



Un modèle pour l'intégration sémantique de données géolocalisées liées à l'observation de la terre (Atelier SAGEO 2016)

Helbert Arenas, Nathalie Aussenac-Gilles, Catherine Comparot, Cassia Trojahn dos Santos

► To cite this version:

Helbert Arenas, Nathalie Aussenac-Gilles, Catherine Comparot, Cassia Trojahn dos Santos. Un modèle pour l'intégration sémantique de données géolocalisées liées à l'observation de la terre (Atelier SAGEO 2016). Conference Spatial Analysis and Geomatics (SAGEO 2016), Dec 2016, Nice, France. pp.1-14. hal-01709127

HAL Id: hal-01709127

<https://hal.science/hal-01709127>

Submitted on 14 Feb 2018

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.



Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/>
Eprints ID : 18827

The contribution was presented at SAGEO 2016 :
<https://sageo2016.sciencesconf.org/>

To cite this version : Arenas, Helbert and Aussenac-Gilles, Nathalie and Comparot, Catherine and Trojahn, Cassia *Un modèle pour l'intégration sémantique de données géolocalisées liées à l'observation de la terre*. (2016) In: 12eme edition de la conference SAGEO : Spatial Analysis and Geomatics - Atelier EXtraction de Connaissances à partir de données Spatialisées (SAGEO - EXCES 2016), 6 December 2016 - 6 December 2016 (Nice, France).

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

Un modèle pour l'intégration sémantique de données géolocalisées liées à l'observation de la terre

Helbert Arenas, Nathalie Aussenac-Gilles,
Catherine Comparot, Cassia Trojahn

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

RÉSUMÉ. Le domaine de l'observation de la terre est en forte évolution. L'ESA a récemment lancé les satellites Sentinel 1 et Sentinel 2 qui livrent entre 8 à 10 To de données par jour, ce qui ouvre de nouvelles opportunités d'applications métier. Un des challenges à relever est alors l'intégration de données hétérogènes provenant de différentes sources. Les technologies du web sémantique, en fournissant une infrastructure basée sur RDF et des ontologies, jouent un rôle clé dans ce contexte. Nous présentons une approche pour enrichir les métadonnées habituelles des images satellites avec des données externes (e.g. la région représentée sur l'image a été récemment affectée par une canicule) de manière à disposer d'un ensemble plus important de critères pour les retrouver. Nous proposons un vocabulaire sémantique et une formalisation des relations spatio-temporelles support au processus d'intégration des différentes données géolocalisées associées à des observations de la terre.

ABSTRACT. Earth Observation is a rapidly evolving domain. ESA recently launched the Sentinel 1 and 2 satellites, which will deliver between 8 and 10 TB of data per day, opening the opportunity to emerging new domain-oriented applications. One of the challenges to be addressed in the field concerns the integration of heterogeneous data coming from different sources. In this context, semantic technologies play a key role by providing an infrastructure based on RDF and ontologies. We present an approach to enrich traditional metadata of satellite imagery with external data (e.g., the region depicted in the image has been recently affected by extreme weather) which allows to identify relevant images using a larger set of matching criteria. We propose a semantic vocabulary and a formalization of spatio-temporal relationships to be used as basis for the integration process of various geo-spatial data linked to earth observation.

MOTS-CLÉS : intégration de données spatio-temporelles, modèle sémantique, images satellites

KEYWORDS: data integration, semantic model, satellite images, spatio-temporal data

1. Introduction

L'observation de la terre offre une valeur ajoutée à une grande diversité de domaines. Récemment l'Agence Spatiale Européenne (ESA) a lancé deux nouveaux satellites, Sentinel 1 et Sentinel 2, qui diffuseront en continu des images de la Terre captées selon différentes technologies. Il est prévu que Sentinel 1 et Sentinel 2 transmettent quotidiennement des images de haute qualité, soit entre 8 à 10 To de données. Ce volume de données ouvre de nombreuses perspectives économiques grâce à de nouvelles applications dans des domaines aussi variés que l'agriculture, l'environnement, l'urbanisme, l'océanographie ou encore la climatologie. Cependant, comme mentionné dans (EARSC, 2016), il manque encore des outils pour permettre aux utilisateurs d'accéder facilement aux données produites. Ainsi, pour pouvoir tirer pleinement avantage de ces nouvelles sources de données, un des challenges est de trouver l'information pertinente au sein de cet amas de plus en plus grand de données générées de façon continue. La recherche des images peut être facilitée par une description plus complète à l'aide de données provenant de sources hétérogènes (métadonnées d'images de télédétection acquises à partir de satellites, d'informations sur le sol, sur le climat, etc.). Cependant, l'hétérogénéité de ces données rend plus complexe leur intégration et leur analyse.

Les approches classiques pour décrire et identifier des images satellites exploitent des ensembles de métadonnées. Ces approches sont limitées car elles considèrent les ensembles de métadonnées comme des entités isolées et stables. Par exemple, ce qui ne permet pas d'indiquer que l'image représente une région récemment touchée par un phénomène tel qu'un tremblement de terre ou une canicule. Pour disposer d'une information contextuelle de ce genre il est nécessaire de formaliser les relations spatio-temporelles entre les ensembles de métadonnées et des sources de données externes. Un terme générique pour ce genre de scénario est *big data*. Ce terme fait référence aux ensembles de données caractérisés de : 1) volumineux, 2) variés (hétérogènes), and véloces (mis à jour fréquemment) (Zaslavsky *et al.*, 2013). Ces caractéristiques du big data nécessitent une ré-évaluation des techniques et méthodes précédentes (Miller, 2010). C'est le cas quand on traite les nouvelles images d'observation de la terre. Ces collections sont à la fois volumineuses, enrichies au fil du temps, et hétérogènes dès qu'on leur associe des données externes. Dans ces scénarii, les technologies du web sémantique peuvent améliorer l'accès, la gestion et l'analyse de données géographiques. L'utilisation de la sémantique dans le domaine de l'Observation de la Terre a reçu une attention particulière ces dernières années. Comme indiqué dans (Reitsma, Albrecht, 2005), l'expression et l'utilisation de la sémantique lors de la modélisation de la terre conduit à de nouvelles perspectives pour les systèmes environnementaux étudiés et une meilleure compréhension des méthodes utilisées pour représenter ces systèmes. Par ailleurs, la

découverte et la recherche sémantiques d'informations géographiques valorisent les données temporelles et spatiales en tant que supports au raisonnement et à une découverte automatisée.

Cet article présente un processus d'intégration sémantique de données géolocalisées à partir d'observations terrestres. Nous proposons un vocabulaire et une formalisation des relations spatio-temporelles qui servent de base pour intégrer aux métadonnées d'images différentes données (données de capteurs météo ou sismiques, nature des sols, etc). Nous proposons d'utiliser des ontologies pour définir formellement les relations entre les enregistrements de métadonnées d'une image et les sources de données externes. Cette information additionnelle, spatio-temporellement proche de l'image, lui fournit un contexte et améliore sa description. Nous montrons, à travers un cas d'étude, comment l'intégration de la sémantique peut améliorer l'accès aux données. Ce travail est mené dans le cadre du projet SparkInData¹ visant à construire une plateforme cloud de données d'observations de la Terre. Le cas d'étude s'appuie sur l'annotation sémantique de métadonnées d'images brutes fournies par le CNES (Centre National d'Etudes Spatiales).

Après l'exposé de travaux connexes (partie 2), nous décrivons, dans la partie 3, les vocabulaires utilisés ainsi que le modèle de données proposé pour l'intégration de données à partir de relations spatio-temporelles. Dans la partie 4, nous montrons, à travers un cas d'usage, l'exploitation de ce vocabulaire pour annoter des métadonnées d'images, et illustrons en partie 5 comment ces métadonnées permettent de répondre à des besoins utilisateurs.

2. Travaux liés

Dans le domaine de la sémantique pour l'imagerie satellitaire, notons le projet européen TELIOS (Virtual Observatory Infrastructure for Earth Observation Data)² dont l'objectif était de faciliter l'accès à des données d'observation de la terre à des agences telles que la NASA ou l'ESA. Son cadre applicatif est la gestion des incendies en Grèce (The TELEIOS Team, 2012). Pour cela, les images sont classifiées et vectorisées, et à chaque vecteur est associée une valeur d'estimation des risques d'incendie. Les vecteurs sont liés à des sources de données externes par le biais d'ontologies (TELEIOS, 2016), notamment le standard stRDF qui étend RDF pour les données spatiales et temporelles, mais les analyses réalisées ne prennent pas en compte la composante temporelle.

Le travail présenté dans (Espinoza-Molina, Datcu, 2013) et (Espinoza-Molina *et al.*, 2015) a été influencé par TELEIOS. Un système permet d'ajouter des annotations sémantiques à des produits issus de radars à synthèse d'ouverture

1. SparkInData fait partie du programme français d'Investissement d'Avenir (FUI).

2. <http://www.earthobservatory.eu> Accès : 2016-12-01

(SAR) selon le processus suivant : 1) extraction des informations pertinentes à partir des fichiers de métadonnées XML ; 2) division de l'image en dalles de tailles différentes, organisation de ces dalles sous une forme pyramidale ; et 3) extraction de caractéristiques de chaque dalle, ensuite transformées en vecteurs. Ce système est capable de trouver de l'information en utilisant des requêtes portant sur les métadonnées, la sémantique et le contenu spatial des images. Cette approche rejoint notre objectif puisqu'elle associe aux images d'observation de la terre des données provenant des sources externes, comme Corine Land Cover, Urban Atlas, Geonames et LinkedGeoData (Espinoza-Molina *et al.*, 2015). Elle réutilise l'ontologie SAR qui étend celle présentée dans (TELEIOS, 2016). La composante temporelle n'y est pas présente.

L'approche de (Keßler, Farmer, 2015) utilise à la fois des sources de données externes et des relations temporelles. Elle s'appuie sur un concept appelé *prisme spatio-temporel* ("space-time prism") qui décrit les caractéristiques des trajectoires. Un outil très connu pour l'analyse spatiale est le *buffer* ; il s'agit d'une zone située autour d'une région d'intérêt, mesurée dans une certaine unité spatiale. Le *prisme spatio-temporel* est similaire à un *buffer* auquel on a ajouté une dimension spatiale. Une fois la zone autour de la caractéristique d'intérêt définie, il est possible d'identifier d'autres propriétés contenues dans le *prisme spatio-temporel*. Dans les trajectoires étudiées dans (Keßler, Farmer, 2015), certains points sont caractérisés par des valeurs temporelles de type *xsd:dateTime*. Les valeurs extraites déterminent si un élément est contenu ou non dans un *prisme spatio-temporel* donné.

Notre contribution va au delà de ces travaux en exploitant à la fois deux composantes, temporelle et spatiale pour annotation sémantiquement des métadonnées d'images. Ceci nous a conduit à définir un nouveau vocabulaire et une approche associée pour traiter les relations temporelles.

3. Un vocabulaire pour intégrer les données spatio-temporelles

Notre objectif est d'utiliser des métadonnées d'image, des observations de la terre fournies par des mesures de capteurs et des données contextuelles spécifiques, ayant à la fois une dimension spatiale et temporelle, pour annoter des images par des données sémantiques. La sémantique des données repose sur leur typage et leur mise en relation à l'aide d'un vocabulaire unique. Comme le recours à des standards améliore l'interopérabilité, ce vocabulaire réutilise trois ontologies reconnues et maintenues : 1) DCAT (Data Catalog Vocabulary), un standard du W3C conçu pour permettre l'accès à des enregistrements de catalogues via le web (Maali, Erickson, 2014) ; 2) SSN (Semantic Sensor Network), un autre standard du W3C conçu pour décrire les capteurs et observations issues de ces capteurs (Compton *et al.*, 2012) ; et GeoSPARQL, un standard de l'OGC conçu comme une ontologie minimale pour interroger des données géolocalisées sur le web sémantique (Kolas *et al.*, 2013).

3.1. Les données et vocabulaires utilisés

3.1.1. DCAT

DCAT³ est un vocabulaire RDF recommandé par le W3C pour publier des catalogues sur le web. Il combine des éléments d'autres vocabulaires tels que Dublin Core Terms (dct), Time, SKOS et VCard. La figure 1 montre les principales classes et propriétés de DCAT que nous avons utilisées. La classe *dcat:Dataset* représente un enregistrement de métadonnées. Elle possède de nombreuses propriétés telles que *dct:title*, *dct:description*, *dct:issued*, *dct:language*, *dct:identifier* permettant de décrire un enregistrement de métadonnées (Maali, Erickson, 2014). Nous avons retenu d'autres propriétés, en particulier *dct:temporal* et *dct:spatial* qui rendent compte respectivement des dimensions temporelle et spatiale d'un ensemble de données. Le codomaine de la propriété *dct:spatial* est *dct:Location* qui définit une position. Celui de la propriété *dct:temporal* est un intervalle de temps représenté par la classe *dct:PeriodOfTime*, qui fait appel aux propriétés *time:hasEnd* et *time:hasBeginning* et à la classe *time:Instant* de l'ontologie OWL-Time⁴.

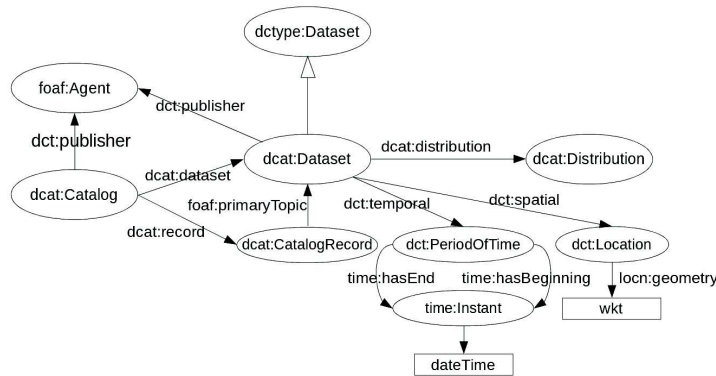


FIGURE 1. Principales classes de DCAT.

3.1.2. GeoSPARQL

GeoSPARQL (Kolas *et al.*, 2013) (Battle, Kolas, 2012), standard de l'OGC qui définit une petite ontologie pour la représentation de caractéristiques, de relations et de fonctions spatiales, joue un rôle majeur dans le domaine de la sémantique géospatiale. Il existe des alternatives à GeoSPARQL comme GeoRDF

3. <https://www.w3.org/TR/vocab-dcat/>

4. <https://www.w3.org/TR/owl-time/>

qui permet de représenter des données simples telles que la latitude, la longitude, l'altitude, comme des propriétés de points (en utilisant WGS84 comme référentiel), ou encore GeoOWL qui permet d'exprimer des objets plus spatiaux (lignes, rectangles, polygones). Nous avons retenu GeoSPARQL car il offre des possibilités pour raisonner sur des géométries, à savoir proposer des relations entre des entités sur la base des relations entre géométries (inclusion, recouvrement, ...). De plus, certains entrepôts RDF (ou triplestores) assurent la gestion de données géométriques à l'aide de GéoSPARQL : OWLIM, Open Sahara (un wrapper pour PostgreSQL+PostGIS+Sesame Sail layer), Parliament (technologies Raytheon BBN), et AllegroGraph. On peut également citer Apache Marmotta, Strabon, Oracle Spatial, ainsi que Stardog qui propose GeoSPARQL depuis sa version 4.0 en 2015.

Dans l'ontologie GeoSPARQL (figure 2), la classe *geo:Feature* représente toute entité du monde réel ayant une localisation spatiale (Perry, Herring, 2012). Cette localisation est décrite par une géométrie (point, polygone, etc), qui est une instance de la classe *geo:Geometry*. Une entité est liée à sa géométrie via la propriété *geo:hasGeometry*. Les coordonnées d'une géométrie sont décrites via la propriété *geo:asWKT*. Il est possible de lier GeoSPARQL à une ontologie de n'importe quel domaine en spécialisant la classe *geo:Feature* par une classe de l'ontologie de ce domaine.

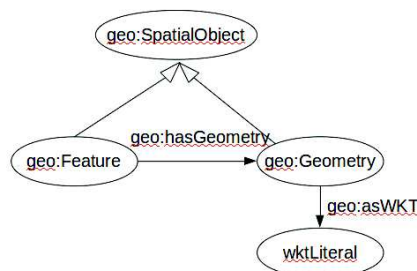


FIGURE 2. Principales classes de GeoSPARQL.

3.1.3. SSN, Semantic Sensor Network

Le deuxième vocabulaire réutilisé est SSN, un vocabulaire de description de capteurs numériques en réseau. En effet, il y a actuellement un grand nombre de capteurs déployés qui récupèrent sur la terre de l'information de domaines divers tels que la gestion de l'environnement, la sécurité, la météo, etc. Ces capteurs sont des sources indéniables pour associer un contexte à des images. L'information fournie par les capteurs est hétérogène et soulève des difficultés d'intégration à d'autres sources de données. En 2009, le groupe SSN du W3C a présenté cette ontologie pour décrire les fonctionnalités, propriétés et mesures

de capteurs numériques (Compton *et al.*, 2012). SSN est alignée avec DOLCE-UltraLite (DUL) pour limiter les interprétations possibles. Elle contient 41 concepts et 39 propriétés d'objets (*ObjectProperty*). L'ontologie permet de décrire des capteurs à n'importe quel niveau de détail. Elle est générique : elle ne définit pas toutes les propriétés possibles des capteurs et des observations. Par contre, il est possible de la lier à des ontologies ou des vocabulaire spécifiques à un domaine tels que SWEET (Calbimonte *et al.*, 2012). Dans (Atemezing *et al.*, 2013), les auteurs présentent une méthodologie pour publier des informations météorologiques de l'office météorologique espagnol (AEMET)⁵ en utilisant SSN couplée à l'ontologie OWL-Time⁶. La figure 3 montre les principales classes du réseau d'ontologies qu'ils proposent.

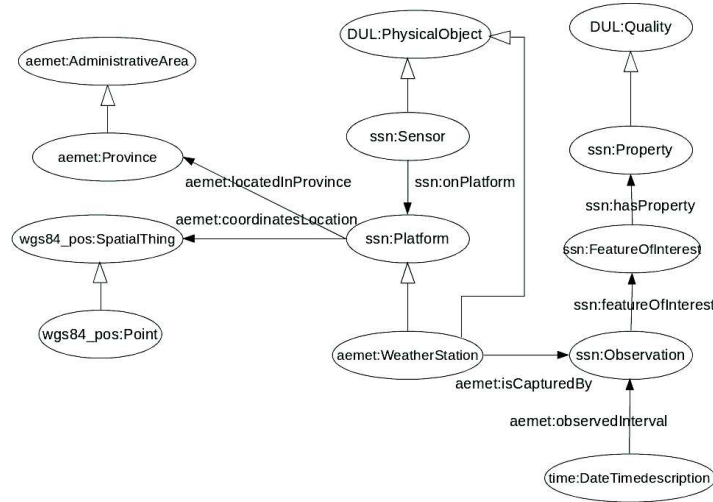


FIGURE 3. Principales classes du réseau d'ontologies Aemet.

3.2. Un modèle des données pour intégrer des données géo-temporelles

Le modèle que nous proposons est présenté sur la figure 4. La classe *SatelliteImgMetadataRecord* est une sous-classe de *dcat:Dataset*. Cette spécialisation permet de représenter des métadonnées spécifiques à l'imagerie satellitaire en tant que données de catalogue. Nous avons associé trois propriétés à la classe *SatelliteImgMetadataRecord* :

5. The Aemet Network of Ontologies, <http://aemet.linkeddata.es/aemet-content/index.html>

6. <https://www.w3.org/TR/owl-time/>

- *hasPlatform* (1 en Fig.4) : la classe *Platform* correspond au satellite qui a produit l'image. Il y a actuellement deux types de satellites Sentinel déployés : deux satellites Sentinel 1 (S1A et S1B), en orbite polaire, produisent des images en utilisant un radar à synthèse d'ouverture (RSO); deux satellites Sentinel 2 (S2A et S2B) produisent des images en utilisant un instrument multispectral passif.
- *hasProcessingLevel* (2 en Fig.4) : cette propriété associe à l'image une entité de la classe *ProcessingLevel* qui représente les différents niveaux de traitement que peut réaliser le satellite. En fait, il y a différents niveaux de traitement propres à chaque type de satellite Sentinel 1 (*LEVEL0*, *LEVEL1*, *LEVEL2*) ou Sentinel 2 (*LEVEL1C*, *LEVEL2A*, *LEVEL3A*). Chaque niveau de traitement fournit des indications particulières sur l'image. Par exemple, *LEVEL0* pour une image SAR indique que l'on dispose des informations sur le bruit, la calibration interne, l'orbite et l'altitude de l'image (ESA, 2016a). De même, *LEVEL2A* pour une image Sentinel 2 signifie que l'image a été orthorectifiée et que les valeurs de ses pixels expriment la réflectance de la surface, avec des masques de nuages et d'ombres de nuages (ESA, 2016b) (Gatti, Bertolini, 2013).
- *hasSensorMode* (3 en Fig.4) : indique le mode opératoire du satellite pour cette image. Sentinel 1 et Sentinel 2 ont des modes opératoires différents utilisés en fonction du type de surface terrestre à observer. Ainsi Sentinel 1 utilise le mode de capteur *WV* (Wave Mode) quand il opère au-dessus des océans, le mode *emphIW* (Interferometric Wide-Swath Mode) pour la terre ou encore *EW* (Extra-Wide-Swath Mode) pour les zones polaires (ESA, 2016a) (Sentinel-1 Team, 2013).

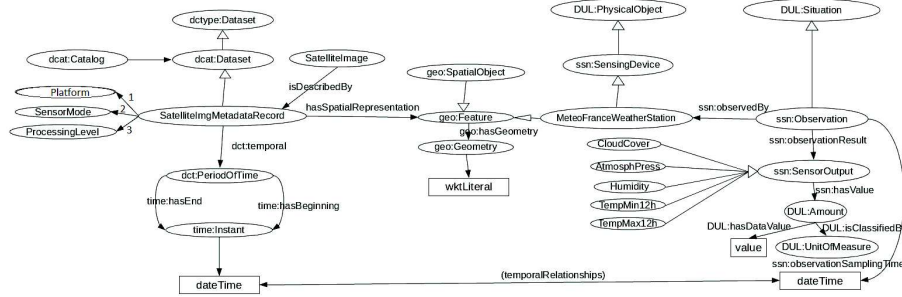


FIGURE 4. Principales classes de notre modèle de données.

En plus des données alphanumériques qui décrivent une image satellite, un enregistrement de métadonnées d'image contient une dimension spatiale et une dimension temporelle. Notre modèle intègre la dimension spatiale avec la propriété *hasSpatialRepresentation* et l'ontologie GeoSPARQL. Cette propriété a pour domaine la classe *SatelliteImgMetadataRecord* et pour codomaine la classe *geo:Feature*. Grâce à cette propriété, il est possible d'utiliser des fonctions

de GeoSPARQL pour lier les enregistrements de métadonnées avec n’importe quelle autre information ayant une composante spatiale, et définie comme une *geo:Feature*. GeoSPARQL fournit des définitions formelles pour les relations topologiques conformes aux relations topologiques RCC8 (Randell *et al.*, 1992). Les triplestores qui proposent GeoSPARQL offrent aussi des fonctions spatiales telles que *geo:distance*, *geo:intersection*, *geo:difference*, etc. GeoSPARQL permet de calculer des relations topologiques entre l’empreinte d’une image satellite (le *footprint*), et la composante spatiale d’autres entités.

Notre premier cas d’usage d’évaluation de ce modèle exploite une seule source externe de données, le site de *Météo France*. Depuis, nous avons étendu l’expérience en rajoutant des données sismiques (Arenas *et al.*, 2016). Les mesures météorologiques⁷ constituent des informations contextuelles complémentaires des métadonnées associées aux images. Tout autre source fournissant des informations datées et propres à une zone de la surface terrestre (nature du sol, sismographie, population, etc.) pourrait être ajoutée selon un même procédé.

Le site de *Météo France* fournit des données relevées toutes les trois heures dans chacune des 62 stations françaises de position fixe. Nous modélisons ces mesures grâce à la partie du modèle adaptée de SSN. Une station météo est représentée à l’aide de la classe *MeteoFranceWeatherStation*, qui est sous-classe à la fois de *ssn:SensingDevice* et de *geo:Feature*. On peut ainsi représenter la position des stations météorologiques et utiliser les fonctions de GeoSPARQL pour les relier à d’autres entités spatiales.

La propriété temporelle d’un enregistrement de métadonnées identifie le moment où l’image a été prise. Les données externes, dans notre cas les informations météo relevées périodiquement par les stations, ont aussi une dimension temporelle. Nous utilisons la classe *ssn:Observation* (figure 4) pour typer les variables mesurées et leur associer une composante temporelle (i.e. date de la mesure a été effectuée). Grâce aux propriétés spatiales et temporelles, il est possible d’établir des relations spatio-temporelles entre métadonnées et données externes (mesures météo). On peut ainsi déterminer quelles observations correspondent aux lieux couverts par l’image durant un temps donné.

4. Le processus d’intégration de données

4.1. Extraction des données

Nous exploitons des enregistrements de métadonnées d’images Sentinel 2 obtenus grâce à RESTO, un moteur de recherche d’images d’Observation de

7. Les variables disponibles sont décrites sur le site de Météo France : https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=90&id_rubrique=32. L’information est fournie en License Open ce qui autorise les utilisateurs à reproduire, diffuser, modifier et transformer les données originales.

la Terre qui est déployé sur un serveur du CNES et utilise un accès REST (Gasperi, 2014). Les métadonnées enregistrées dans une base de données relationnelle, sont publiés en JSON. On peut y accéder à l'aide de requêtes comme celle ci-dessous, qui recherche des enregistrements pour la région de Toulouse :

`https://peps.cnes.fr/resto/api/collections/S2/search.json?q=Toulouse`

Nous avons implémenté en Java un mécanisme qui réécrit en RDF les enregistrements ainsi obtenus. Chaque enregistrement devient une instance de la classe *SatelliteImageMetadataRecord* du modèle de données. Les triplets RDF ainsi obtenus sont insérés dans un entrepôt Stardog. De même, une application Java convertit les informations météorologiques des fichiers CSV de *Météo France* vers le modèle de données et les stocke en RDF dans le triplestore.

4.2. Recherche de métadonnées

Nous pouvons alors utiliser des filtres spatio-temporels pour rechercher des enregistrements de métadonnées pertinents. Nous avons développé un prototype sous forme d'application web qui montre la pertinence du modèle de données pour assurer ce type de recherche. L'interface de cette application (figure 5) permet à l'utilisateur de spécifier une recherche en deux temps : pour exprimer la contrainte spatiale, il dessine un rectangle sur la carte et indique ainsi la zone qui l'intéresse ; pour indiquer la période temporelle recherchée, il saisit deux dates dans les éditeurs situés à droite de la carte. Ainsi l'utilisateur n'a besoin ni d'aller consulter chacune des sources séparément ni de connaître le langage de requête donnant accès aux données de l'entrepôt.

5. Interrogation de l'entrepôt de données

L'intégration de données permet à l'utilisateur d'avoir un point d'entrée unique pour accéder à un ensemble d'informations (métadonnées d'origines et données externes) associées aux images, et ce l'aide de l'interface qui génère des requêtes GeoSPARQL. Le rectangle dessiné par l'utilisateur spécifie des contraintes spatiales (qui utilisent les relations spatiales) et la période de temps qu'il saisit définit un buffer temporel. A partir de ces informations, le système génère une requête qui permet de retrouver les données externes (observations météo), les méta-données et les images dont la composante temporelle recouvre le buffer temporel et dont la localisation vérifie les contraintes spatiales (plus précisément, l'empreinte de ces images a une intersection non nulle avec le rectangle choisi par l'utilisateur).

La requête SPARQL ci-dessous relie des enregistrements de métadonnées d'images à 1) des mesures météorologiques correspondant aux stations situées dans l'empreinte de l'image (Cf. *FILTER(geof:sfContains(?img_md_geo,?ws_geo))*), et 2) à des données collectées durant une période donnée. L'utilisateur a défini

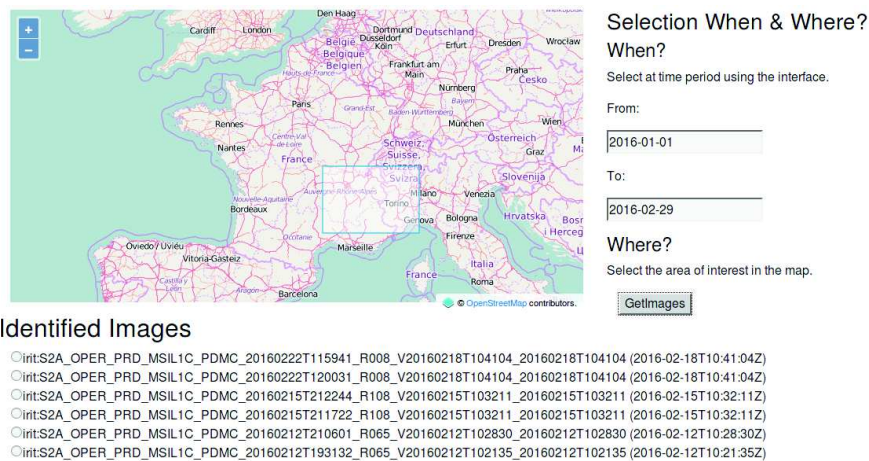


FIGURE 5. Interface de l'application de navigation

une période d'intérêt d'une semaine après la prise d'image (24x7=168 heures, 'PT168H') ; cette période est utilisée pour filtrer les données pertinentes (Cf. les trois dernières clauses FILTER). Faute de place, nous n'avons pas indiqué les éléments filtrant les images présentes dans la zone spécifiée par l'utilisateur⁸.

```
select ?img_md ?img_identifieur ?obs_time ?humidityDataVal ?humidityUnits
where{
  ?img_md a sparkindata:SatelliteImgMetadataRecord.
  ?img_md dct:identifieur ?img_identifieur.
  ?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.
  ?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))
}
```

8. Une vidéo montre le fonctionnement du prototype ici : <http://geo-space.info/demos/metadataBrowserDataMeteoFrance.mp4>

```
FILTER (?diffDateTime<'PT168H'^^xsd:dayTimeDuration)
FILTER (?diffDateTime>'PT0H'^^xsd:dayTimeDuration)}
```

La figure 6 montre l'information météorologique associée à une image faisant partie de la liste retournée par la requête. L'empreinte de l'image est représentée par un polygone orange. Chaque cercle bleu correspond à une station météorologique. En bas de la carte figurent des séries temporelles qui représentent l'évolution de la variable *Temperature* pendant la période d'intérêt choisie.

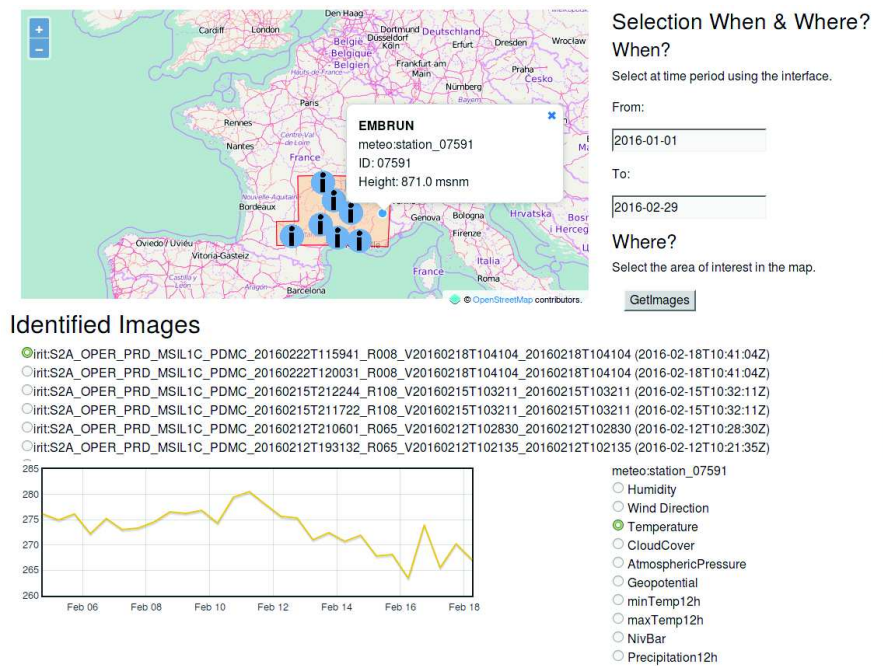


FIGURE 6. Affichage des informations météorologiques associées à une image satellite en réponse à la requête.

6. Conclusions et perspectives

Nous avons présenté une approche pour intégrer sémantiquement des données géolocalisées et datées ainsi que des méta-données d'images d'observations de la terre. Cette approche repose sur un modèle de connaissances générique qui réutilise des vocabulaires existants (SSN, DCAT, GeoSPARQL). Ce modèle décrit les données externes et les méta-données comme des entités ayant deux composantes, spatiale et temporelle. Nous avons montré l'intérêt de ce modèle

à l'aide d'un prototype qui permet de rechercher, via une interface simple, des métadonnées d'images satellites et des données météorologiques associées.

Pour aller plus loin, il serait intéressant d'identifier des patrons qui augmenteraient la pertinence de l'image, en s'appuyant sur des règles et du raisonnement. On pourrait par exemple analyser les séries temporelles de température et détecter une canicule. Pour le moment, la spécialisation du modèle générique, à savoir l'alignement entre métadonnées des images et propriétés /classes du vocabulaire, est faite à la main. Dans le futur, nous envisageons d'intégrer des alignements automatiques pour mettre en correspondance les métadonnées d'images et les vocabulaires, mais également les données annotées et d'autres sources du LOD. Nous prévoyons aussi de considérer plusieurs sources de données d'un domaine (e.g. l'agriculture et les productions agricoles) et de fournir des règles et des outils de raisonnement pour assister les analyses.

Bibliographie

- Arenas H., AussenacGilles N., Comparot C., Trojahn C. (2016). Visualisation d'événements enrichis par des images satellites et des données ouvertes : application aux données sismiques. In *Sageo 2016 - démonstrations - nice (f)*, 6 déc. 2016.
- Atemezing G., Corcho O., Garijo D., Mora J., Poveda-Villalón M., Rozas P. *et al.* (2013). Transforming meteorological data into linked data. *Semantic Web*, vol. 4, n° 3, p. 285-290.
- Battle R., Kolas D. (2012). Enabling the Geospatial Semantic Web with Parliament and GeoSPARQL. *Semantic Web*, vol. 3, n° October 2012, p. 355-370.
- Calbimonte J.-p., Jeung H., Corcho O., Aberer K. (2012). Enabling Query Technologies for the Semantic Sensor Web. *International Journal on Semantic Web and Information*, vol. 8, n° 1, p. 43-63.
- Compton M., Barnaghi P., Bermudez L., García-Castro R., Corcho O., Cox S. *et al.* (2012). The SSN ontology of the W3C semantic sensor network incubator group. *Journal of Web Semantics*, vol. 17, p. 25-32. Consulté sur <http://dx.doi.org/10.1016/j.websem.2012.05.003>
- EARSC E. (2016). *Creating a European Marketplace for Earth Observation Services*. Rapport technique. eoSERVICESindustry EARSC.
- ESA. (2016a). *Sentinel 1*. <https://sentinel.esa.int/web/sentinel/missions/sentinel-1>. (Accessed:2016-07)
- ESA. (2016b). *Sentinel 2*. <https://sentinel.esa.int/web/sentinel/missions/sentinel-2/>. (Accessed:2016-07)
- Espinoza-Molina D., Datcu M. (2013). Earth-observation image retrieval based on content, semantics, and metadata. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, n° 11, p. 5145-5159.
- Espinoza-Molina D., Nikolaou C., Dumitru C. O., Bereta K., Koubarakis M., Schwarz G. *et al.* (2015). Very-High-Resolution SAR Images and Linked Open Data

- Analytics Based on Ontologies. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, n° 4, p. 1696 – 1708.
- Gasperi J. (2014). Semantic Search Within Earth Observation Products Database Based on Automatic Tagging of Image Content. In *Proceedings of the conference on big data from space*, p. 4–6.
- Gatti A., Bertolini A. (2013). *Sentinel-2 Products Specification Document*. Rapport technique. European Space Agency. Consulté sur https://sentinel.esa.int/documents/247904/349490/S2_MSI_Product_Specification.pdf/ed1970b9-94ca-421f-9edc-de08af2a98da?version=1.1
- Keßler C., Farmer C. J. Q. (2015). Querying and integrating spatial-temporal information on the Web of Data via time geography. *Journal of Web Semantics*, vol. 35, p. 25–34. Consulté sur <http://dx.doi.org/10.1016/j.websem.2015.09.005>
- Kolas D., Perry M., Herring J. (2013). *Getting started with GeoSPARQL*. Rapport technique. OGC. Consulté sur http://www.ssec.wisc.edu/meetings/geosp_sem/presentations/GeoSPARQL_Getting_Started-KolasWorkshopVersion.pdf
- Maali F., Erickson J. (2014). *Data Catalog Vocabulary (DCAT)*.
- Miller H. J. (2010, feb). The Data Avalanche Is Here. Shouldn't We Be Digging? *Journal of Regional Science*, vol. 50, n° 1, p. 181–201. Consulté sur <http://doi.wiley.com/10.1111/j.1467-9787.2009.00641.x>
- Perry M., Herring J. (2012). *OGC GeoSPARQL-A geographic query language for RDF data*. Rapport technique. Open Geospatial Consortium. Consulté sur <http://www.opengis.net/doc/IS/geosparql/1.0>
- Randell D. A., Cui Z., Cohn A. G. (1992). A spatial logic based on regions and connection. In *Proceedings 3rd international conference on knowledge representation and reasoning*.
- Reitsma F., Albrecht J. (2005). Modeling with the semantic web in the geosciences. *IEEE Intelligent Systems*, vol. 20, n° 2, p. 86–88.
- Sentinel-1 Team. (2013). *ESA Sentinel 1 handbook*. Rapport technique, vol. 1. European Space A.
- TELEIOS. (2016). *Virtual Observatory Infrastructure for Earth Observation Data*. http://www.earthobservatory.eu/Ontologies/_and/_Linked/_Geospatial/_Data. (Accessed: 2016-12-01)
- The TELEIOS Team. (2012). Building Remote Sensing Applications Using Scientific Database And Semantic Web Technologies. In *Image information mining conference: Knowledge discovery from earth observation data*, p. 2–5. Consulté sur <http://www.earthobservatory.eu/publications/iim1.pdf>
- Zaslavsky A., Perera C., Georgakopoulos D. (2013). Sensing as a Service and Big Data. In *Proceedings of the international conference on advances in cloud computing acc*. Bangalore.