station. In this section we provide a general overview of how we describe the main context of where our weather station is located. Besides the SSN Ontology (our weather station is an instance of the $ssn : System$ and $ssn : Platform$ classes) we use two others vocabulary: the GeoSPARQL vocabulary in order to relate the weather station to its geometry, and the ISA Core Location Vocabulary (LOCN) for the description of the address where the Montoldre experimental farm is located. Note that the $geosparql : Geometry$ instance uses the $geosparql : asWKT$ property, with a WKT string to specify the corresponding geographical point. Note that the LOCN vocabulary propose to use the $dcterms : Location$ class to identify spatial objects.

To combine the LOCN vocabulary with the GeoSPARQL one, we add an instantiation link from the $dcterm : Location$ instance to the $geosparql : SpatialObject$ class.

Figure 1 provides a graphical summary of our model and individuals.

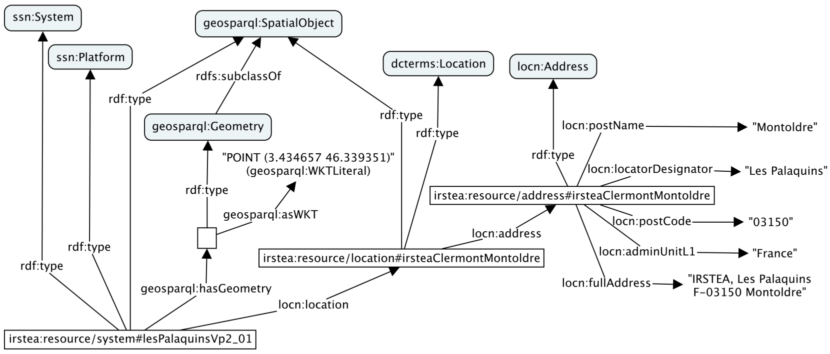

Fig. 1. Description of our weather station at Montoldre

Describing the Sensors deployed in the Weather Station. We have also focused on describing each of the sensors that are assembled in our weather station. Figure 2 provides a graphical overview of one of these sensors, more specifically the anemometer.

This anemometer is identified by a URI that finishes with $WindGauge\_01$. It is an instance of the class $aws : CupAnemometer$, which is a subclass of $aws : WindSensor$, in order to specify clearly the type of sensor that the anemometer represents. This anemometer is a subsystem of the weather station, expressed with the property $ssn : hasSubsystem$ between the weather station and the sensor. Finally, we specify that the sensor observes the individual $cf-property : wind\_speed$.

An important aspect to note here is to be able to connect the individuals of $cf-feature$ ontology to the class $ssn : Property$. The definition of $ssn : Sensor$ states that a sensor observes only instance of $ssn : Property$. Thus a reasoner will deduce automatically that the individual $cf - property : wind\_speed$ is an instance of $ssn : Property$. Even if the class $qu : QuantityKind$ is not defined as a subclass of $ssn : Property$, we can deduce that, in the case of our description of sensors, any instance of $qu : QuantityKind$ will also be an instance of $ssn : Property$. This inference is represented in figure 2 by a dash arrow.

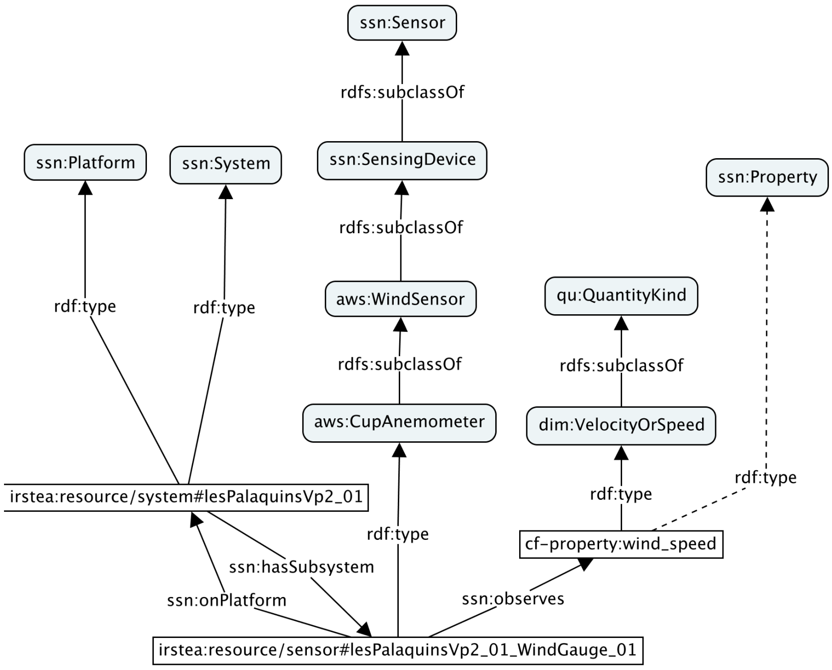

Fig. 2. Description of the wind gauge installed in our weather station

Describing the Observations. Observations allow describing the context of a measurement done by a sensor, and their description lies at the core of the SSN Skeleton module, and it is one of the most well documented patterns to follow. Figure 3 represents an observation done by the anemometer, described previously, on the wind speed at a given point in time. We can see here that we use the properties $ssn : observedProperty$, $ssn : featureOfInterest$, $ssn : observedBy$ and $ssn : observationResult$ to relate our specific observation with its corresponding observed property, feature or phenomenon, sensor used to obtain the measurement, and result of the measurement, respectively. As for the result of the observation, it is important to note that this is a pattern that is not so well documented and hence there are multiple ways to represent the observation values, together with their corresponding units. For example, the QU ontology proposes the properties $qu : numericalValue$ and $qu : unit$. As shown in Figure 3 we prefer to use the properties already contained in ssn:

- the property $ssn : hasValue$ to relate the result with the corresponding $ssn : ObservationValue$,
- the property $DUL : hasDataValue$ to express the actual value,
- the property $DUL : isClassifiedBy$ to express the unit of the measurement.

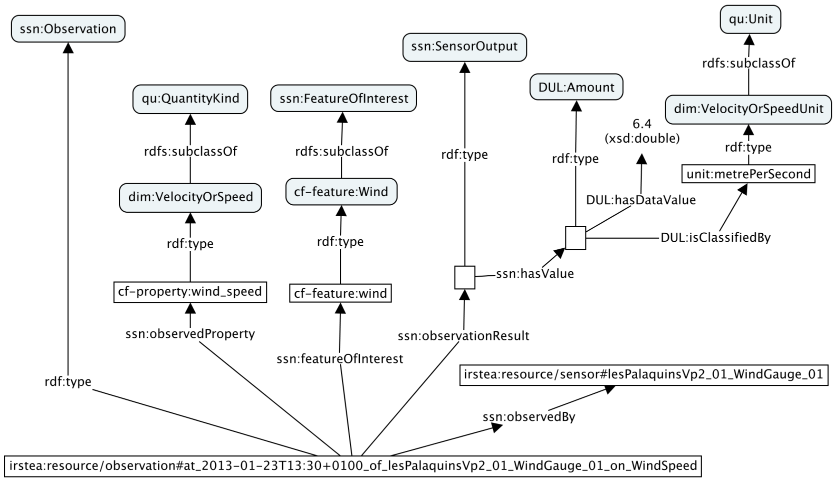

Fig. 3. Description of an observation of wind speed

Describing the Time Instants or Intervals associated to the Observations. This is also a pattern that is not well documented in the existing SSN Ontology documentation, and hence many options are available for the generation of the corresponding data. An observation can be related to a time instant or describe the result of an aggregation of values (average, sum, maximum, minimum, etc.) over a time interval. The relationship of the observation with the corresponding time instant or time interval is done with the property $ssn : observationResultTime$, and the terms that are needed to specify time instants and intervals are provided in the W3C Time Ontology.

Figure 4 presents an example where the observation is related to a time instant (for instance, in the case of a temperature measurement). In this case, the range of the property $ssn : observationResultTime$ is an instance of the class $time : Instant$ which is connected to the corresponding $xsd : dateTime$ value according to the ISO 8601 format via the property $time : inXsdDateTime$. Besides, we provide in our examples additional properties defined by the W3C Time Ontology, such as $time : year$, $time : month$, $time : day$, $time : hour$, and $time : minute$, so as to allow creating GROUP BY queries more easily, which

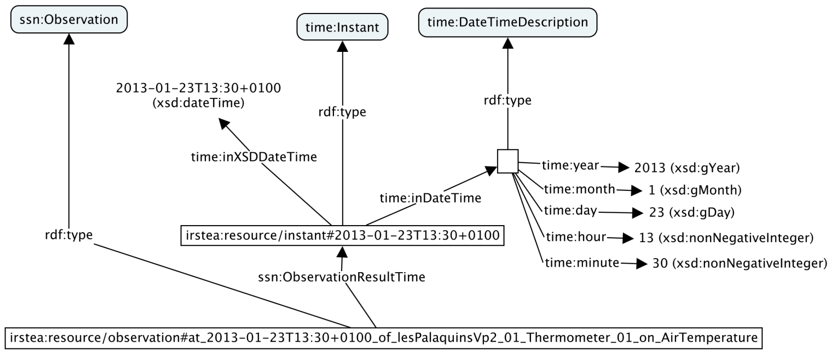

Fig. 4. Description of an observation related to a time instant

can be used for aggregation purposes by data consumers when accessing our SPARQL endpoint.

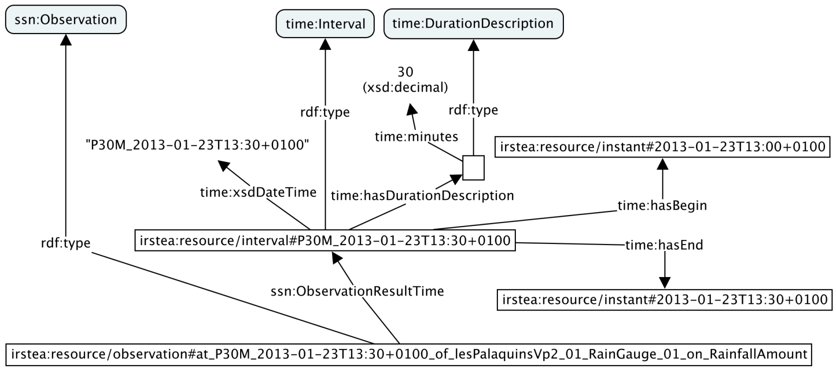

Fig. 5. Description of an observation related to a time interval

Figure 5 presents an example where the observation is related to a time interval (for instance, in the case of rainfall amount). In this case, we follow a similar pattern, but the value of the property $ssn : observationResultTime$ is instead an instance of the class $time : Interval$, and we also use the properties $time : hasDurationDescription$ and $time : hasEnd$ to specify the duration of the interval and its end, besides the beginning of the interval.

Table 4 contains a summary of the type of timestamp that we have associated to each type of measurement that our weather station sensors perform.

6 Implementation of the Data Transformation Process

Now we describe briefly the process followed for the generation of the RDF data according to the design principles described in the previous section. All the data that we have generated are available at our SPARQL endpoint[20].

As explained in the Introduction, data measured by the weather station sensors are stored in CSV files. These measurements are performed every half or quarter an hour. Figure 6 shows the main data processing flow that we apply.

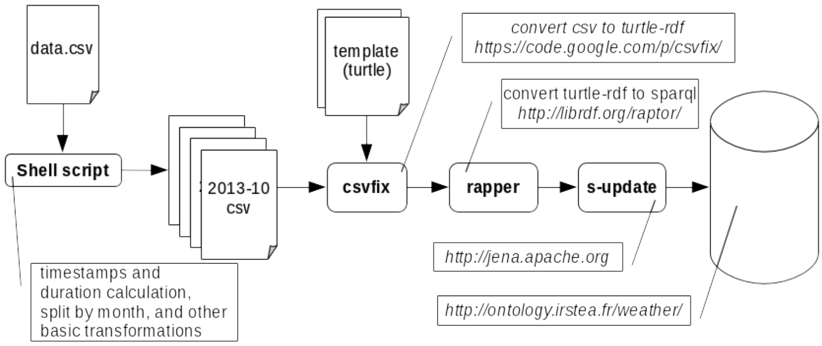

Fig. 6. Data Processing Flow for the Generation of RDF Data from the Weather Station Measurements

Pre-processing and date-related operations. A first pre-processing step is done using shell functions, which allows easy and quick conversions and substitutions, such as deleting line headers, detection and removal of ill-formed lines, and the substitution of initial tab characters by a semicolon, better accepted by the next tools in the pipeline (e.g., the csvfix tool). Moreover, we used standard Unix tools, for the following purposes:

- to convert dates as they were stored in the CSV file into the ISO 8601 format,
- to calculate the time interval between two timestamps,
- to convert the wind direction in degree.

The last pre-processing step is to group data by month, as a pragmatic way of processing batches of data and allowing an easier load into our Fuseki server.

CSV to RDF Transformations. The conversion of CSV data to RDF Turtle is made using the *csvfix*[21] toolbox for CSV file, which allows converting CSV data to other formats using template files.

[20] `http://ontology.irstea.fr/weather/query/`
[21] `https://code.google.com/p/csvfix/`

We have created some RDF Turtle format template files which are used in this process[22]. Finally, the data are sent to the server using *s-update*, which is part of SOH (SPARQL over HTTP) from the Jena framework[23].

7 Conclusions, Perspectives and Recommendations

As discussed in the Introduction, our intention with this work lied mainly at providing an implementation report about the usage of a set of ontologies that have been so far associated to the generation and publication of Linked Data from weather stations. This work tries to reflect which parts of the existing documented examples need to be refined or extended, and which parts are already well described and reproducible for a weather station.

We hope that our examples and code can provide further information to people that will be involved in a similar processing pipelines. Besides that, we have the following recommendations for the following versions of this set of ontologies:

- In general, we think that we still need a better documentation for many of these ontologies, in order to facilitate their use. This covers both documenting some of the classes and properties that are defined in these ontologies and providing more examples of usage. This does not say that the current documentation is bad at all. However, the documentation is still sometimes difficult to understand by people who are not specialists in ontologies. It would be interesting to associate it with concrete examples showing good practices in publication of data collected by sensors, such as what we try to do in this paper, or what is available in many other papers as well as in the W3C SSN Incubator Group implementation report. The lack of documentation is also a limitation of the $cf - feature$ ontology, as discussed in this paper, what makes it sometimes difficult to take decisions about which instances of properties to use for a specific type of measurement.
- There is still too much heterogeneity in ontologies that can be used to express geospatial information (W3C WGS84, GeoSPARQL, the ISA Core Location Vocabulary, etc.). This forces us to spend a lot of time deciding which ontologies to use for describing each type of geospatial information (locations, spatial points, polygons, etc.), and in some cases many alternatives exist, what may make some applications and SPARQL queries to work for some SPARQL endpoints and fail in others.
- The $cf - feature$ and $cf - property$ ontologies should be merged in one ontology in order to clarify their usage.
- The pattern to use for the specification of time instants and intervals associated to observations has to be unified as well with clear guidelines, something that is not done because of the lack of a range for the ssn :

[22] These template files and the processing pipeline code will be published on `http://ontology.irstea.fr`

[23] `http://jena.apache.org/documentation/serving\_data/soh.html`

observationResultTime property. We have seen examples in other datasets where the observation time is provided directly as an xsd:dateTime, whereas in our case we have preferred using the W3C Time Ontology so as to allow easier SPARQL queries when aiming at doing aggregations through the GROUP BY operator available in SPARQL.

In the coming months we will do similar processes to other weather stations owned by Irstea in other sites in France, so that they can be used by other researchers outside our institution in their research. We will also work with the association of quality indicators to data measured by the weather station or calculated by different processes over the initial data that have been captured.

References

1. Best practices for publishing linked data. http://www.w3.org/TR/ld-bp/. Accessed: 2014-07-08.
2. Climate and forecast (cf) features. `http://www.w3.org/2005/Incubator/ssn/ssnx/cf/cf-feature`. Accessed: 2014-06-30.
3. Isa programme location core vocabulary. `http://www.w3.org/ns/locn#`. Accessed: 2014-06-30.
4. Library for quantity kinds and units: schema, based on qudv model omg sysml(tm), version 1.2. `http://www.w3.org/2005/Incubator/ssn/ssnx/qu/qu`. Accessed: 2014-06-30.
5. Ogc geosparql - a geographic query language for rdf data. `http://www.opengeospatial.org/standards/geosparql`. Accessed: 2014-06-30.
6. Ontology for meteorological sensors. `http://www.w3.org/2005/Incubator/ssn/ssnx/meteo/aws`. Accessed: 2014-06-30.
7. Owl code of the time ontology. `http://www.w3.org/2006/time`. Accessed: 2014-06-30.
8. Semantic sensor network xg final report. `http://www.w3.org/2005/Incubator/ssn/XGR-ssn-20110628/`. Accessed: 2014-06-30.
9. G Atemezing, O Corcho, D Garijo, J Mora, Poveda-Villalon M., P Rozas, D Vila-Suero, and B Villazon-Terrazas. Transforming meteorological data into linked data. *Semantic Web journal, Special Issue on Linked Dataset descriptions, 2012. IOS Press, ISSN: 1570-0844*, 07 2012.
10. J-P Calbimonte, H Jeung, O Corcho, and K Aberer. Semantic sensor data search in a large-scale federated sensor network. In K Taylor, A Ayyagari, and D De Roure, editors, *SSN*, volume 839 of *CEUR Workshop Proceedings*, pages 23–38. CEUR-WS.org, 2011.
11. Jerry R. Hobbs and Feng Pan. Time ontology in owl. *W3C Working Draft*, 09 2006.
12. L. Lefort, J. Bobruk, A. Haller, K. Taylor, and A. Woolf. A linked sensor data cube for a 100 year homogenised daily temperature dataset. In *In 5th International Workshop on Semantic Sensor Networks (SSN-2012), CEUR-Proceedings*, 2012.
13. M. Stocker, Baranizadeh E., H Portin, M. Komppula, M. Rönkkö, A. Hamed, A. Virtanen, K. Lehtinen, A. Laaksonen, and M. Kolehmainen. Representing situational knowledge acquired from sensor data for atmospheric phenomena. *Environmental Modelling and Software*, 58:27–47, 2014.

measurement	feature of interest instance of $cf-feature$	property instance of $cf-property$
outdoor temperature	air	air_temperature
atmospheric pressure	air	air_pressure
humidity	air	relative_humidity
wind direction	wind	wind_from_direction
wind speed	wind	wind_speed
quantity of precipitation	rainfall	rainfall_amount
speed of precipitation	rainfall	rainfall_rate
solar radiation	ground_level_soil	downward_heat_flux_ at_ground_level_in_soil

Table 2. Weather measurement properties

Object	Resource class	local ID pattern
Station	$ssn:System$	⟨location Name⟩⟨ station type ⟩_⟨number ID⟩
Windwane	$ssn:Sensor$	⟨station ID⟩_⟨sensor type⟩_⟨number ID⟩
Instant	$time:Instant$	⟨date⟩T⟨time⟩⟨time zone⟩ (ISO 8601 format)
Interval	$time:Interval$	P⟨period⟩⟨unit⟩_⟨date⟩T⟨time⟩⟨time zone⟩
Observation	$ssn:Observation$	at_⟨timestamp⟩_of_⟨sensor⟩_on_⟨property⟩
Examples		
URI station	`irstea:resource/system#lesPalaquinsVp2_01`	
URI windwane	`irstea:resource/sensor#lesPalaquinsVp2_01_WindVane_01`	
URI instant	`irstea:resource/instant#2013-01-23T13:30+0100`	
URI interval	`irstea:resource/interval#P30M_2013-01-23T13:30+0100`	
URI observation	`irstea:resource/observation#at_2012-06-19T15:45+0200_` `of_lesPalaquinsVp2_01_Thermometer_01_on_AirTemperature`	

Table 3. URI generation templates for resources

Measured property	timestamp
exterior temperature	instant
atmospheric pressure	instant
air humidity	instant
wind direction	interval
wind speed	interval
rainfall amount	interval
rainfall speed	interval
solar radiation	interval

Table 4. Types of timestamps associated to the measured properties of our weather station