



HAL
open science

Un modèle pour l'intégration spatiale et temporelle de données géolocalisées

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

► **To cite this version:**

Helbert Arenas, Cassia Trojahn dos Santos, Catherine Comparot, Nathalie Aussenac-Gilles. Un modèle pour l'intégration spatiale et temporelle de données géolocalisées. *Revue Internationale de Géomatique*, 2018, 28 (2), pp.243-266. 10.3166/rig.2018.00055 . hal-02398679

HAL Id: hal-02398679

<https://hal.science/hal-02398679v1>

Submitted on 8 Jan 2020

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.

Un modèle pour l'intégration spatiale et temporelle de données géolocalisées

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

► **To cite this version:**

Helbert Arenas, Cassia Trojahn, Catherine Comparot, Nathalie Aussenac-Gilles. Un modèle pour l'intégration spatiale et temporelle de données géolocalisées. *Revue Internationale de Géomatique*, Lavoisier, 2018, 28 (2), pp.243-266. 10.3166/RIG.00.1-24 . hal-02398679

HAL Id: hal-02398679

<https://hal.archives-ouvertes.fr/hal-02398679>

Submitted on 8 Jan 2020

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.

Un modèle pour l'intégration spatiale et temporelle de données géolocalisées

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

IRIT, CNRS et Université 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'Agence Spatiale Européenne a récemment lancé les satellites des séries Sentinel qui livrent 8 à 10 To de données par jour, ce qui ouvre de nouvelles opportunités dans les domaines de l'environnement, l'urbanisme, l'océanographie, la climatologie, etc. Les applications métier nécessitent de coupler ces images à des données issues de diverses sources, afin de fournir une meilleure aide à la décision pour des actions à effectuer sur les zones observées. Un des défis à relever est alors l'intégration de ces données hétérogènes. Les technologies du Web Sémantique fournissent une infrastructure basée sur des standards de représentation de données et des ontologies, qui jouent un rôle clé dans ce contexte. Notre approche sémantique d'intégration de données enrichit les métadonnées des images satellites par des données de sources pertinentes, décrivant les observations de la Terre en réponse à un besoin donné. Nous proposons un vocabulaire sémantique, formalisant des relations spatio-temporelles et support à l'intégration de divers jeux de données géolocalisées. Nous illustrons notre approche sur un cas d'étude basé sur des données météorologiques.

ABSTRACT. Earth Observation is a rapidly evolving domain. The European Space Agency recently launched the satellites of the Sentinel series, which will deliver between 8 and 10 TB of data per day, opening emerging opportunities in domains ranging from agriculture and environmental monitoring to urban planning, oceanography and climate studies. Domain-oriented applications require that images be enriched with data coming from various sources, that allow to better support decision-making processes related to the observed areas. One of the challenges to be addressed then concerns the integration of heterogeneous data. In this context, semantic technologies play a key role by providing an infrastructure based on standard models for representing data and on ontologies. We present a semantic approach to integrate data with the aim of enriching metadata of satellite imagery with data from various sources that are relevant to describe Earth observations for a particular need. We propose a semantic vocabulary, a formalization of spatio-temporal relationships as well as an integration process of various geo-spatial data sets. We illustrate our approach through a case study exploiting meteorological data.

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 types de satellites, Sentinel-1 et Sentinel-2, qui diffuseront en continu des images de la Terre captées selon différentes technologies et accessibles à large échelle. Ces satellites sont censés transmettre quotidiennement des images de haute qualité, soit entre 8 à 10 To de données. La disponibilité de ces 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. Ces applications métier ont cependant un fort besoin de coupler ces images à des données sur les zones observées, provenant de diverses sources (données sur le sol, sur le climat, etc.). À titre d'exemple, indiquer qu'une image contient une région touchée par un phénomène naturel tel qu'un tremblement de Terre ou une canicule s'avère pertinent pour décider des actions à effectuer sur cette zone ou pour réaliser des analyses à plus long terme. Par ailleurs, en exploitant les caractéristiques spatio-temporelles du phénomène sur une image (son empreinte spatiale et sa date), il devient possible de savoir qu'une entité géolocalisée au sein de cette empreinte (une ville par exemple), a subi le même phénomène. Dans ce cadre, parce que les images sont déjà décrites par des métadonnées satellitaires, un des défis à relever concerne l'intégration de données hétérogènes venant de sources diverses.

Dans ces scénarios, 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 permet de gérer la mise en correspondance temporelle, géographique et sémantique de données et d'images, ce qui explique l'attention particulière qu'elle a reçue ces dernières années. Selon (Reitsma, Albrecht, 2005), grâce à l'utilisation de la sémantique pour modéliser des données d'observation de la Terre, de nouvelles perspectives s'ouvrent pour étudier plus précisément les systèmes environnementaux et mieux comprendre les méthodes utilisées pour représenter ces systèmes. En outre, la découverte et la recherche sémantiques d'informations géographiques valorisent les données temporelles et spatiales en tant que supports au raisonnement et à la découverte automatisée de connaissances.

Cet article présente un processus d'intégration sémantique de données géolocalisées, dont des données d'observations terrestres. Nous proposons d'utiliser des ontologies pour définir formellement les relations entre les ensembles de métadonnées d'une image et les données provenant de sources externes (données de capteurs météo ou sismiques, nature des sols, etc). Pour cela, nous avons défini un vocabulaire en deux parties. La première partie formalise les concepts et les relations spatio-temporelles nécessaires à la description des métadonnées et des propriétés spatiales et temporelle. C'est sur cette représentation que s'appuie le processus d'intégration de ces différentes données. La deuxième partie permet de décrire des données issues de capteurs avec leur composante spatiale et temporelle, comme un type particulier de données à intégrer. La principale contribution de ce travail est un modèle générique de connaissances qui permet la description sémantique et homogène de données externes et de méta-

données d'images satellites comme des entités ayant deux composantes, spatiale et temporelle. Ce vocabulaire repose sur des vocabulaires existants (SSN, DCAT, GeoSPARQL). Ainsi, notre travail contribue à la sémantique géospatiale (Janowicz *et al.*, 2012) permettant l'interopérabilité des sources de données.

Nous illustrons notre approche à travers un cas d'étude exploitant des données météorologiques. Pour cette étude, la représentation sémantique des métadonnées associées aux images brutes fournies par le CNES (Centre National d'Etudes Spatiales) et des données contextuelles (météorologiques, sanitaires, etc.) est liée aux informations météorologiques relatives à la zone et à la date de la prise de vue. Ce travail est mené dans le cadre du projet SparkInData¹ visant à construire une plate-forme cloud de données d'observations de la Terre.

Après l'exposé de travaux connexes (partie 2), nous décrivons, dans la partie 3, les vocabulaires utilisés ainsi que le vocabulaire formel que nous proposons pour l'intégration de métadonnées d'images et de données, dans lequel les relations spatiales et temporelles jouent un rôle central. Dans la partie 4, nous montrons, à travers le cas d'usage, comment ce vocabulaire permet de produire une représentation sémantique des métadonnées d'images et de données externes tirées de capteurs. Nous expliquons, en partie 5, comment ces métadonnées sémantiques permettent de mieux répondre à des besoins d'utilisateurs.

2. État de l'art: modèles sémantiques pour les métadonnées d'images satellites

L'intégration de données spatiales a été très tôt au coeur du projet du web sémantique, comme le formule Egenhofer dans sa vision d'un web sémantique géospatial *Semantic GeoSpatial Web* (Egenhofer, 2002). Les ontologies sont identifiées comme des outils adaptés pour modéliser les données géospatiales (Janowicz *et al.*, 2012), et les standards du web sémantique comme des technologies facilitant l'interopérabilité. En effet, les technologies du web sémantique offrent une approche rationnelle non seulement pour modéliser les données associées à des images d'observation de la Terre, mais aussi pour lier et intégrer ces métadonnées à des données de nature différente, à condition qu'elles aient une dimension spatiale et une dimension temporelle : ces deux dimensions sont nécessaires pour associer les bonnes données à la bonne image. Les données doivent avoir été relevées dans la zone couverte par l'image à une date égale ou proche de celle de la prise de vue. Nous mentionnons ici les vocabulaires mis au point dans l'état de l'art pour réaliser cette modélisation.

Avant même l'utilisation de formats du web sémantique, plusieurs standards ont été mis au point pour représenter des métadonnées d'images satellites (Mitchell *et al.*, 2014), notamment le répertoire commun des métadonnées de la NASA, le CMR (Common Metadata Repository), les métadonnées utilisées pour décrire les missions hétérogènes de l'ESA (ESA Heterogeneous Missions Access project) (Voges, 2014),

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

ou encore le profil proposé par l'OGC, appelé Earth Observation Metadata Profile of Observations and Measurements (EOMOM) (Gasperi *et al.*, 2016). Certains travaux utilisent des vocabulaires non-spécifiques, comme dans (Desconnets *et al.*, 2014), où un profil d'application pour les métadonnées des images d'observation de la Terre est basé sur Dublin Core et le framework *Singapore Dublin Core*. D'autres approches font appel à des vocabulaires définis précisément pour les métadonnées géolocalisées, comme GeoDCAT (Voges *et al.*, 2016), qui spécialise le vocabulaire de métadonnées DCAT (Data Catalog Vocabulary), un standard du W3C conçu pour permettre l'accès à des catalogues de métadonnées (Maali, Erickson, 2014).

TELIOS (Virtual Observatory Infrastructure for Earth Observation Data)² est un des projets les plus aboutis visant à exploiter la sémantique pour les images de satellites, dans l'objectif de faciliter l'accès à des données d'observation de la Terre par des agences telles que la NASA ou l'ESA. Dans le cadre de ce projet, les images sont classifiées et représentées sous forme de vecteurs liés à des sources de données externes par le biais d'ontologies (TELEIOS, 2016). Ces ontologies s'appuient notamment sur le standard stRDF qui étend RDF pour les données spatiales et temporelles, bien que les analyses réalisées ne prennent pas en compte la composante temporelle. Un des objectifs de TELIOS concerne les images produites par les radars à synthèse d'ouverture (SAR) (Espinoza-Molina, Dacu, 2013) et (Espinoza-Molina *et al.*, 2015) et a pour but de leur associer des données provenant de différentes sources de données, comme Corine Land Cover (données sur la nature de la couverture du sol), Urban Atlas, Geonames et LinkedGeoData (taxonomies de toponymes). La représentation sémantique de ces données et métadonnées vise à faciliter leur recherche au sein du système Earth Observation Image Retrieval. Pour cela, le projet propose un modèle ontologique, SAR, qui réutilise la taxonomie TerraSARX et étend celle présentée dans (TELEIOS, 2016). Ce modèle est alimenté par un processus d'extraction automatique d'informations et de typage sémantique qui exploite les métadonnées des images d'une part, et les données des sources choisies d'autre part. Les données des sources externes sont alors liées par leurs propriétés spatiales aux métadonnées des images situées sur une zone géographique incluant celle des données. L'entrepôt de données sémantiques ainsi constitué est ensuite exploité pour rechercher des images ou des informations en formulant des requêtes complexes combinant les métadonnées d'origine et celles tirées des données. Cette approche rejoint notre objectif puisqu'elle associe aux images d'observation de la Terre des données provenant du web. Cependant, la composante temporelle n'y est pas présente.

Plus récemment, plusieurs projets se sont intéressés à la mise à disposition de données géolocalisées et sémantisées au sein du LOD (Linked Open Data) sur le web, avec par exemple la mise sous forme de données liées des données de OpenStreetMap dans le projet LinkedGeoData (Stadler *et al.*, 2012). L'avantage principal de publier des jeux de données au sein du LOD est de fournir un accès plus large à ces données, et de les rendre ainsi disponibles pour de nouvelles applications (Blower *et al.*, 2016).

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

D'autres projets, à l'inverse, utilisent des données liées existant dans le LOD. Par exemple, dans le cadre du projet MELODIES, la combinaison de sources de données hétérogènes dans le domaine de l'observation de la Terre s'appuie sur des vocabulaires tirés du Web des données liées (Blower *et al.*, 2014). De même, le projet CHARMe a eu pour objectif d'ajouter des principes et données du LOD et de l'Open Data pour enrichir les données Copernicus. Dans un premier temps, la caractérisation des métadonnées a eu pour but de décrire des services climatiques de haute qualité (Lankester *et al.*, 2014), la qualité des données étant un des enjeux centraux. Un deuxième objectif était d'utiliser ces données pour des analyses climatiques et pour produire des prévisions météorologiques (Clifford *et al.*, 2014).

Finalement, de grandes quantités de données d'observation de la Terre sont disponibles, mais peu de jeux de données sont présents sous forme de données liées dans le LOD. Le projet Linked Open Earth Observation Data for Precision Farming (LEO) cherche à étendre le travail sur le LOD à partir de TELEIOS. Il vise aussi le développement d'une méthodologie et d'outils logiciels pour intégrer au LOD des jeux de données d'observation de la Terre (Koubarakis, 2014), (Kyzirakos, Manegold *et al.*, 2014) (Kyzirakos, Vlachopoulos *et al.*, 2014) (Kyzirakos *et al.*, 2015).

Nous avons identifié une approche qui vise la description sémantique de données de capteurs produites au fil du temps (Atemezing *et al.*, 2013). Ses auteurs réutilisent et adaptent l'ontologie SSN³ (Semantic Sensor Network), couplée à l'ontologie OWL-Time⁴, pour publier des informations météorologiques de l'office météorologique espagnol (AEMET)⁵. Bien que notre approche soit proche de cette dernière proposition (Atemezing *et al.*, 2013), notre contribution va au delà de ces travaux en exploitant à la fois deux composantes, temporelle et spatiale, pour la mise en correspondance des métadonnées d'images et de données externes représentées selon un format sémantique (sous forme de données liées). Nous définissons un nouveau vocabulaire qui réutilise des vocabulaires existants, et une approche associée qui gère les deux types de relations, spatiales et temporelles.

3. Un vocabulaire pour intégrer les données spatio-temporelles

Notre objectif est d'utiliser des métadonnées d'images d'observation de la Terre, conjointement à des mesures fournies par des 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 type et leur mise en relation à l'aide d'un vocabulaire unique. Comme le recours à des standards améliore l'interopérabilité, nous proposons un vocabulaire qui réutilise trois ontologies reconnues et maintenues : DCAT (Data Catalog Vocabulary), un standard du W3C conçu pour permettre l'accès à des catalogues de métadonnées (Maali,

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

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

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

Erickson, 2014) ; GeoSPARQL, un standard de l’OGC conçu comme un vocabulaire minimal pour interroger des données géolocalisées (Kolas *et al.*, 2013) ; et enfin SSN (Semantic Sensor Network), un autre standard du W3C conçu pour décrire les capteurs et observations issues de ces capteurs (Compton *et al.*, 2012). Avant de présenter le vocabulaire que nous proposons (Section 3.2), nous introduisons les trois vocabulaires sur lesquels notre vocabulaire repose (Section 3.1).

3.1. Les vocabulaires réutilisés

3.1.1. Data Catalog Vocabulary (DCAT), un vocabulaire pour décrire des métadonnées

DCAT⁶ est un vocabulaire recommandé par le W3C pour publier des métadonnées de catalogues sur le Web (Maali, Erickson, 2014). DCAT combine des éléments d’autres vocabulaires tels que Dublin Core Terms⁷ (dct) (un vocabulaire de métadonnées documentaires), OWL-Time⁸ (une ontologie minimale pour représenter en OWL des dates et des durées), SKOS⁹ (un vocabulaire simple pour représenter des taxonomies de termes) et VCard (un format standard ouvert pour décrire des données personnelles). La figure 1 montre les principales classes et propriétés de DCAT que nous avons utilisées. La classe *dcat:Dataset* représente un enregistrement de métadonnées décrit à l’aide des nombreuses propriétés de cette classe, telles que *dct:title*, *dct:description*, *dct:issued*, *dct:language*, *dct:identifier* (pas présentes sur la figure 1). Nous avons également exploité les propriétés *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-Time.

Lorsque l’on s’intéresse à des métadonnées géolocalisées, il existe des alternatives à DCAT comme GeoDCAT et EOMPOM. GeoDCAT (Perego *et al.*, 2017) est un vocabulaire permettant de compléter la description d’une couverture spatiale en référence à un lieu (*dct:Location*), par une représentation géométrique sous forme de polygone (e.g. *BoundingPolygon*) sur lequel on peut appliquer des fonctions topologiques. Ce modèle a été développé par le *EU Programme Interoperability*, qui a publié la dernière version (GeoDCAT 1.0.1) en août 2016. Cependant, GeoDCAT a été plutôt conçu comme un mécanisme de *mapping* entre ISO 19115 (Geographic Information Metadata, une norme ISO pour la description des données géolocalisées) et DCAT, et il est difficile de l’adapter à d’autres formes d’enregistrements de métadonnées. Le deuxième vocabulaire, EOMPOM (Earth Observation Metadata Profile of Observa-

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

7. https://terms.tdwg.org/wiki/Dublin_Core

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

9. <https://www.w3.org/2004/02/skos/>

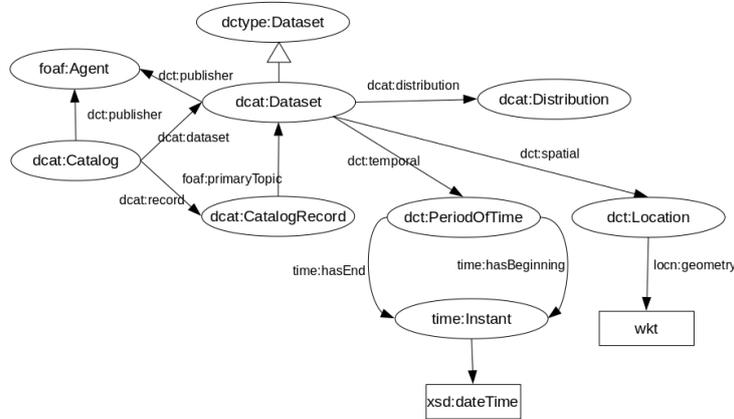


FIGURE 1. Principales classes de DCAT.

tions and Measurements), a été produit par l’OGC comme un profil de métadonnées pour les observations de la Terre, en réponse aux besoins de standardisation d’industriels du domaine spatial (Gasperi *et al.*, 2016). L’objectif de ce profil est de définir un standard pour décrire et cataloguer les produits de capteurs de satellites d’observation de la Terre. Dans ce profil, les produits sont groupés en collections. Les produits d’une collection se différencient les uns des autres par la date d’acquisition, la localisation, et les caractéristiques du capteur. EOMPOM est disponible sous forme d’un ensemble de schémas XML (XML Schema), mais pas encore sous forme d’un vocabulaire sémantique à ce jour. Par conséquent, son utilisation comme vocabulaire pour la construction d’ontologies n’est pas simple et sera considérée dans nos travaux ultérieurs. Le tableau 1 fournit une comparaison entre les trois vocabulaires. Nous avons choisi de retenir DCAT car il présente les avantages d’être disponible dans un format du web sémantique, d’être reconnu comme un standard et de permettre de gérer à la fois les dimensions spatiale et temporelle des données. Nous avons décidé de l’étendre pour représenter plus précisément des métadonnées des images de télédétection.

Tableau 1. Comparaison entre les vocabulaires de métadonnées.

	Dcat	GeoDcat	Earth Observation Metadata Profile
Focus	Générique	Générique spatial	Spécifique
Standard	oui (W3C)	no (draft EU)	oui (OGC)
Mise en œuvre	RDFS	RDF	XML
Spatial	dct:Location	GeoSPARQL	gml:Polygon
Temporal	dct:temporal	dct:temporal	gml:TimeInstant

3.1.2. GeoSPARQL

Les ontologies ont été utilisées à des fins diverses pour l'information géo-spatiale. Une norme bien connue dans ce domaine est GeoSPARQL (Kolas *et al.*, 2013) (Battle, Kolas, 2012). Définie par l'OGC comme une norme, cette ontologie représente des propriétés, des relations et des fonctions spatiales. La Figure 2 présente les principales classes de GeoSPARQL. La classe *geo:Feature* représente toute entité du monde réel ayant une empreinte spatiale (Perry, Herring, 2012). Cette empreinte 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. En revanche, GeoSPARQL ne traite pas la dimension temporelle des données. Pour traiter cette dimension, nous avons utilisé conjointement OWL-Time, qui permet d'associer une propriété temporelle à une classe de données.

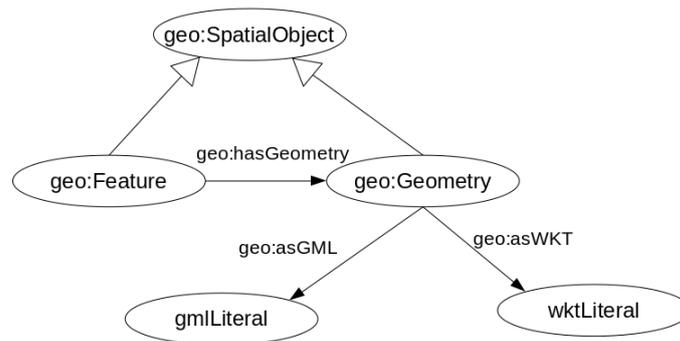


FIGURE 2. Principales classes de GeoSPARQL.

Parmi les alternatives à GeoSPARQL, la plus complète est le vocabulaire de requête stSPARQL (Koubarakis, Kyzirakos, 2010) (Bereta *et al.*, 2013). stSPARQL est une extension de SPARQL conçue pour interroger des graphes RDF spatio-temporels. stSPARQL gère une composante spatiale comme GeoSPARQL mais permet en plus d'interroger des quadruplets au format stRDF, i.e. des triplets RDF auxquels est associée une variable temporelle, de type *xsd:dateTime* (un instant) ou *strdf:period* (un intervalle de temps), indiquant le temps de validité du triplet ("valid time"). stSPARQL propose des fonctions temporelles (relations temporelles d'Allen, relations topologiques, agrégats temporels) permettant par exemple de rechercher un fait qui s'est produit entre deux dates.

Le tableau 2 fournit une comparaison de GeoSPARQL et stSPARQL qui ont été développés plus en moins en même temps. Dans les deux cas, la représentation spatiale utilise des littéraux GML (Geographic Markup Language) et WKT (Well Known

Text). Nous y avons ajouté OWL-Time¹⁰ pour montrer sa complémentarité afin de traiter la dimension temporelle des données. Nous avons retenu GeoSPARQL car il offre des possibilités pour raisonner sur des géométries, à savoir calculer des relations entre des entités sur la base des relations entre leur représentation spatiale (inclusion, recouvrement, ...), en association avec OWL-Time qui nous suffit pour le moment pour représenter la composante temporelle des données.

Tableau 2. Comparaison des vocabulaires geoSPARQL, stSPARQL et OWL-Time

	GeoSPARQL	stSPARQL	OWL-Time
Représentation spatiale	oui	oui	non
Représentation temporelle	non	oui	oui
Mécanisme de réécriture des requêtes	oui	non	non
Agrégation spatiale	non	oui	non
Évaluation des relations topologiques	oui	non	non
Vocabulaire standardisé	oui (OGC)	non	oui (W3C)
Pris en charge en base de données	Apache Marmotta, Parliament, Strabon, OpenSahara, Oracle, Stardog, Virtuoso	Strabon	

3.1.3. Semantic Sensor Network (SSN)

Le troisième vocabulaire réutilisé est SSN (Compton *et al.*, 2012), qui décrit des capteurs numériques en réseau, ainsi que les fonctionnalités, propriétés et mesures de ces capteurs. En effet, il y a actuellement un grand nombre de capteurs déployés qui récupèrent sur ou à proximité de la Terre de l'information de domaines divers tels que l'environnement, la sécurité, la météo, etc. Ces capteurs sont des sources indéniables pour associer un contexte à des images, ou portions d'images. L'information fournie par les capteurs est hétérogène, tant par la nature des données que par leur format de stockage, et soulève des difficultés d'intégration à d'autres sources de données. L'ontologie SSN est alignée avec une ontologie formelle DOLCE-UltraLite (DUL) qui définit de façon formelle des classes de plus haut niveau, et cela pour limiter les interprétations possibles. SSN contient 41 concepts et 39 propriétés d'objets (*Object-Property*), et permet de décrire des capteurs à n'importe quel niveau de détail. Elle est générique au sens où elle ne définit pas toutes les propriétés spécifiques à des capteurs et des observations particuliers. En revanche, il est possible de la lier à des ontologies ou des vocabulaires propres à un domaine tels que l'ontologie SWEET¹¹ qui décrit des données environnementales (Raskin, 2006).

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

11. <http://sweet.jpl.nasa.gov/ontology/>

Une alternative intéressante à SSN est sa version adaptée à la représentation de données météorologiques, AEMET¹², développée autour de 2010 pour représenter les données de l'office météorologique espagnol (Atemezing *et al.*, 2013). Cependant, ce vocabulaire n'est pas très documenté, et n'a pas été maintenu, ce qui justifie que nous ne l'ayons pas retenu. La figure 3 montre les principales classes qui constituent cette ontologie, qui réutilise les vocabulaires SSN, DUL et wgs, ainsi que OWL-Time, et qui est spécialisée pour les données de capteurs météorologiques. Selon cette représentation, une station météorologique est une plate-forme regroupant des capteurs dont les mesures relevées sont assimilées à des observations, au sens de SSN. Chaque plate-forme a une géolocalisation qui permet de géolocaliser aussi les données restituées par les capteurs.

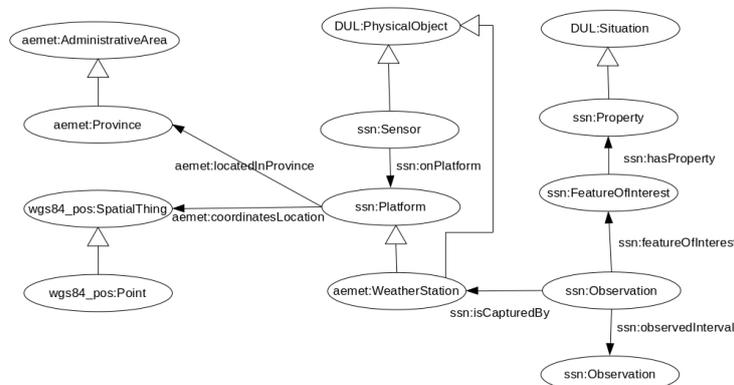


FIGURE 3. Principales classes du réseau d'ontologies AEMET.

3.2. Vocabulaire proposé

La principale contribution de cet article est la définition d'un vocabulaire sémantique (il ne possède pas encore toutes les propriétés d'une ontologie) facilitant l'intégration de données spatio-temporelles, et en particulier de données issues de capteurs. Nous le présentons en deux parties. Tout d'abord, un modèle générique fournit un noyau focalisé sur les métadonnées d'images satellite et tout autre type de données ayant une composante spatiale et une dimension temporelle. Cette partie du modèle peut être étendue pour répondre à des besoins spécifiques, essentiellement en reliant les métadonnées à des données provenant d'une autre source et possédant ces deux dimensions. La figure 4 présente cette première partie du modèle proposé. Une deuxième partie du modèle sert à décrire tout type de données de capteur en lui associant aussi des propriétés spatiales et temporelles. Cette partie du modèle se lie donc

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

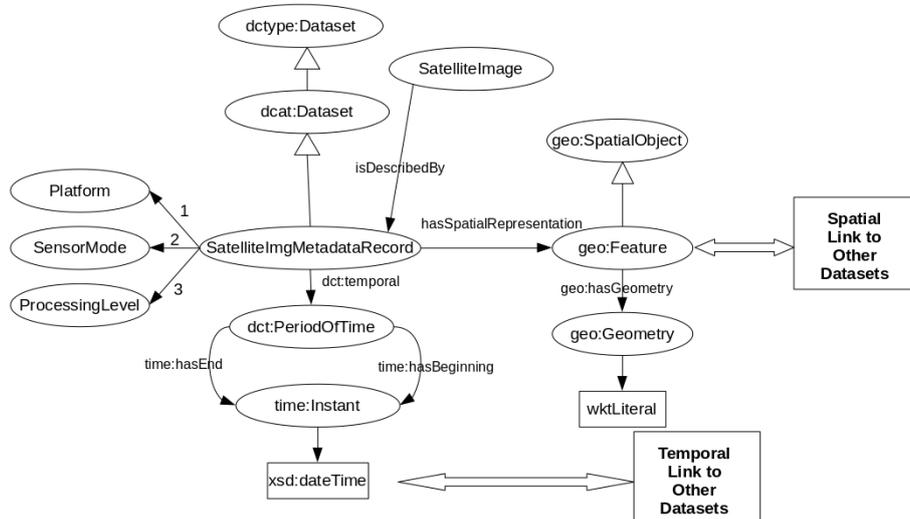


FIGURE 4. Principales classes du modèle de données dédiées aux métadonnées.

facilement à la première. La version présentée sur la figure 5, a été spécialisée pour décrire des données météorologiques.

Nous détaillons dans les quatre parties suivantes comment sont représentées les métadonnées d’images, les aspects spatiaux et les aspects temporels des données, ainsi que les données de capteurs.

3.2.1. Représentation des métadonnées d’images

Comme décrit ci-dessus, nous avons utilisé et adapté DCAT pour la représentation riche des métadonnées des images de télédétection. Comme présenté dans la figure 4, la classe *SatelliteImgMetadataRecord* est une sous-classe de *dcat:Dataset*. Grâce à cette spécialisation, les métadonnées spécifiques à l’imagerie satellitaire sont représentées comme des métadonnées de catalogue. Nous avons associé trois propriétés à *SatelliteImgMetadataRecord* :

- *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. Plus précisément, il y a différents niveaux de traitement propres à chaque type de satellite Sentinel-1 (*LEVEL0*, *LEVEL1*, *LEVEL2*) ou Sentinel-2 (*LE-*

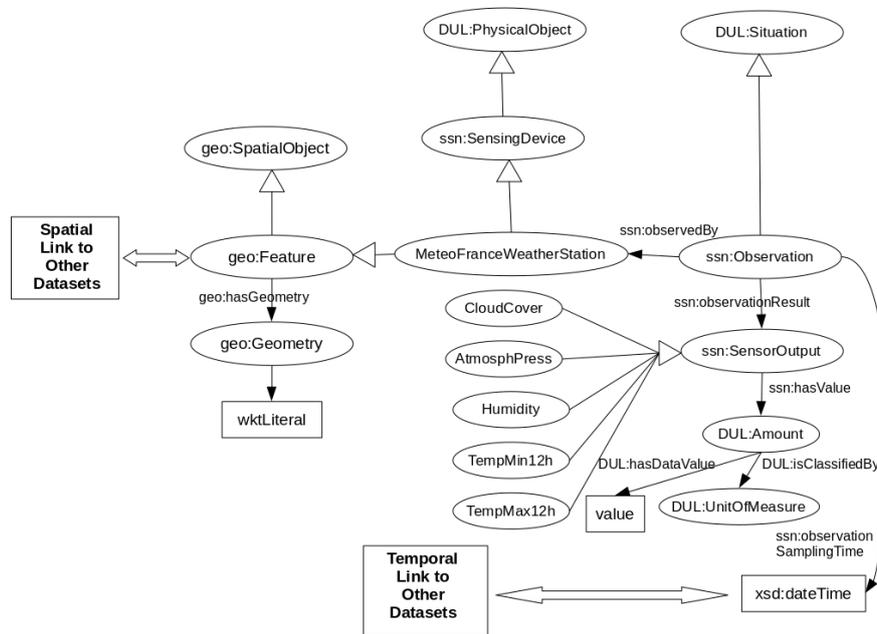


FIGURE 5. Principales classes du modèle de données dédiées aux informations météorologiques.

VELIC, 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 (European Space Agency, 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 (European Space Agency, 2016b) (Gatti, Bertolini, 2015).

– *hasSensorMode* (3 en Fig.4) : cette propriété indique le mode opératoire du satellite pour cette image. Sentinel-1 et Sentinel-2 ont des modes opératoires différents en fonction du type de surface terrestre à observer. Ainsi, Sentinel-1 utilise le mode "senseur" WV (Wave Mode) quand il opère au-dessus des océans, le mode IW (Interferometric Wide-Swath Mode) pour la Terre ou encore EW (Extra-Wide-Swath Mode) pour les zones polaires (European Space Agency, 2016a) (Sentinel-1 Team, 2013).

En plus de ces propriétés alphanumériques décrivant l'image, un enregistrement de métadonnées, tout comme les données venant d'autres sources, possède une dimension spatiale et une dimension temporelle que nous décrivons dans les deux parties sui-

vantes. Grâce à ces propriétés, il est possible d'établir des relations spatio-temporelles entre les métadonnées et des données externes. Ceci permet, comme nous le montrons dans ce qui suit, de déterminer quelles observations peuvent être associées aux lieux couverts par l'image durant un temps donné.

3.2.2. Représentation de la dimension spatiale

Notre modèle sémantique 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*. Comme introduit ci-dessus, 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. Une fois les données et métadonnées ainsi représentées, GeoSPARQL peut calculer des relations topologiques entre l'empreinte d'une image satellite (le *footprint*), et la composante spatiale d'autres données.

Notre premier cas d'usage d'évaluation de ce modèle exploitait comme source externe de données, le site de *Météo France*, d'où la classe *MeteoFranceWeatherStation* introduite dans le modèle comme spécialisation 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.

3.2.3. Représentation de la dimension temporelle

La propriété temporelle d'un enregistrement de métadonnées identifie le moment où l'image a été prise. Les données externes, en particulier les données de capteurs comme les informations météo relevées périodiquement par les stations, sont aussi datées. Pour les données de capteurs, nous utilisons la classe *ssn:Observation* (figure 5) pour typer les valeurs mesurées et ainsi leur associer, par la relation *ssn:observation-SamplingTime*, une composante temporelle correspondant à la date de la mesure et de type *xsd:dateTime*.

3.2.4. Représentation de données de capteurs

La partie du modèle adaptée de SSN permet aussi de représenter les mesures météorologiques (température, humidité, vent, etc.), données relevées toutes les trois heures dans chacune des 62 stations françaises de position fixe et fournies par le site de *Météo France*

¹³. Les données météorologiques constituent des informations contextuelles complémentaires des métadonnées associées aux images. Toute autre source fournissant des informations (nature du sol, sismographie, population, etc.) datées et propres à une zone de la surface terrestre, pourrait être ajoutée selon un même procédé. Depuis, nous avons étendu l'expérience en rajoutant des données sismiques (Arenas *et al.*, 2016).

3.3. *Processus d'intégration de données grâce à ce modèle*

Une fois ce vocabulaire sémantique défini, le processus d'intégration des données comporte les trois tâches suivantes, la première concernant les métadonnées et les deux suivantes les données de sources autres :

1. Pour les métadonnées des images de télédétection, le vocabulaire spécialise déjà les méta-données DCAT pour qu'elles soient adaptées aux méta-données des images. Il suffit donc d'instancier ce modèle. Ceci revient à définir des instances des classes `geo:Feature` pour expliciter que ces méta-données ont une dimension spatiale, et de donner une valeur à la classe `geo:Geometry` à partir de la géométrie associée à l'image. De même, il s'agit de représenter les informations temporelles connues pour les métadonnées en utilisant les classes `dct:PeriodOfTime` (de DCAT) et `time:Instant` (de OWL-time). Ces données ainsi représentées en RDF amorcent un entrepôt RDF.

2. Pour chaque source de données à intégrer, il convient d'abord d'adapter le vocabulaire à la description de ces données. Si ce sont des données de capteurs, comme les données météorologiques, océanographiques ou sismologiques par exemple, alors les données et les capteurs qui les récupèrent peuvent être décrits en spécialisant la partie SSN du modèle par des classes spécifiques, comme nous allons le montrer avec les données météorologiques dans la suite. Si les données ne sont pas des données de capteurs (comme des informations tirées de textes ou un relevé cadastral) alors il suffit de construire un modèle décrivant ces données, et d'indiquer que la ou les classes principales sont géolocalisées en les reliant à la classe `geo:Feature` par la relation `has-GeoFeature` de notre modèle.

3. L'étape suivante consiste à décrire l'ensemble des données pertinentes de chaque jeu de données à intégrer selon la spécialisation de l'ontologie définie pour cette source. Ce processus permet de générer des données au format RDF qui peuvent être toutes stockées dans un même entrepôt RDF. Le processus a été mis en oeuvre sur les métadonnées dans la tâche 1. Il doit être réalisé sur les données autres en utilisant les adaptations spécifiques du vocabulaire faites pour chaque source (tâche 2). Au cours de ce processus, à partir de l'information géographique des données sources, on renseigne les coordonnées de l'instance de la classe `geo:Feature` associée au capteur ou de la classe décrivant les données. Toutes les données fournies par un capteur ont les mêmes données géographiques. Pour le temps, la date de mesure des données (fournie par la source de données) est représentée avec des propriétés de OWL-Time.

13. Les variables disponibles sont décrites sur le site de Météo France : https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=90&id_rubrique=32. Fournies en License Open, les données originales peuvent être reproduites, diffusées, modifiées et transformées.

4. Cas d'étude : intégration de métadonnées d'images Sentinel et de données météorologiques

Nous illustrons ici notre approche à l'aide d'un cas d'étude exploitant des données météorologiques. Nous introduisons d'abord le processus d'intégration des données utilisées au sein d'un entrepôt de données, puis nous montrons l'intérêt de la formalisation pour interroger les données de l'entrepôt. Comme sources de données, nous utilisons la base de données PEPS pour les métadonnées des images satellite et les données météorologiques de *Météo France*.

4.1. Le processus d'extraction

Les métadonnées que nous récoltons sont stockées dans une base de données relationnelle gérée par PEPS, qui contient des métadonnées pour environ 1 228 421 images de Sentinel-1, 3 534 69 pour Sentinel-2 et 497 911 images pour Sentinel-3. Nous exploitons des enregistrements de métadonnées d'images Sentinel-2, qui sont obtenus grâce à RESTO, un moteur de recherche d'images d'observation de la Terre qui est déployé sur un serveur du CNES et qui utilise un accès REST (Gasperi, 2014). Les métadonnées sont publiées en JSON. Nous y accédons 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
```

Les métadonnées ainsi sélectionnées sont ensuite représentées à l'aide du modèle sémantique : chaque enregistrement de métadonnées d'une image devient une instance de la classe *SatelliteImageMetadataRecord*. Les triplets RDF ainsi obtenus sont insérés dans un entrepôt de données Stardog. Le code suivant est un exemple d'enregistrement de métadonnées au format RDF sérialisé selon la notation N3¹⁴.

```
sparkindata:img001 a sparkindata:SatelliteImageMetadataRecord.
sparkindata:img001 dct:identifier "img01"^^xsd:string.
sparkindata:img001 dct:title "Example image"^^xsd:String.
sparkindata:img001 dct:description
  "This is the description of the example image."^^xsd:String.
sparkindata:img001 dct:temporal sparkindata:img001_temporal.
sparkindata:img001_temporal a dct:PeriodOfTime.
sparkindata:img001_temporal time:hasBeginning
  sparkindata:img001_temp_beginning.
sparkindata:img001_temp_beginning a time:Instant.
sparkindata:img001_temp_beginning time:inXSDdateTime
  "2015-01-01T00:00:00-05:00"^^xsd:dateTime.
sparkindata:img001_temporal time:hasEnd sparkindata:img001_temp_end.
sparkindata:img001_temp_end a time:Instant.
sparkindata:img001_temp_end time:inXSDdateTime
  "2015-01-01T00:00:02-05:00"^^xsd:dateTime.
```

14. <https://www.w3.org/TeamSubmission/n3/>

```

sparkindata:img001 sparkindata:hasProcessingLevel sparkindata:LEVEL1.
sparkindata:img001 sparkindata:hasPlatform sparkindata:S1A.
sparkindata:img001 sparkindata:hasSensorMode sparkindata:EW.

sparkindata:img001 sparkindata:hasSpatialRepresentation
    sparkindata:img001_feature.
sparkindata:img001_feature a geo:Feature.
sparkindata:img001_feature geo:hasGeometry sparkindata:img001_geo.
sparkindata:img001_geo geo:asWKT "POLYGON (( ....))^geo:wktLiteral".

```

Les données météorologiques de *Météo France*, quant à elles, sont disponibles au format CSV. Une application Java convertit les informations météorologiques des fichiers CSV vers le modèle sémantique et les stocke en RDF dans le triplestore en tant qu'instances de la classe *ssn:Observation*. Le code suivant est un extrait de station météorologique avec une observation associée, transcrit en RDF/N3.

```

sparkindata:ws_01 a sparkindata:MeteoFranceWeatherStation.
sparkindata:ws_01 a geo:Feature.
sparkindata:ws_01 geo:hasGeometry sparkindata:ws_01_geo.
sparkindata:ws_01_geo geo:asWKT "POINT (( ....))^geo:wktLiteral".

sparkindata:obs_001 ssn:observedBy sparkindata:ws_01.
sparkindata:obs_001 a ssn:Observation.
sparkindata:obs_001 ssn:observationSamplingTime
    "2015-01-05T00:15:00-00:00"^^xsd:dateTime.

sparkindata:obs_001 ssn:observationResult
    sparkindata:obs_001_cloudcover.
sparkindata:obs_001_cloudcover a sparkindata:CloudCover.
sparkindata:obs_001_cloudcover ssn:hasValue
    sparkindata:obs_001_cloudcover_val.
sparkindata:obs_001_cloudcover_val a DUL:Amount.
sparkindata:obs_001_cloudcover_val DUL:hasDataValue "30.0"^^xsd:double.
sparkindata:obs_001_cloudcover_val DUL:isClassifiedBy
    sparkindata:Percentage.
sparkindata:Percentage a DUL:UnitOfMeasure.

sparkindata:obs_001 ssn:observationResult
    sparkindata:obs_001_atmPressure.
sparkindata:obs_001_atmPressure a sparkindata:AtmosphPress.
sparkindata:obs_001_atmPressure ssn:hasValue
    sparkindata:obs_001_atmPressure_val.
sparkindata:obs_001_atmPressure_val a DUL:Amount.
sparkindata:obs_001_atmPressure_val DUL:hasDataValue
    "101.325"^^xsd:double.
sparkindata:obs_001_atmPressure_val DUL:isClassifiedBy sparkindata:kPa.
sparkindata:kPa a DUL:UnitOfMeasure.

```

Notons que, au cours des deux processus de représentation sémantique des données (métadonnées ou informations météorologiques), nous assignons aux données une structure et des types sémantiques définis dans notre vocabulaire, de façon à les rendre compatibles entre elles et à assurer des opérations spatiales. Par exemple,

ws_01 est une instance de *MeteoFranceWeatherStation*, mais c'est également une instance de la classe *geo:Feature*. De ce fait, elle peut être liée par la propriété *hasGeometry* à l'élément *ws_01_geo*, qui lui-même est lié par la propriété *geo:asWKT* à un littéral qui contient les coordonnées géographiques réelles de cette entité. Un chemin similaire est développé pour chaque enregistrement de métadonnées. Parce que les deux types d'éléments (données météo et métadonnées d'image) ont maintenant leur représentation spatiale définie dans des termes similaires, il est possible de les comparer en utilisant GeoSPARQL.

4.2. Interrogation de l'entrepôt de données

L'intégration de données permet à l'utilisateur d'avoir un point d'entrée unique pour accéder aux diverses informations (métadonnées d'origine et données externes) associées aux images. L'interface lui permet de dessiner un rectangle pour spécifier des contraintes spatiales (en utilisant les relations spatiales) et de saisir une période de temps pour définir un empan temporel. A partir de ces informations, le système génère une requête GeoSPARQL qui permet de retrouver les données externes (observations météo), les métadonnées et les images dont la composante temporelle recouvre l'empan 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 sélectionne des enregistrements de métadonnées d'images dont les mesures météorologiques associées respectent deux contraintes :

1. elles proviennent de stations situées dans l'empreinte de l'image (?img_md_geo étant la géométrie d'une image et ?ws_geo celle d'une station météo, la géométrie de l'image doit contenir celle de la station météo, ce qu'indique la relation *geof:sfContains* dans le filtre *FILTER (geof:sfContains(?img_md_geo, ?ws_geo))*), et

2. elles ont été collectées durant une période donnée ; l'utilisateur a défini une période d'intérêt d'une semaine après la prise d'image (24x7=168 heures, 'PT168H'), qui est utilisée au sein des deux dernières clauses *FILTER*.

Faute de place, nous n'avons pas indiqué les éléments de la requête filtrant les images présentes dans la zone spécifiée par l'utilisateur¹⁵.

```
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.
```

15. Une vidéo montre le fonctionnement du prototype ici : <http://melodi.irit.fr/videos/metadataBrowserDataMeteoFrance.mp4>

```

?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))
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.

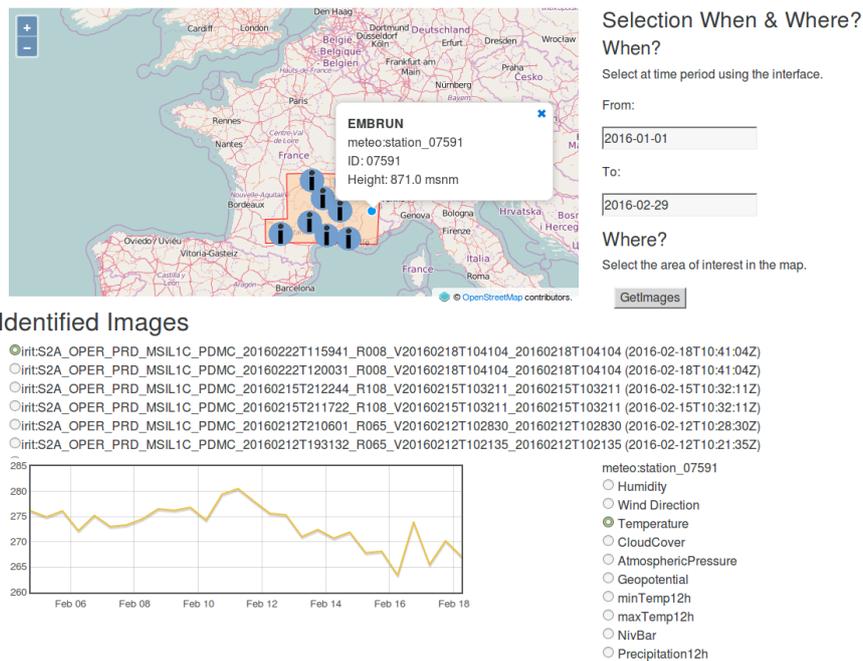


FIGURE 6. Affichage des informations météorologiques associées à une image satellite en réponse à la requête.

4.3. Raisonnement sur des données additionnelles

Le modèle a été conçu de façon à permettre l'intégration d'autres types de données spatio-temporelles. En comparant la dimension spatiale de l'image avec la composante spatiale d'autres ensembles de données, il est possible de déduire des relations topologiques. Si l'on ajoute des comparaisons similaires pour le domaine temporel, des raisonnements de haut niveau peuvent être faits. Par exemple, en prenant en compte les unités administratives de France (Régions, Départements, Communes, etc.), qui ont une représentation spatiale qui peut être représentée à l'aide de GeoSPARQL, nous avons les faits suivants :

```
Occitanie a Region .
HauteGaronne a Departement .
Revel a Commune .

Occitanie hasSpatialRepresentation Occitanie_geo .
HauteGaronne hasSpatialRepresentation HauteGaronne_geo .
Revel hasSpatialRepresentation Revel_geo .
```

À partir de ces faits et considérant que `Occitanie_geo`, `HauteGaronne_geo` et `Revel_geo` sont des instances de `geo:Feature`, il est possible de définir les relations topologiques suivantes entre les représentations :

```
Occitanie_geo geo:sfContains HauteGaronne_geo .
HauteGaronne_geo geo:sfContains Revel_geo .
```

Étant donné que les métadonnées d'images sont décrites en utilisant GeoSPARQL, il est possible de calculer les relations topologiques entre une image et une unité administrative. Dans cet exemple, nous pourrions avoir un enregistrement de métadonnées `MdImg01` défini comme ci-dessous :

```
Img01 isDescribedBy MdImg01 .
MdImg01 a dcat:Dataset .
MdImg01 hasSpatialRepresentation MdImg01_geo .
MdImg01_geo geo:sfContains HauteGaronne_geo .
```

De ces faits, le système peut inférer, par transitivité de la relation `geo:sfContains` que la zone géographique de Revel est aussi incluse dans celle de l'image :

```
MdImg01_geo geo:sfContains Revel_geo .
```

De plus, si on considère un évènement qui s'est produit dans une commune spécifique, mais pour laquelle il n'y a pas de coordonnées géographiques précises, on peut exploiter la composante spatiale de la commune pour inférer de nouveaux faits sur l'évènement. Par exemple, le cas d'une `MaladieX` a été rapporté dans la commune `CommuneXYZ` ; on associe la représentation spatiale de la commune à la `MaladieX` comme suit :

```
MaladieX hasSpatialRepresentation MaladieX_geo .
MaladieX_geo owl:sameAS CommuneXYZ_geo .
ou
CommuneXYZ_geo geo:sfContains MaladieX_geo .
```

Ensuite, considérant que :

```
Departement1234_geo geo:sfContains CommuneXYZ_geo .
```

le système peut inférer que

```
Departement1234_geo geo:sfContains MaladieX_geo .
```

Selon le type d'analyse, d'autres relations spatiales peuvent être pertinentes. Par exemple, si nous savons que `MaladieX` a été identifié dans `CommuneXYZ`, un agent agricole pourrait accorder une attention particulière à toutes les communes qui partagent une frontière avec `CommuneXYZ` :

```
CommuneXYZ_geo geo:sfTouches ?NeighborCommune_geo .
```

5. Conclusions et perspectives

Nous avons présenté une approche pour intégrer sémantiquement des données géo-localisées et datées ainsi que des métadonnées 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 métadonnées 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 images à partir des informations qui leur sont associées sous forme de métadonnées sémantiques. Ces métadonnées correspondent aux métadonnées des images satellites et à des données de capteurs relevant des informations sur la zone couverte par l'image, dans notre cas des données météorologiques. Nous disposons ainsi d'un cadre qui permet d'enrichir progressivement la description des images en fonction des besoins et des usages. Du point de vue de la recherche sur le web sémantique, une première contribution de ce travail est de montrer l'intérêt de partager et diffuser des modèles structurés grâce à une ontologie non seulement pour standardiser et mieux interpréter les données, mais aussi pour raisonner et interroger ces données grâce à GeoSPARQL. Ce travail souligne également l'intérêt actuel du couplage des dimensions spatiales et temporelles dans les modèles sémantiques, à rapprocher des travaux sur la modélisation des événements. Enfin, notre contribution inclut le modèle ontologique lui-même, disponible en ligne ¹⁶.

16. <http://sparkindata.irit.fr/>

Pour aller plus loin, il serait intéressant de présenter non plus seulement des données brutes mais des interprétations pertinentes de ces données en fonction des besoins des utilisateurs. Dans le projet SparkinData, un premier travail a été mené dans cette direction avec des documents en langage naturel : au lieu de présenter le document pdf dans son ensemble, nous en extrayons des données pertinentes qui sont associées aux images concernées. Dans notre cas, il s'agit de bulletins agricoles d'où sont extraites des informations sur les attaques constatées sur des cultures. Une autre approche possible est d'identifier des patrons de requête qui augmenteraient la pertinence des images et informations retournées en réponse à une requête, en s'appuyant sur des règles et du raisonnement. Une facette complémentaire est de produire des analyses préalables sur les données et présenter des vues synthétiques ou des résultats de plus haut niveau. Par exemple, avec les données que nous avons utilisées, on pourrait analyser les séries temporelles de températures et détecter ainsi une canicule. Dans cette perspective, nous sommes en train d'implémenter des modules d'analyse des données (calculs de valeurs moyennes ou de référence, de tendances et comparaison aux valeurs associées à une image ou une séries d'images) mais aussi de recueillir des connaissances expertes d'interprétation des données et de les modéliser. Nous sommes également en train d'intégrer d'autres sources de données propres au domaine de l'agriculture et des productions agricoles, y compris des données extraites de textes, et de fournir des règles et des outils de raisonnement pour assister les analyses. Ainsi, nous mettrons à l'épreuve l'utilisabilité du modèle pour d'autres données, et sa capacité à rendre compte de divers types de connaissances, autres que des données des capteurs.

Pour le moment, la spécialisation du modèle générique, à savoir l'alignement entre métadonnées des images et propriétés ou classes du vocabulaire, est faite à la main. Dans le futur, nous envisageons d'intégrer des procédures d'alignement automatique pour identifier et typer les éléments ayant une dimension spatiale et temporelle dans les données à intégrer. Nous prévoyons également d'associer les données annotées par les vocabulaires que nous utilisons et celles provenant d'autres sources du LOD.

Acknowledgement

The authors are partially supported by the French FUI SparkinData project, that designed the SparkinData cloud framework, see <https://sparkindata.com/>. Our contributions to this project are available on line at <http://melodi.irit.fr/>.

Bibliographie

- Arenas H., Aussenac-Gilles 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.
- Bereta K., Smeros P., Koubarakis M. (2013). Representation and querying of valid time of triples in linked geospatial data. In P. Cimiano, O. Corcho, V. Presutti, L. Hollink, S. Rudolph (Eds.), *The semantic web: Semantics and big data: 10th international conference, eswc 2013, montpellier (f), may 26-30, 2013. proceedings*, p. 259–274. Berlin, Heidelberg, Springer Berlin Heidelberg. Consulté sur https://doi.org/10.1007/978-3-642-38288-8_18
- Blower J., Clifford D., Goncalves P., Koubarakis M. (2014). The MELODIES project: Integrating diverse data using linked data and cloud computing. In *Big data from space*, p. 244–247. Frascati, Italy.
- Blower J., Riechert M., Pace N., Koubarakis M. (2016). Big Data meets Linked Data - What are the opportunities? In *Big data from space*. Santa Cruz de Tenerife, Spain.
- Clifford D., Blower J., Alegre R., Philips R. (2014). Annotating Climate Data with Commentary: The CHARMe Project. In *Big data from space*, p. 251–254. Frascati, Italy.
- 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>
- Desconnets J.-C., Chahdi H., Mougnot I. (2014). Application Profile for Earth Observation Images. In *Metadata and semantics research: 8th research conference, mtsr 2014 karlsruhe, germany, november 27-29, 2014 proceedings 13*, vol. 478.
- Egenhofer M. J. (2002). Toward the semantic geospatial web. In *Proceedings of the 10th acm international symposium on advances in geographic information systems*, p. 1–4. New York, NY, USA, ACM. Consulté sur <http://doi.acm.org/10.1145/585147.585148>
- 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.
- European Space Agency. (2016a). *Sentinel 1*. <https://sentinel.esa.int/web/sentinel/missions/sentinel-1>. (Accessed: 2017-04)
- European Space Agency. (2016b). *Sentinel 2*. <https://sentinel.esa.int/web/sentinel/missions/sentinel-2/>. (Accessed: 2017-04)
- 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.
- Gasperi J., Houbie F., Woolf A., Smolders S. (2016). *Earth Observation Metadata profile of Observations & Measurements*. Rapport technique. Open Geospatial Consortium. Consulté sur <http://docs.opengeospatial.org/is/10-157r4/10-157r4.html>
- Gatti A., Bertolini A. (2015). *Sentinel-2 Products Specification Document*. Rapport technique. European Space Agency. Consulté sur https://sentinel.esa.int/documents/247904/349490/S2_MSI_Product_Specification.pdf/ (Accessed: 2017-04)

- Janowicz K., Scheider S., Pehle T., Hart G. (2012). Geospatial Semantics and Linked Spatio-temporal Data – Past , Present , and Future. *Semantic Web*, n° 3, p. 321–332.
- 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
- Koubarakis M. (2014). Linked Open Data in the Earth Observation Domain: The Vision of Project LEO. In *Big data from space*, p. 366–369. Frascati, Italy.
- Koubarakis M., Kyzirakos K. (2010). Modeling and Querying Metadata in the Semantic Sensor Web: The Model stRDF and the Query Language stSPARQL. In L. Aroyo *et al.* (Eds.), *The Semantic Web: Research and Applications: 7th Extended Semantic Web Conference, ESWC 2010, Proceedings, Part I*, p. 425–439. Berlin, Heidelberg, Springer Berlin Heidelberg. Consulté sur http://dx.doi.org/10.1007/978-3-642-13486-9_29
- Kyzirakos K., Manegold S., Vlachopoulos I., Savva D., Koubarakis M., Nikolaou C. *et al.* (2014). *LEO Deliverable 2.1: Data models and languages for mapping EO data to RDF*. Rapport technique. European Union.
- Kyzirakos K., Savva D., Manegold S., Koubarakis M. (2015). *LEO Deliverable 2.2.2 : Tools for transforming EO products to RDF-v2*. Rapport technique. European Union.
- Kyzirakos K., Vlachopoulos I., Savva D., Pirk H., Manegold S., Koubarakis M. (2014). *LEO Deliverable 2.2.1: Tools for transforming EO products to RDF-v1*. Rapport technique. European Union.
- Lankester T., Coleman E., Hubbard S., Vesconi F., Ottavianelli G., Martin J. (2014). Linked Data Approach to handling commentary metadata for Copernicus Quality control. In *Big data from space*, p. 248–250. Frascati, Italy.
- Maali F., Erickson J. (2014). Data Catalog Vocabulary (DCAT). In *W3c recommendation 16 january 2014*.
- Mitchell A., Lowe D., Murphy K., Pilone D. (2014). Evolving Metadata in NASA Earth Science Data Systems to Support Big Data. In P. Soille, P. Marchetti (Eds.), *Big data from space*, p. 236–239. Frascati, Italy, Joint Research Centre.
- Perego A., Friis-Christensen A., Lutz M. (2017). *GeoDCAT-AP: Use Cases and Open Issues*. Rapport technique. European Commission, Joint Research Centre. Consulté sur https://www.w3.org/2016/11/sdsvoc/SDSVoc16_paper_25 (Accessed: 2017-03)
- 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*.
- Raskin R. (2006). *Guide to SWEET Ontologies*. Rapport technique n° 9, vol. 53. Pasadena, CA, USA, NASA/Jet Propulsion Lab.
- 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 Agency.

- Stadler C., Lehmann J., Höffner K., Auer S. (2012). LinkedGeoData: A Core for a Web of Spatial Open Data. *Semantic Web*, vol. 3, n° October 2012, p. 333–54.
- TELEIOS. (2016). *Virtual Observatory Infrastructure for Earth Observation Data*. http://www.earthobservatory.eu/Ontologies_and_Linked_Geospatial_Data. (Accessed: 2017-04)
- Voges U. (2014). Providing ESA HMA Catalogues EO Collection and Product Metadata as Linked Open Data. In *Big data from space*, p. 240–243. Frascati (I), Joint Research Centre.
- Voges U., Becker P., Einspainer U. (2016). Linking big data from space to apps on earth. In *Big data from space*. Santa Cruz de Tenerife, Spain.