Semantic Integration of Geospatial Data from Earth Observations

Helbert Arenas, Nathalie Aussenac, Catherine Comparot, and Cassia Trojahn

Institut de Recherche en Informatique de Toulouse, Toulouse, France
{helbert.arenas, nathalie.aussenac-gilles, catherine.comparot, cassia.trojahn}@irit.fr

Abstract. We propose an approach to semantically enrich metadata records of satellite imagery with external data. As a result we are able the identify relevant images using a larger set of matching criteria. Conventional methods for annotating data sets are usually based on metadata records (with attributes such as title, provider, access mode, and spatio-temporal characteristics), which offer a narrow view of the world. Enriching metadata with contextual information (i.e., the region depicted in the image has been recently affected by extreme weather) requires formalizing spatio-temporal relationships between metadata records and external data sources. Semantic technologies play a key role in such scenarios by providing an infrastructure based on RDF and ontologies.

1 Introduction

Traditional approaches for the description and identification of satellite imagery make use of metadata records. However they are limited in the sense that they consider metadata records as isolated entities. In our work, we propose the use of ontologies to define formal relationships between metadata records and external data sources. Thus we are able to create links to useful pieces of information in the image spatio-temporal proximity. This new information provides the image with context, enhancing its description. This work is carried out within the SparkInData project that aims at building a cloud platform for Earth Observation data. SparkInData is part of the "Investing for the Future" French program.

In order to cover our annotation needs (image metadata, earth observation, sensors, and specific contextual data, with special attention to spatio-temporal data), we use three well recognized and supported ontologies: 1) Data Catalog Vocabulary (DCAT), a W3C standard designed to handle catalog records on the web [6]; 2) Semantic Sensor Network (SSN) another W3C standard designed to describe sensors and observations [1]; and 3) GeoSPARQL, an OGC standard designed as a minimum ontology to query geospatial data on the Semantic Web [5]. The usage of standard-based vocabularies guarantees maximum interoperability.

This paper details the proposed model for spatial and temporal data integration. To illustrate our approach we use weather information as an external data source. We model weather information using elements from SSN, while the linkage between weather information and satellite images metadata records is done using GeoSPARQL.

2   Data sets

The weather information is provided by *Meteo France*. This organization offers data as monthly compiled CSV zipped files. The observations are taken every three hours for each one of the 62 weather stations in France. A separate file contains a list of the weather stations with their position as points encoded as geographic coordinates [7]. We have developed an application in Java to map the information contained in the CSV files into our Data Model as RDF.

The satellite image metadata records are obtained from RESTO, a data service managed by CNES (Centre National d'Etudes Spatiales) [4]. The metadata records are stored in a relational database and published as JSON. Each image has its own spatial representation encoded as a polygon with geographic coordinates. The revisit time for Sentinel 1 is 12 days, while for Sentinel 2 it is 5 days [2] [3]. We have developed a software tool to translate the records to RDF, it uses the Jena and GeoTools libraries.

3   A data model for enriched geospatial metadata records

Figure 1 depicts the proposed data model. The class *SatelliteImgMetadataRecord* is a subclass of *dcat:Dataset*. This specialization is required in order to represent properties specific of satellite imagery, such as the platform (what Satellite is taking the image), the processing level, the sensor mode (which depends on the satellite and the observed surface). A satellite image metadata record has a spatial dimension and a temporal dimension. Both of these dimensions are used to add external data to the image.

The model handles the **spatial dimension** of a metadata record thanks to the property *hasSpatialRepresentation* and the GeoSPARQL ontology. This property has as range the class *geo:Feature* and as domain the class *SatelliteImgMetadataRecord*. Thanks to this property we are able to use GeoSPARQL functions to relate the metadata records with any other information with a spatial component and also defined as a *geo:Feature*. GeoSPARQL provides formal definitions of topological relationships (RCC8). Triplestores that support GeoSPARQL, also offer a set of spatial functions that enable geospatial operations, such as *geof:distance, geof:intersection, geof:difference*, etc. By using GeoSPARQL, we can create statements that describe the topological relationships between a satellite image footprint and other entities with a spatial nature.

In order to provide contextual information to the metadata records we link them to weather information that we model using both the Semantic Sensor Network (SSN) and the GeoSPARQL vocabularies. We represent a weather station as an instance of the class *MeteoFranceWeatherStation*, which is at the same time a subclass of *ssn:SensingDevice* and as a subclass of *geo:Feature*. Then it is possible to compare the two instances of the class *geo:Feature* and to determine their spatial relationships.

The *temporal dimension* of an image metadata record identifies the moment when the image has been captured. The external datasource, the weather information, also has a temporal dimension. Weather stations record measurements

periodically. So we use the class *ssn:Observation* as a way to relate the measured variables to the weather station while at the same time, providing a temporal dimension for the observations (See Figure 1). Then, it is possible to establish temporal relationships (before, after) between an image metadata record and weather measurements or to establish periods of interest (for instance, one week after the image was created). Using the model, the user can define a relevant period of time and link an image metadata record to the available weather information. The user defined period works as a temporal buffer that provides context to the image metadata record.

The following code depicts a SPARQL query that links image metadata records with weather measurements corresponding to stations located inside the image footprint (cf. *FILTER(geof:sfContains(?img_md_geo,?ws_geo))*) and the data collected during a given period. The user defined a period of interest of one week after the image was taken (24x7=168 hours, 'PT168H') that is used to filter the relevant data (cf. the last three FILTER lines) [1].
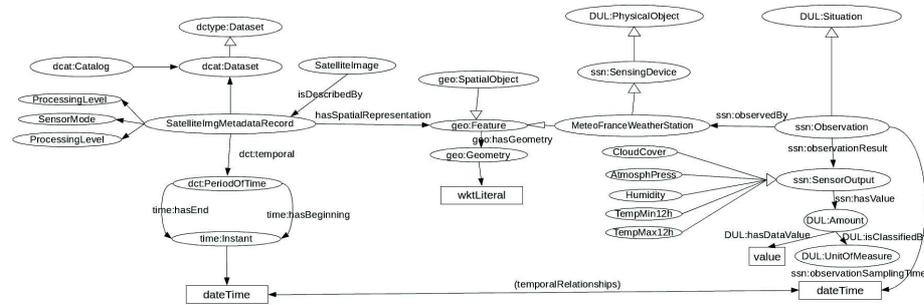

Fig. 1. Main classes used in our model.

```
select ?img_md ?img_identifier ?obs_time ?humidityDataVal ?humidityUnits
where{
?img_md a sparkindata:SatelliteImgMetadataRecord.
?img_md dct:identifier ?img_identifier.
?img_md sparkindata:hasSpatialRepresentation ?img_md_feat.
?img_md_feat geo:hasGeometry ?img_md_geo.
?img_md dct:temporal ?img_md_temp.
?img_md_temp time:hasEnd ?img_md_end.
?ws a sparkindata:MeteoFranceWeatherStation.
?ws geo:hasGeometry ?ws_geo.
?obs ssn:observedBy ?ws.
?obs ssn:observationSamplingTime ?obs_time.
?obs ssn:observationResult ?humidity.
```

---

[1] A video showing the implementation of our approach is available at: http://geo-space.info/demos/metadataBrowserDataMeteoFrance.mp4

```
?humidity a sparkindata:Humidity.
?humidity ssh:hasValue ?humidityVal.
?humidityVal DUL:hasDataValue ?humidityDataVal.
?humidityVal DUL:isClassifiedBy ?humidityUnits.
BIND (?img_md_end-?obs_time as ?diffDateTime)
FILTER(geof:sfContains(?img_md_geo,?ws_geo))
FILTER (?diffDateTime<'PT168H'^^xsd:dayTimeDuration)
FILTER (?diffDateTime>'PT0H'^^xsd:dayTimeDuration)}
```

4   Concluding remarks and future work

We have presented an approach for semantic annotation of satellite images that relies on a model providing context to image metadata records by using external data sources. The proposed approach is rather generic and is based on well-known existing vocabularies. We show the feasibility of this approach by using weather information as the contextual data. Currently, we are working on extending this approach to use as contextual information phytosanitary reports and seismic information. As future work, we plan to integrate automatic matching approaches for matching raw image metadata and vocabularies as well as to align semantic annotated data with other sources of data in the Linked Open Data. We plan as well to consider domain-oriented sources of data for a particular use case (agriculture and data sources as agricultural reports) and provide rules and reasoning capabilities that help the specific domain analysis.

Acknowledgement

References

1. M. Compton, P. Barnaghi, L. Bermudez, R. García-Castro, O. Corcho, S. Cox, J. Graybeal, M. Hauswirth, C. Henson, A. Herzog, V. Huang, K. Janowicz, W. D. Kelsey, D. Le Phuoc, L. Lefort, M. Leggieri, H. Neuhaus, A. Nikolov, K. Page, A. Passant, A. Sheth, and K. Taylor. The SSN ontology of the W3C semantic sensor network incubator group. *Journal of Web Semantics*, 17:25–32, 2012.
2. ESA.   Sentinel 1.   `https://sentinel.esa.int/web/sentinel/missions/sentinel-1`, 2016. Accessed:2016-07.
3. ESA.   Sentinel 2.   `https://sentinel.esa.int/web/sentinel/missions/sentinel-2/`, 2016. Accessed:2016-07.
4. J. Gasperi. Semantic Search Within Earth Observation Products Database Based on Automatic Tagging of Image Content. In *Proceedings of the Conference on Big Data from Space*, pages 4–6, 2014.
5. D. Kolas, M. Perry, and J. Herring. Getting started with GeoSPARQL. Technical report, OGC, 2013.
6. F. Maali and J. Erickson. Data Catalog Vocabulary (DCAT), 2014.
7. Ministere de l'Ecologie du Developpement durable et de l'Energie.   Meteo France,Donnes Publiques.   `https://donneespubliques.meteofrance.fr`, 2016. Accessed:2016-07.