



Ontologie pour l'intégration de données d'observation de la Terre et contextuelles basée sur les relations topologiques

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. Ontologie pour l'intégration de données d'observation de la Terre et contextuelles basée sur les relations topologiques. 29es Journées Francophones d'Ingénierie des Connaissances, IC 2018, AFIA, Jul 2018, Nancy, France. pp.5-20. hal-01839542

HAL Id: hal-01839542

<https://hal.science/hal-01839542>

Submitted on 23 Jul 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.

Ontologie pour l'intégration de données d'observation de la Terre et contextuelles basée sur les relations topologiques

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

IRIT, CNRS, UNIVERSITÉ TOULOUSE 2 ET UNIVERSITÉ DE TOULOUSE France
{prenom.nom}@irit.fr

Résumé : Nous proposons une ontologie pour l'intégration de données d'observation de la Terre et de données contextuelles, à l'aide de relations topologiques spatiales et temporelles. Cette ontologie spécialise des standards, notamment SOSA, GeoSPARQL et OWL-Time. La base de connaissance décrite par cette ontologie est alimentée grâce à un processus qui sélectionne, transforme et intègre de données géospatiales hétérogènes (méta-données d'image satellite, données météorologiques, unités administratives, couverture terrestre, etc.). Ce processus s'appuie sur le tuilage des images pour traiter les données ayant une composante spatiale fixe, alors que les relations temporelles sont calculées à la volée à partir d'une topologie temporelle. Nous validons cette approche par un cas d'étude exploitant des méta-données d'image satellite Sentinel.

Mots-clés : Ontologies, intégration de données, données spatiales et temporelles, observation de la Terre, SOSA.

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 le programme Sentinel avec deux types de satellites, Sentinel-1 and Sentinel-2, qui transmettent des images de haute qualité (entre 8 à 10 To de données quotidiennement). Ces images de la Terre sont captées selon différentes technologies et libres d'accès. Cette disponibilité des 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 néanmoins un besoin de coupler les images avec des données sur les zones observées. Ces données sont accessibles à partir de différentes sources dans des formats hétérogènes et des temporalités différentes : elles peuvent être statiques, comme les données sur la couverture terrestre, ou dynamiques, comme les observations météo. Elles peuvent être utiles par exemple pour indiquer qu'une image contient une région touchée par un phénomène tel qu'un tremblement de terre ou une canicule, et ensuite pour décider des actions à mener dans cette zone ou conduire à des analyses à plus long terme. Plus encore, en exploitant les caractéristiques spatio-temporelles d'un phénomène (son empreinte spatiale et sa date), il devient possible de savoir si une entité localisée dans l'empreinte de l'image (e.g. une ville) a subi le même phénomène.

Les images satellites étant décrites par des méta-données, le problème revient à intégrer à ces méta-données des données provenant de sources variées et très hétérogènes (format, représentation). L'apport des technologies sémantiques pour faciliter cette tâche a été démontré dans des travaux antérieurs (Reitsma & Albrecht, 2005; Sukhobok *et al.*, 2017), en particulier grâce à l'utilisation d'ontologies comme représentations formelles des connaissances d'un domaine donné. Ainsi, dans la continuité des travaux sur l'accès et l'intégration de données via les ontologies (i.e. OBDA pour "Ontology-based Data Access" et OBDI pour "Ontology-based Data Integration") (Lenzerini, 2011; Lefrançois *et al.*, 2017; Console & Lenzerini, 2014), nous avons conçu un vocabulaire sémantique permettant de représenter les données des différentes sources envisagées et d'y accéder de façon homogène. Notre approche a consisté à construire une ontologie modulaire permettant de répondre aux besoins

d'exploitation propres à chaque source de données. Cette approche permet de réduire partiellement la complexité et l'hétérogénéité des données, et ainsi de faciliter le peuplement de l'ontologie. Une caractéristique commune aux observations de la Terre est qu'elles peuvent être liées via leurs propriétés topologiques spatiales et temporelles. Un préalable au peuplement de l'ontologie (i.e. la production d'entrepôts RDF), a été de convertir les différents ensembles de données géo-spatiales fournis dans des formats hétérogènes (shapefile, KML, CSV, GeoJSON, TIFF), en un format commun, JSON, voire d'assurer, le cas échéant, la compatibilité de leurs propriétés et relations spatiales et temporelles, puis de les stocker dans une base de données NoSQL MongoDB.

Nous présentons dans cet article le vocabulaire que nous avons défini pour assurer la description sémantique homogène des différents types de données sous forme d'entités ayant des propriétés spatiales et temporelles. Une partie des données géospatiales à intégrer aux méta-données d'image sont des données contextuelles mesurées à la surface de la Terre, que nous traitons comme des données de capteurs. Ce vocabulaire spécialise ainsi des vocabulaires connus du LOD, dont SOSA¹ pour les données de capteurs, GeoSPARQL (Kolas *et al.*, 2013) pour gérer les relations et coordonnées spatiales, et OWL Time² pour traiter la dimension temporelle. De fait, nous nous appuyons sur une topologie des entités, en distinguant notamment les données dynamiques, dont la validité est fournie par la composante temporelle, et les données statiques, pour lesquelles la composante spatiale joue un rôle plus primordial.

Nous montrons, à travers un cas d'étude, comment cette ontologie permet de prendre en compte les spécificités de différentes sources en termes de vocabulaire, périodicité et volume, et comment elle peut améliorer l'accès aux données. A titre d'exemple, nous avons généré des entrepôts RDF qui mettent en relation des informations sur la couverture terrestre (% de forêts, d'eau, etc.) d'une part, et des relevés de stations météorologiques (température, humidité, etc.) d'autre part, avec les images satellites. Ce travail est mené dans le cadre du projet SparkinData visant à construire une plate-forme cloud de données d'observations de la Terre. Le cas d'étude s'appuie sur l'annotation sémantique de méta-données d'images brutes fournies par le CNES (Centre National d'Etudes Spatiales).

Le reste de l'article est organisé ainsi : la partie 2 expose des travaux liés ; en partie 3, nous présentons le modèle d'intégration ; son exploitation pour représenter en RDF les diverses sources d'informations envisagées est l'objet de la partie 4 ; la partie 5 décrit différentes stratégies mises en oeuvre pour intégrer les données des entrepôts RDF via les relations topologiques. Enfin, nous concluons et présentons des perspectives à ce travail en partie 6.

2 Travaux liés

2.1 Modèle sémantique pour l'imagerie satellitaire

Le projet européen TELIOS (Virtual Observatory Infrastructure for Earth Observation Data)³ a été pionnier dans l'utilisation de représentations sémantiques pour décrire des images de satellites et faciliter l'accès à des données d'observation de la Terre par des agences telles que la NASA ou l'ESA. Le cadre applicatif de ce projet est la gestion des incendies en Grèce (The TELEIOS Team, 2012). Les images sont classifiées et représentées sous forme vectorielle, et à chaque vecteur est associée une estimation des risques d'incendie. Les vecteurs sont liés à des sources de données externes par le biais d'ontologies (TELEIOS, 2016). Ces ontologies exploitent notamment le standard stRDF qui étend RDF pour les données spatiales et temporelles ; seule la composante spatiale est exploitée dans les analyses réalisées.

Le travail de Espinoza-Molina & Datcu (2013) et Espinoza-Molina *et al.* (2015) a été influencé par TELEIOS. Leur système permet d'ajouter des annotations sémantiques à des images produites par des radars à synthèse d'ouverture (SAR) selon le processus suivant :

1. https://www.w3.org/2015/spatial/wiki/SOSA_Ontology

2. [https://www.w3.org/TR/owl-time/\(10/2017\)](https://www.w3.org/TR/owl-time/(10/2017))

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

1) extraction des informations pertinentes à partir des fichiers de métadonnées des images (fichiers XML); 2) division de l'image en *tuiles* de tailles différentes et organisation de ces tuiles sous une forme pyramidale; 3) extraction puis transformation sous forme de vecteurs des caractéristiques de chaque tuile. Ce système est capable de trouver de l'information en utilisant des requêtes portant sur les méta-données, la sémantique et le contenu spatial des images. Cette approche associe aux images d'observation de la Terre, des données provenant des sources externes comme Corine Land Cover, Urban Atlas, Geonames et LinkedGeoData. Elle s'appuie sur un modèle ontologique, SAR, qui réutilise la taxonomie TerraSARX et étend celle de TELEIOS (2016).

L'approche de Keßler & Farmer (2015) utilise à la fois des sources de données externes et des relations temporelles. Elle s'appuie sur le concept de *prisme spatio-temporel* ("space-time prism") qui décrit les caractéristiques de 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 temporelle. 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*. Ainsi, dans les trajectoires étudiées par 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.

Compte-tenu du volume d'information à gérer, la recommandation RDF DataCube du W3C (Brizhinev *et al.*, 2017) suggère de lier les images à des tuiles de telle sorte à faire des assertions sur les tuiles. Dans ce cadre, les tuiles sont des zones carrées géo-localisées définies suite à une décomposition de la surface de la terre à partir d'une grille. Chaque image fournie par Sentinel-2 Single Tile (S2ST) a déjà une tuile. Nous nous appuyons également sur la notion de tuile dans notre processus d'intégration.

2.2 Données géographiques et LOD

La collecte et l'intégration de données géographiques produites par une diversité de disciplines est au cœur du projet Digital Earth (Gore, 1998). Rendre les données géographiques disponibles puis interopérables à un niveau sémantique est une préoccupation qui conduit à appliquer les principes du Linked Data (Blázquez *et al.*, 2014), et ainsi exposer, partager, et intégrer les données sur le Web via des URI déréférencables (Heath & Bizer, 2011). Les guides du W3C pour publier des données spatiales sous forme de LOD (Linked Open Data), attirent l'attention sur la représentation des relations spatiales et des systèmes de référence (CRS, pour "Coordinate Reference Systems") (Tandy *et al.*, 2017). De nombreuses ontologies et vocabulaires sont recommandés pour représenter des données raster géospatiales volumineuses dans le LOD (Linked Open Data). Le W3C suggère l'ontologie RDF Data Cube (QB), introduite ci-dessus, combinée à d'autres ontologies standards du W3C et de l'OGC dont SSN (Semantic Sensor Network)⁴, OWL-Time⁵, SKOS⁶, PROV-O⁷ et la récente extension de DataCube pour les entités spatio-temporelles, QB4ST⁸.

L'OGC a introduit la notion de *données géo-liées* ("geolinked data") pour faire référence aux données liées géographiquement. Dans les premiers travaux, la géométrie était stockée dans un ensemble de données géospatiales séparé, et non directement comme valeur d'attributs. Cette option étant trop contraignante lorsqu'il faut comparer la géométrie de chaque entité, les entrepôts actuels mémorisent ensemble, une représentation RDF de la géométrie et une représentation RDF des entités spatiales. Atemezing (2015) a identifié divers types de géométries (point, ligne ou polygone) et divers outils pour construire une représentation RDF

4. <http://purl.oclc.org/NET/ssnx/ssn>

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

6. <http://www.w3.org/2004/02/skos/core>

7. <https://www.w3.org/TR/prov-o>

8. <https://www.w3.org/TR/qb4st/>

de la géométrie (comme Geometry2RDF⁹ ou TripleGeo¹⁰). Il modélise aussi quatre vocabulaires pour représenter les CRS, les entités topographiques et leurs géométries. Ces ontologies étendent des vocabulaires existant et offrent deux avantages supplémentaires : une utilisation explicite du CRS identifié par des URI pour la géométrie, et la possibilité de décrire des géométries structurées en RDF. Les données sont publiées comme la base de données française GEOFLA.

Une autre considération à intégrer pour représenter des données d'observations de la Terre est la différence de validité temporelle des données. Certaines données, comme la position des stations météorologiques, des villes et de la plupart des lieux administratifs, et même la couverture terrestre, sont valides pour une très longue période, plus longue que celle de l'application, et peuvent être considérées stables ou statiques. Les flux de données, au contraire, fournissent en continu de nouvelles données à intervalles de temps réguliers. Par exemple, les mesures de température sont données toutes les 3 heures par les bulletins Météo France, et des dizaines d'images satellites et leurs méta-données sont disponibles sur le serveur PEPS chaque jour.

3 Modèle d'intégration

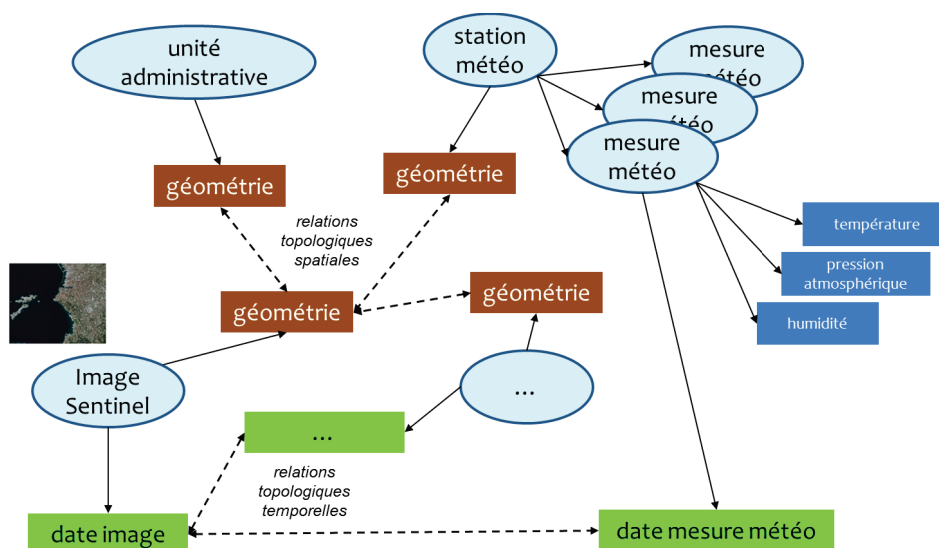


FIGURE 1 – Illustration du rôle pivot de la géométrie et de la datation dans le modèle de données pour associer les données par des relations topologiques.

Le modèle retenu pour intégrer des données aux métadonnées des images repose sur l'hypothèse que ces données sont géolocalisées et datées. Pour chacune, on doit connaître a minima un point déterminé par sa latitude / longitude, ou une zone géolocalisée (appelée « géométrie » dans la Figure 1). C'est par le biais de la géométrie que des données peuvent être associées à une même image ou à une zone dans une image. Par exemple sur la Figure 1, les mesures météorologiques (humidité, température, pression) ainsi que les unités administratives (villes, régions, etc.) sont géolocalisées et couvrent une superficie plus ou moins grande. Pour comparer les géométries des données à lier, il faut les ramener à une même unité, ce qui permet de savoir quelle zone de l'image est concernée ou décrite par ces données. Les images étant datées, dès lors que les autres données sont aussi datées, comme les mesures météorologiques, il est possible de les lier par des relations temporelles, et de savoir par exemple

9. <https://github.com/boricles/geometry2rdf>

10. <https://github.com/GeoKnow/TripleGeo>

quelles sont les températures relevées sur la zone couverte par l'image durant les 3 jours qui ont précédés la prise d'image.

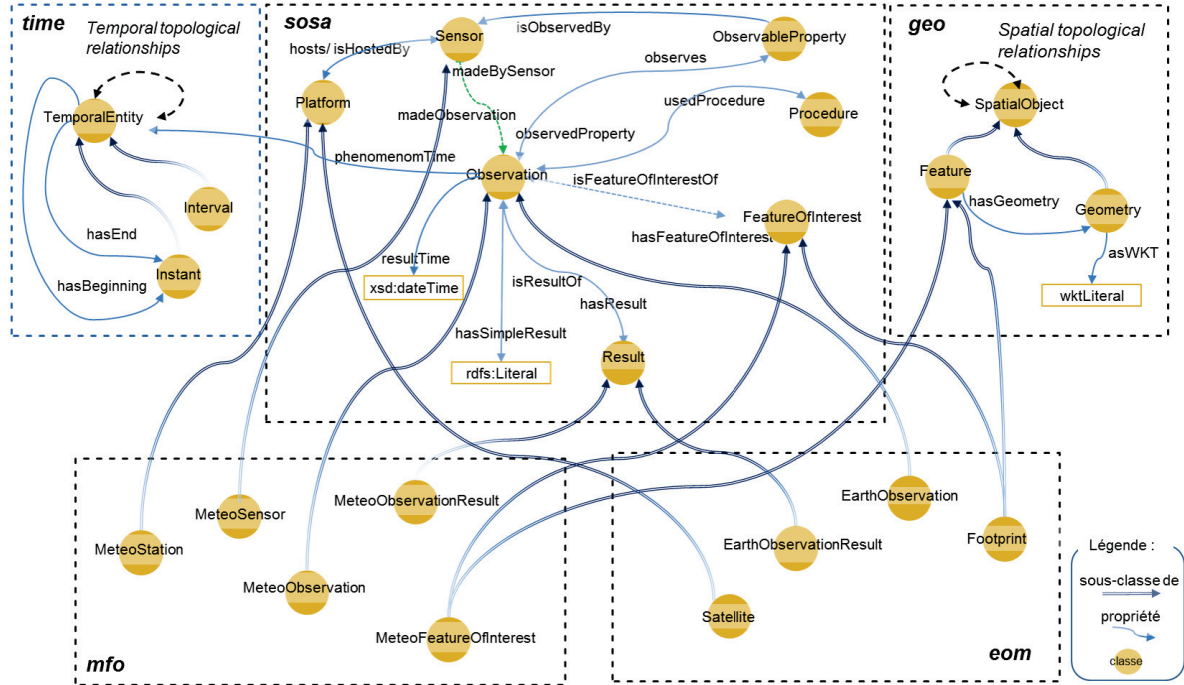


FIGURE 2 – Le modèle d'intégration de données.

Nous détaillons à présent le modèle conçu comme support à cette approche. Pour permettre l'intégration de divers types de données, ce modèle, présenté sur la Figure 2 comporte une partie générique (les trois cadres *time*, *sosa* et *geo* de la partie haute de la figure), composée de classes et de propriétés tirés de vocabulaires existants (respectivement OWL-Time, SOSA, GeoSPARQL), et une partie spécifique aux données à intégrer (cadres de la partie basse de la Figure 2). La partie spécifique comprend a minima un vocabulaire pour décrire les métadonnées d'images (cadre *eom* pour "Earth Observation Model") et autant de vocabulaires que de types de données à intégrer (un seul vocabulaire est mentionné ici, *mfo* "pour MétéoFrance Observation", représentant les données météorologiques).

3.1 Les vocabulaires réutilisés

Pour faciliter le partage de connaissances, une bonne pratique dans la construction d'ontologies consiste à identifier des ontologies existantes à réutiliser, à les connecter ou les étendre. Nous commençons donc par décrire les vocabulaires sur lesquels s'appuie la partie générique de notre modèle : OWL-Time, SOSA et GeoSPARQL, désignés par la suite respectivement par les préfixes *time*, *sosa* et *geo*. D'autres vocabulaires peuvent être réutilisés en fonction des données à intégrer pour la partie spécifique.

3.1.1 Vocabulaire pour représenter méta-données d'images et données de capteurs

Un grand nombre de modèles ont été développés pour représenter des méta-données d'observations par les communautés des sciences de la Terre et de l'environnement. O&M (Observations and Measurements), développé par le groupe SWE de l'OGC, fournit une vision utilisateur des observations. L'OGC Sensor Observation Service and Web Feature Service en a fourni une implémentation XML (spécification XSD). Plus général, DCAT est un vocabulaire recommandé par le W3C pour publier des métadonnées de catalogues sur le Web et disponible au format RDF (Maali & Erickson, 2014).

Pour faciliter l'intégration de données provenant de différents types de capteurs, l'ontologie SSN¹¹ décrit des réseaux de capteurs numériques ainsi que les fonctionnalités, propriétés et mesures de ces capteurs. Construite comme une sous-couche de l'ontologie formelle DOLCE-UltraLite (DUL), SSN est générique au sens où elle ne définit pas toutes les propriétés spécifiques à des capteurs et des observations particuliers. En revanche, on peut la lier à des ontologies ou des vocabulaires propres à un domaine tels que l'ontologie SWEET (Semantic Web for Earth and Environmental Terminology) qui décrit des données environnementales (Raskin, 2006). Pour faciliter sa réutilisation, et éviter les risques d'incompatibilité notamment avec les déclarations DUL, une version modulaire de SSN a été publiée. Elle est composée de plusieurs ontologies couvrant le même domaine, mais avec une portée sémantique différente pour s'adapter à des contextes d'utilisation différents. SOSA (Sensor, Observation, Sample, and Actuator) est le module ontologique noyau de la nouvelle version de SSN ; à ce titre, il regroupe les termes centraux de l'ontologie. Cox (2017) propose des alignements entre différents modèles de données d'observations, dont SOSA, à l'aide de l'ontologie PROV-O, pour faciliter l'intégration des données produites avec chacun d'eux. Les axiomes d'alignement sont fournis en RDF dans un module de la spécification SOSA/SSN. Ils s'appuient notamment sur le fait qu'une observation dans le vocabulaire SOSA (*sosa:Observation*) est considérée comme une activité (*prov:Activity*) de capteur, i.e. tout processus (calcul, simulation, interprétation, etc.) appliquant une procédure pour obtenir une estimation d'une valeur de propriété. La classe *sosa:Observation* y est ainsi définie comme une sous-classe de *prov:Activity*.

Nous avons dans un premier temps fait le choix de DCAT et SSN pour représenter respectivement les catalogues de méta-données d'images et les données issues de stations météorologiques (Arenas *et al.*, 2016a,b). SOSA étant pertinent pour une vaste gamme d'applications, dont l'imagerie satellite, nous l'avons finalement retenu pour représenter à la fois les méta-données d'image et les données météorologiques.

3.1.2 Vocabulaire pour représenter la composante spatiale des données

Pour représenter l'information géospatiale, une des normes les plus connues est GeoSPARQL, un standard de l'OGC qui définit une petite ontologie pour la représentation de caractéristiques, de relations et de fonctions spatiales (Kolas *et al.*, 2013; Battle & Kolas, 2012; Perry & Herring, 2012). Le cadre *geo* de la Figure 2 présente les principales classes de GeoSPARQL. La classe *geo:Feature* représente toute entité du monde réel ayant une empreinte spatiale. Cette empreinte est décrite par une "géométrie" (point, polygone, etc.), 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 domaine considérée.

Parmi les alternatives à GeoSPARQL, GeoRDF sert à représenter des propriétés géographiques de points telles que la latitude, la longitude, l'altitude (en utilisant WGS84 comme référentiel) ; GeoOWL permet de représenter l'empreinte d'objets spatiaux plus complexes (lignes, rectangles, polygones) ; stSPARQL est une autre extension de SPARQL conçue pour interroger des graphes RDF spatio-temporels (Koubarakis & Kyzirakos, 2010; Bereta *et al.*, 2013). Nous avons retenu GeoSPARQL car il offre la possibilité de raisonner sur des entités (*geo:Feature*) ayant une forme géométrique (*geo:Geometry*), à savoir proposer des relations entre des entités sur la base des relations entre leurs géométries (inclusion, recouvrement, etc). GeoSPARQL assure l'expression de relations topologiques spatiales binaires sous forme de propriétés des *geometries* et des *features*, et de fonctions servant à les comparer sur la base des propriétés. Parmi les fonctions spatiales de GeoSPARQL, on peut citer *equals*, *disjoint*, *intersects*, *touches*, *within*, *contains*, *overlaps*, et *crosses*. Grâce à des requêtes utilisant ces fonctions, il est possible de comparer, à la volée, deux "*geometries*" ou deux "*features*", ou de comparer une ressource géolocalisée à une position fournie explicitement. La requête GeoSPARQL ci-après recherche les "objets" situés dans (fonction GeoSPARQL *sfWithin*) un polygone (un triangle) défini par les coordonnées de ses trois sommets :

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

```
select ?obj_loc
WHERE {
  ?obj_loc a geo:Feature .
  ?obj_loc geo:hasGeometry ?geo_obj .
  FILTER ( geof:sfWithin( ?geo_obs ,
    "POLYGON ((-3.562096 42.073807,
      -0.442349999999998 42.476536,
      -0.789792999999997 43.973148))" ^geo:wktLiteral) )
}
```

3.1.3 Vocabulaire pour représenter la composante temporelle des données

Parmi les modèles de données et langages de requêtes du web sémantique permettant de représenter la composante temporelle de données, on trouve OWL-Time, le standard du W3C, et SWRL Temporal Ontology. Le cadre *time* de la Figure 2 présente les principales classes de OWL Time¹². La classe *time:TemporalEntity* représente toute entité ayant une temporalité, i.e. un début (propriété *time:hasBeginning*) et une fin (propriété *time:hasEnd*), et donc une durée (propriété *time:hasDuration*). Une entité dont la durée n'est pas nulle ("début" et "fin" sont différents) est un *intervalle*, sinon il s'agit d'un *instant* (*time:Interval* et *time:Instant* étant des spécialisations de *time:TemporalEntity*). Les entités temporelles peuvent être liées par des relations binaires comme *meets*, *overlaps*, *during* issues de l'algèbre des intervalles d'Allen et servant de base pour les raisonnements spatio-temporels.

Le temps a été associé aux triplets RDF à diverses fins. Ainsi, pour donner à un triplet RDF une dimension temporelle représentant sa période de validité (*valid-time*), notion issue des Bases de Données temporelles, qui s'oppose à celle de "date de transaction" (*transaction-time*) indiquant quand un élément a été enregistré dans la base, Koubarakis & Kyzirakos (2010); Bereta *et al.* (2013) proposent le langage stRDF. Un fait *y* est représenté par un quadruplet $\langle s, p, o \rangle : t$ où *t* est une variable de type *xsd:dateTime* (un instant) ou *strdf:period* (un intervalle de temps). stSPARQL, le langage de requête associé à stRDF, supporte des fonctions permettant d'établir des relations temporelles et ainsi de rechercher des faits qui se sont produits entre deux dates. Des représentations plus complexes sont également proposées pour représenter les évolutions d'objets au cours du temps. Ainsi le modèle Continuum s'appuie sur une ontologie des *fluent*s et enrichit GeoSPARQL pour modéliser les changements qui s'opèrent sur des entités spatiales (Harbelot *et al.*, 2014).

Pour représenter la date des observations, SOSA intègre une dimension temporelle basée sur OWL-Time. Nous utilisons donc aussi ce vocabulaire.

3.2 Les vocabulaires spécifiques du modèle

Nous avons donc adopté SOSA pour décrire les données d'observations (méta-données d'image, observations météorologiques, etc.). La propriété *sosa:phenomenonTime* ayant *sosa:Observation* pour domaine et *time:TemporalEntity* pour co-domaine sert à représenter la période de temps durant laquelle le résultat d'une observation a été obtenu.

Pour représenter les relations spatiales, nous complétons notre modèle avec GeoSPARQL. Nous partons du principe que pour toute observation géolocalisée, son "*point d'intérêt*" (*FeatureOfInterest*) correspond à la zone observée; cette zone étant caractérisée par une géométrie, elle est également représentée comme un *geo:Feature*. Ainsi, pour les différents types de sources d'observations issues de capteurs, notre approche consiste à définir un nouveau vocabulaire, avec une espace de nom spécifique, approprié à la source des données. Ce vocabulaire comporte au moins une classe qui spécialise à la fois *sosa:FeatureOfInterest* et *geo:Feature*.

Les deux cadres en bas de la Figure 2 correspondent à deux vocabulaires développés selon cette approche pour représenter les méta-données d'images satellites pour l'un (vocabulaire *eom*), les observations de stations météorologiques pour l'autre (vocabulaire *mfo*). Ces vocabulaires sont présentés dans la partie 4.1.1. Dès lors qu'on est capable de représenter la composante spatiale d'une image comme instance de *geo:Feature*, on est capable de la lier à d'autres informations ayant une composante spatiale également définie comme une *geo:Feature*. De même, on peut lier par une relation temporelle un enregistrement de méta-données d'image avec des mesures observées par ailleurs, telles que les données météo, ou avec des périodes d'intérêt (par exemple "une semaine après la prise de l'image").

12. <https://www.w3.org/TR/owl-time/> (10/2017)

4 Entrepôts RDF pour les données d'observations de la Terre

4.1 Les sources de données et leurs composantes spatiale et temporelle

Nous précisons ici comment nous avons adapté les ontologies réutilisées pour concevoir un vocabulaire adapté à chaque source de données. Parmi les données géolocalisées, nous avons distingué les *données dynamiques*, dont la validité est fournie par la composante temporelle, et les *données statiques*, qui ne sont pas datées a priori ou auxquelles sont associés de très longs intervalles de temps. Nous présentons à présent les deux (sous-)modèles résultant et leur instanciation pour enrichir les méta-données d'images. Il est à noter que la propriété *time:hasTime* de OWL-Time est définie comme "un prédicat générique pour associer une entité temporelle à n'importe quoi". Donc même si notre modèle ne le mentionne pas, toute entité décrite par une URI peut être estampillée par une composante temporelle (une date ou un intervalle de temps).

4.1.1 Les observations (données dynamiques)

Le modèle est spécialisé dans 2 modules dédiés à chacune des sources d'observations issues de capteurs décrites ici, à savoir les images satellites et les observations météo.

4.1.1.1 Les méta-données d'images

Dans le projet SparkInData, nous utilisons des enregistrements de méta-données d'images Sentinel¹³. La périodicité de Sentinel-1 est de douze jours, celle de Sentinel-2 étant de cinq jours. Les enregistrements de méta-données sont obtenus au format GeoJSON à partir de RESTO, un service géré par le CNES (Gasperi, 2014). L'API RESTO permet de spécifier les paramètres à retrouver, à savoir des métadonnées d'enregistrements comme la couverture nuageuse, l'intervalle de temps, la zone géographique d'intérêt, etc. Nous collectons ces informations toutes les nuits. L'URL suivante fait appel à RESTO pour retourner tous les enregistrements de métadonnées de la collection S2ST pour la France, captés entre le 19/09/2017 23:00 et le 25/09/2017 00:00 : <https://peps.cnes.fr/resto/api/collections/S2ST/search.json?q=France&startDate=17-09-19T23:00:00&completionDate=2017-09-25T00:00:00>.

Les méta-données d'images sont obtenues sous forme de fichiers GeoJSON ; elles sont converties en RDF et représentées à l'aide du vocabulaire *eom* comme des "observations de la Terre", i.e. des instances de *eom:EarthObservation* (cadre *eom* de la Figure 2). Cette classe étant une sous-classe de *sosa:Observation*, la composante spatiale des observations est fournie via la classe *eom:Footprint*, spécialisation de *geo:Feature* et *sosa:FeatureOfInterest* : un *footprint* est décrit par un polygone fermé (une géométrie) délimitant la zone terrestre couverte par l'image. La classe *eom:Footprint* étant une sous-classe de *geo:Feature*, connaissant l'empreinte des images (leur *footprint*), il est facile d'identifier les caractéristiques d'un autre type de données qui recouvrent la position des images. La *dimension temporelle* d'un enregistrement de méta-données d'image indique le moment où l'image a été prise. Elle est fournie via la propriété *sosa:phenomenonTime* de l'observation correspondante.

Le module est complété par ailleurs pour fournir des données propres aux images satellites telles que le domaine spectral de l'image (infrarouge, UV, etc.), la direction orbitale du satellite (ascendante, descendante) ou encore la couverture nuageuse de l'image.

4.1.1.2 Les observations météorologiques

Comme données contextuelles dynamiques, nous utilisons les informations météo de *SYNOP Météo France*¹⁴. Ces observations sont réalisées toutes les trois heures par chacune des 62 stations françaises. Un fichier séparé contient la liste des stations avec leur position respective (un point fixe repéré par ses coordonnées géographiques). Selon une approche très similaire à celle de Lefort *et al.* (2012), nous utilisons des classes tirées de SOSA que nous spécialisons pour définir des classes propres au domaine de la météorologie. Dans un premier temps, nous sommes restés au plus près du schéma de la source MétéoFrance utilisée. Nous envisageons de faire évoluer ce modèle en intégrant le vocabulaire SWEET (Raskin, 2006), devenu un standard pour les données météo.

Nous représentons une station météo comme une instance de la classe *mfo:MeteoStation*, une sous-classe de *sosa:Platform*. Les capteurs fonctionnant sur une station météo sont représentés comme

13. <https://sentinel.esa.int/web/sentinel/missions/> (07/2016)

14. <https://donneespubliques.meteofrance.fr/> (07/2016)

instances de *mfo:MeteoSensor*, sous-classe de *sosa:Sensor*. La position géographique où s'effectuent les mesures est représentée comme instance de la classe *mfo:MeteoFeatureOfInterest*, une sous-classe de *sosa:FeatureOfInterest*.

Comme pour *eom*, le module *mfo* a été complété pour représenter les variables fournies avec les relevés météo; il y a 28 variables par relevé, mais beaucoup d'entre elles ne sont pas renseignées. Nous avons ainsi instancié la classe *sosa:ObservableProperty* avec notamment des individus représentant la température (*mfo:Temperature*), l'humidité (*mfo:HUMIDITY*) ou la vitesse du vent (*mfo:WIND_SPEED*), variables dont une valeur est relevée toutes les trois heures. Le module est donc enrichi régulièrement par de nouvelles instances (tous les 5 ou 12 jours pour les images satellites, quotidiennement pour les relevés météo). Le code ci-après est un extrait de la représentation RDF d'un relevé de température réalisé le 04/02/2018, entre 3h et 15h :

```
g-mfo:obs_07005_20180204150000_tminsol a mfo:Observation .
g-mfo:obs_07005_20180204150000_tminsol sosa:phenomenonTime
    g-mfo:interval_1517713200-1517756400 .
g-mfo:interval_1517713200-1517756400 a time:Interval .
g-mfo:interval_1517713200-1517756400 time:hasBeginning g-mfo:instant_1517713200 .
g-mfo:interval_1517713200-1517756400 time:hasEnd g-mfo:instant_1517756400 .
g-mfo:instant_1517713200 a time:Instant .
g-mfo:instant_1517713200 time:inXSDDateTime "2018-02-04T03:00:00"^^xsd:dateTime .
g-mfo:instant_1517713200 time:inXSDDateTimeStamp "1517713200"^^xsd:dateTimeStamp .
g-mfo:obs_07005_20180204150000_tminsol sosa:hasResult
    g-mfo:obs_07005_20180204150000_tminsol_result .
g-mfo:obs_07005_20180204150000_tminsol_result a mfo:Result .
g-mfo:obs_07005_20180204150000_tminsol_result a qudt-1-1:QuantitativeValue .
g-mfo:obs_07005_20180204150000_tminsol_result qudt-1-1:unit qudt-unit-1-1:Kelvin .
g-mfo:obs_07005_20180204150000_tminsol_result qudt-1-1:numericValue "274.75"^^xsd:double .
g-mfo:obs_07005_20180204150000_tminsol sosa:hasFeatureOfInterest
    <http://melodi.irit.fr/lod/mfo/foi_07005> .
g-mfo:obs_07005_20180204150000_tminsol sosa:observedProperty g-mfo:Temperature .
g-mfo:obs_07005_20180204150000_tminsol sosa:madeBySensor g-mfo:station_07005_Thermometer .
g-mfo:obs_07005_20180204150000_tminsol sosa:usedProcedure g-mfo:procedure_tminsol .
```

4.1.2 Les données supports (données statiques)

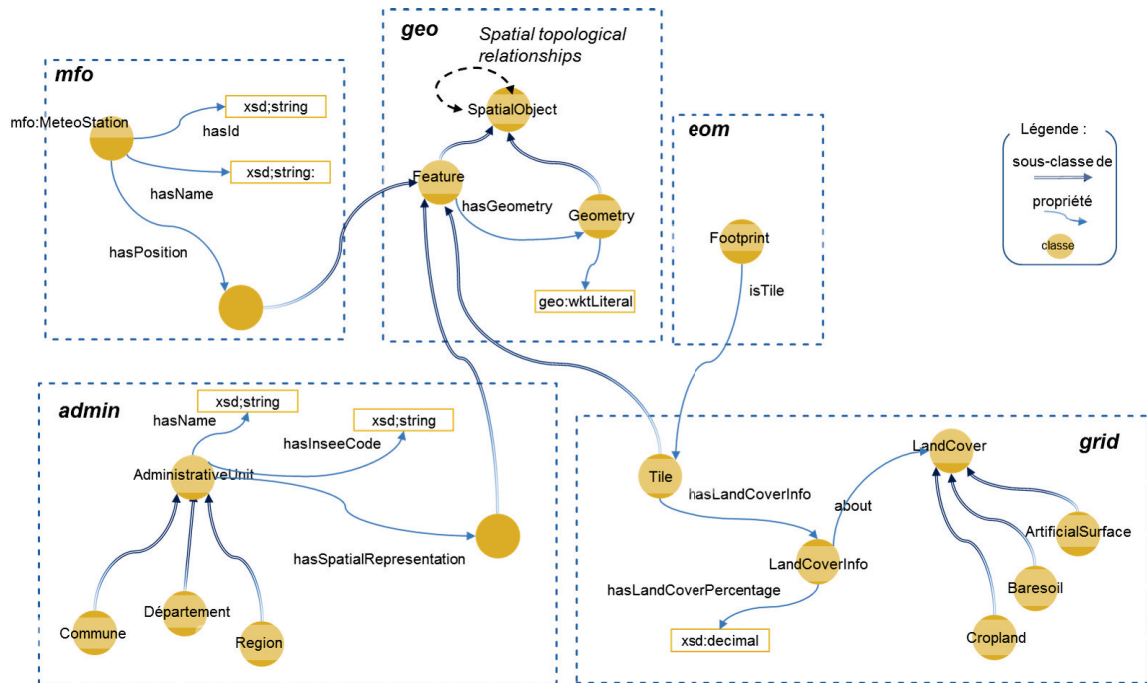


FIGURE 3 – Les vocabulaires admin et grid intégrant des données "statiques".

Le modèle peut être enrichi pour représenter des données dites statiques, i.e. qui ont une composante spatiale mais dont la représentation de la composante temporelle n'est pas nécessaire car leur valeur ne varie pas dans l'échelle de temps considérée. Nous avons créé deux modules de ce type, *admin* et *grid*, présentés sur la Figure 3. Pour mettre en relation ces données avec les données dynamiques

via des relations spatiales, un concept de chacun de ces modules spécialise la classe *geo:Feature*. Les classes de ces modules sont instanciées une fois pour toutes.

4.1.2.1 Les tuiles d'images

Les images Sentinel ont des caractéristiques différentes en fonction du capteur qui les a prises. En septembre 2016, l'ESA a commencé à diffuser des images Sentinel 2 sous forme de paquets de tuiles simples (images S2ST), chaque tuile représentant une zone fixe de la surface du globe de taille approximative de 100 x 100 km. Une image S2ST correspond donc à un fragment d'image S2. L'intérêt par rapport à une image S2 est que l'utilisateur peut mieux sélectionner la surface qui l'intéresse et ne télécharger que l'information souhaitée. Un fichier S2ST est aussi moins volumineux : il peut faire environ 500 Mo alors que celui d'une image S2 avant tuilage peut faire plus de 3 Go.

La grille de l'ESA décrivant le tuilage de la surface terrestre (Cf. Figure 4), i.e l'empreinte de chaque tuile, est fournie dans un fichier Grid au format KML¹⁵. Nous avons converti ce fichier en RDF à l'aide du vocabulaire *grid* décrit en bas de la Figure 3 : chaque tuile est représentée comme une instance de la classe *Tile*, spécialisation de *geo:Feature*, dont la propriété *geo:hasGeometry* correspond à son empreinte. Le module *eom* (Cf Section 4.1.1) a été enrichi avec la propriété *eom:isTile* pour associer une tuile au *footprint* d'une image. Ainsi en reliant des données à une tuile, on associe indirectement ces données à toutes les images associées à cette tuile.

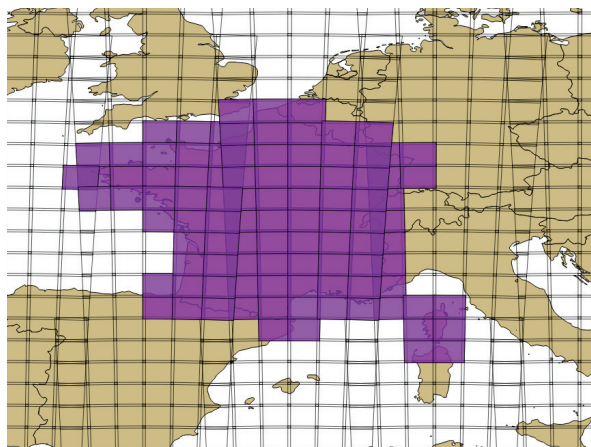


FIGURE 4 – Les tuiles de l'ESA définies pour la France métropolitaine

4.1.2.2 La couverture terrestre

Pour associer des données aux tuiles, nous avons utilisé une autre source de données statiques, le GLC-SHARE (Global Land Cover SHARE) produit par le FAO, qui donne des informations sur la couverture terrestre. Cette dernière est définie à partir d'une nomenclature permettant de classer les zones en fonction du type d'occupation des sols ou du type de surface (surface artificielle, terre cultivée, zone forestière, etc.). Les données du GLC-SHARE sont fournies sous forme d'une image au format TIFF dont chaque pixel correspond à une échelle spatiale approximativement de 1 km². La valeur d'un pixel est un entier indiquant la classe la plus fréquente pour la zone couverte par le pixel. Nous avons donc calculé la composition de la couverture terrestre de chaque tuile de l'ESA de la France. Avec le vocabulaire *grid*, une tuile est liée à un ensemble de graphes RDF (via la propriété *hasLandCoverPercentage*) décrivant chacun le pourcentage d'une classe GLC-SHARE (*CropLand*, *Baresoil*, etc.) sur la surface couverte par la tuile.

15. <https://sentinel.esa.int/web/sentinel/missions/sentinel-2/news/-/article/sentinel-2-tiling-grid-updated>

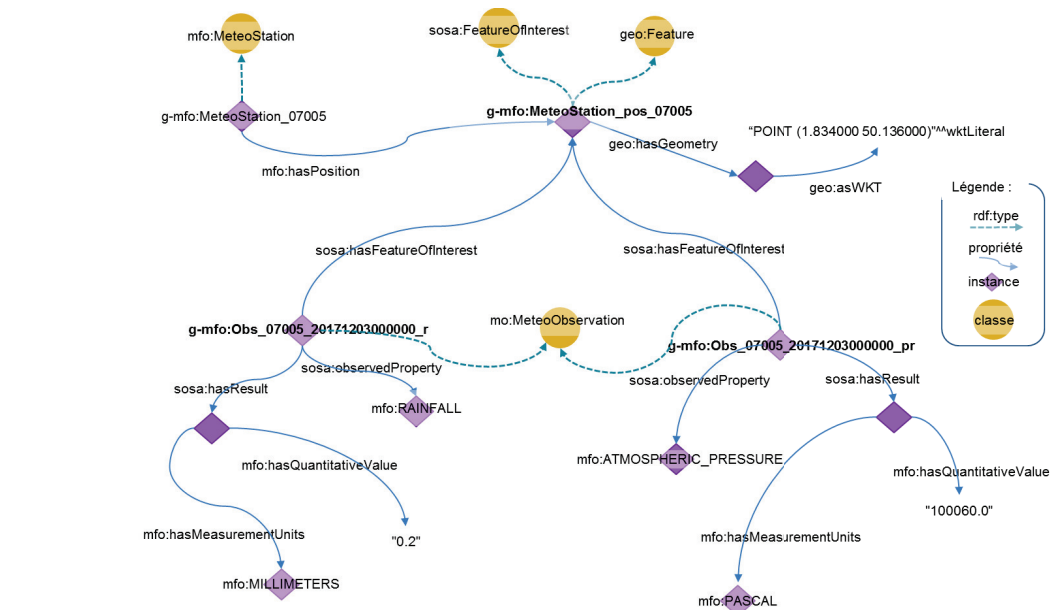


FIGURE 5 – *Instanciation du sous-modèle "mfo" : le "feature of interest" des observations d'une station météo est le lieu où se trouve la station.*

4.1.2.3 Les stations météorologiques

Les informations sur les stations météo, tirées du fichier fourni par *SYNOP Météo France*, fournissent entre autre la position géographique des mesures météo. Le modèle *mfo* (Cf Section 4.1.1) a ainsi été complété notamment avec la propriété *hasPosition* dont le codomaine est une spécialisation de *geo:Feature*. Le code suivant est un extrait de la représentation RDF de la station météorologique d’Embrun :

```

g-mfo:MS_07591      rdf:type          mfo:MeteoStation .
g-mfo:MS_07591      sosa:hosts        g-mfo:Sensor_07591_humidity .
g-mfo:MS_07591      sosa:hosts        g-mfo:Sensor_07591_temperature .
g-mfo:MS_07591      mfo:hasId         "07591"^^xsd:String .
g-mfo:MS_07591      mfo:hasName        "EMBRUN"^^xsd:String .
g-mfo:MS_07591      mfo:hasPosition   g-mfo:MS_pos_07591 .
g-mfo:MS_pos_07591  rdf:type          geo:Feature .
g-mfo:MS_pos_07591  geo:hasGeometry   g-mfo:MS_geo_07591 .
g-mfo:MS_geo_07591  rdf:type          geo:Geometry .
g-mfo:MS_geo_07591  geo:asWKT         "POINT ((6.502333 44.565667))"^^geo:wktLiteral .

```

La Figure 5 est un graphe d’instances illustrant (partiellement) la représentation d’observations météo avec le vocabulaire *mfo* : 2 observations mesurant respectivement les précipitations et la pression atmosphérique, ont été réalisées par la même la station météo. La figure met en évidence qu’une seule URI (*g-mfo:MeteoStation_pos_07005* sur l’exemple) est utilisée pour représenter à la fois la position de la station (propriété *mfo:hasPosition*) et le lieu de chacune des observations (propriété *sosa:hasFeatureOfInterest*).

4.1.2.4 Les unités administratives

Afin de lier les observations de la Terre à des entités administratives (villes, départements, régions, etc.) françaises à partir de leur position géographique (point ou polygone), nous avons enrichi le modèle avec le module *admin* (Cf. Figure 3). La classe *AdministrativeUnit* est caractérisée par une propriété *hasSpatialRepresentation* dont le co-domaine est une sous-classe de *geo:Feature*.

Le modèle a été instancié avec des données provenant de la plate-forme ouverte des données publiques française « data.gouv.fr »¹⁶. Ces données étaient initialement fournies au

16. <https://www.data.gouv.fr>

format shapefile. Les entités administratives de notre triplestore (instances de *AdministrativeUnit*) sont alignées via la propriété *owl:sameAs* avec les territoires français de l'ontologie géographique de INSEE¹⁷; cette ontologie ne contenant pas d'information géométrique, il n'était pas possible de l'utiliser pour établir des relations spatiales. Ci-après un exemple de graphe RDF, celui de la région 54 :

```
admin:region_54 a admin:AdministrativeUnit .
admin:region_54 a admin:Region .
admin:region_54 owl:sameAs <http://id.insee.fr/geo/region/54> .
admin:region_54 admin:hasInseeCode "54"^^xsd:String .
admin:region_54 admin:hasName "Poitou-Charentes"^^xsd:String .
admin:region_54 a geo:Feature .
admin:region_54 geo:hasGeometry 1_admin:region_54_geo .
admin:region_54_geo geo:asWKT "MULTIPOLYGON(((
-1.0988062299633785 45.64032288975508, ...
-1.0988062299633785 45.64032288975508)))"^^wkt:Literal .
```

5 Intégration de données RDF via leurs relations spatiales

Chacun des modules des Figures 2 et 3 est décrit dans un fichier OWL spécifique. Le modèle global est ainsi composé de 4 ontologies¹⁸. Le calcul des relations spatiales et temporelles permet de relier les données de chacun de ces entrepôts. En liant des éléments spatiaux aux tuiles de l'ESA, les connaissances s'y rapportant sont extrapolées au niveau des images. Par exemple, sachant que les images $[img1, img2, img3]$ partagent la tuile $tile_1$, et que cette tuile recouvre $adminUnit_i$, il est possible d'inférer que $[img1, img2, img3]$ recouvrent aussi $adminUnit_i$.

Pour calculer une relation spatiale R (appartenance ou non-intersection, par exemple) entre deux ensembles de données S et T d'une base de connaissances, nous comparons chaque entité s de S à chaque entité t de T , pour vérifier l'existence de $R(s, t)$. Ceci est réalisé à l'aide d'une requête comme celle ci-dessous qui recherche l'ensemble des triplets $\langle s, \text{geo:sfIntersect}, t \rangle$ tels que les géométries (les empreintes géographiques) de s et t ont une intersection non vide :

```
CONSTRUCT { ?s geo:sfIntersects ?t . }
WHERE {
  GRAPH <S> { ?s geo:hasGeometry/geo:asWKT ?s_geometry . }
  GRAPH <T> { ?t geo:hasGeometry/geo:asWKT ?t_geometry . }
  FILTER( geo:sfIntersects( ?s_geometry , ?t_geometry ) )
}
```

Cette approche, consistant à faire le produit cartésien de deux ensembles, est de complexité quadratique si bien que, dès que les jeux de données à combiner sont volumineux, calculer ces relations au moment de l'interrogation, peut être extrêmement coûteux en temps et inacceptable pour les applications temps-réel. C'est pourquoi il est préférable de les pré-calculer. Nous avons distingué le cas des données à composante spatiale seule (statiques) de celles ayant en plus une composante temporelle (dynamiques).

5.1 Intégration des données statiques

Pour les données statiques, il est raisonnable d'envisager d'enregistrer dans l'entrepôt RDF les relations spatiales entre les données RDF correspondantes, si ces jeux de données ont une taille raisonnable. Si le volume est trop conséquent, il est nécessaire d'avoir recours à des techniques d'optimisation. Ainsi, en nous appuyant sur l'indexation spatiale fournie par le tuilage des images S2ST, nous avons calculé les relations entre les données de chacun des jeux

17. <http://rdf.insee.fr/def/index.html>

18. <http://melodi.irit.fr/ontologies/eom.owl> <http://melodi.irit.fr/ontologies/mfo.owl> <http://melodi.irit.fr/ontologies/grid.owl> <http://melodi.irit.fr/ontologies/administrativeUnits.owl>

que nous venons de présenter : stations météo de *mfo*, unités administratives de *admin* et tuiles de *grid*. Ainsi, il est possible de lier les stations météorologiques aux images en calculant uniquement les relations spatiales entre les stations météo et les tuiles de l'ESA. La requête SPARQL ci-dessous interroge les jeux de données instanciant les modèles *eom*, *grid* et *admin*. Elle retrouve les images S2ST avec une couverture nuageuse inférieure à 10%, collectées à un moment donné, et dont l'empreinte géographique couvre (propriété *geo:sfContains*) une zone géographique particulière et ayant un type d'occupation du sol particulier. L'utilisateur a défini une date qui, comme la couverture nuageuse et le type d'occupation du sol, est utilisée pour filtrer les données pertinentes (Cf. les deux dernières clauses FILTER).

```
select distinct ?s2st_result ?tileId
{
  ?s2st a md:EarthObservation .
  ?s2st md:featureOfInterest ?fi .
  ?s2st md:result ?s2st_result .
  ?s2st_result md:product ?s2st_product .
  ?s2st_product md:cloudCoverPercentage ?s2st_cloudCover .
  ?s2st md:phenomenonTime ?s2st_period .
  ?s2st_period time:hasBeginning ?s2st_time .
  ?fi md:isTile ?tile .
  ?tile geo:hasGeometry ?tile_geo .
  ?tile_geo geo:sfContains ?admin_geo .
  ?adminUnit admin:hasSpatialRepresentation ?admin_sr .
  ?admin_sr geo:hasGeometry ?admin_geo .
  ?adminUnit admin:hasName ?adminUnitName .
  ?adminUnitName bif:contains 'Alpes' .
  ?tile grd:hasLandCoverInfo ?tile_lc .
  ?tile_lc grd:hasLandCoverPercentage ?lc_perc .
  ?lc_perc grd:about ?lc_class .
  ?lc_perc grd:percentage_value ?lc_value .
  ?tile grd:tileId ?tileId .
  FILTER (?s2st_time = ?user_time)
  FILTER (?lc_value>1 and ?lc_class=grd:ArtificialSurface
          and ?s2st_cloudCover<10)
}
```

Nous avons testé deux méthodes pour calculer les relations spatiales et créer les triplets RDF les représentant. Nous venons d'exposer la première qui met en oeuvre des requêtes SPARQL incluant des fonctions GeoSPARQL. La seconde a consisté à développer un programme en Python, à l'aide du module Shapely qui permet de faire des opérations spatiales sur des données au format WKT. Avec la solution Python, il faut 0,50s (sans optimisation particulière) pour calculer toutes les relations spatiales entre 1 élément d'un jeu de données *S* et 50 éléments d'un jeu de données *T*, alors qu'il faut 11s avec GeoSPARQL pour calculer une seule relation entre 1 élément de *S* et 50 éléments de *T*.

5.2 Intégration des données ayant une composante temporelle

Il est possible d'établir des relations temporelles entre un enregistrement de méta-données d'image et des données ayant une indication temporelle comme les relevés météo, ou des périodes d'intérêt définies par l'utilisateur. A défaut de disposer d'entités qui servent de référentiel temporel (le pendant des tuiles de l'ESA pour la composante spatiale), nous exploitons l'intervalle de temps défini par un utilisateur lors de la recherche d'images, comme un buffer temporel fournissant un contexte aux enregistrements de méta-données.

La Figure 6 illustre ce principe. Une interface permet à l'utilisateur de dessiner un rectangle pour spécifier des contraintes 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 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). Les images recherchées sont des images Sentinel 2 "classiques", qui ne sont pas liées à une unique tuile comme les S2ST, et donc pour lesquelles

les relations spatiales n'ont pas été pré-calculées. La requête SPARQL ci-dessous sélectionne des enregistrements de méta-données d'images dont les mesures météorologiques associées respectent deux contraintes :

1. elles proviennent de stations situées dans l'empreinte de l'image : ?img_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 fonction `geof:sfContains` dans le filtre `FILTER(geof:sfContains(?img_geo,?ws_geo))` ;
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`.

```
SELECT ?img ?obs_time ?humidityDataVal ?humidityUnits
WHERE{
  ?img a eom:EarthObservation .
  ?img eom:featureOfInterest ?img_foi .
  ?img_foi geo:hasGeometry ?img_geo .
  ?img eom:phenomenonTime ?img_temp .
  ?img_temp time:hasEnd/time:inXSDDateTime ?img_end .
  ?ws a mfo:MeteoStation .
  ?ws mfo:hasPosition ?ws_pos .
  ?ws_pos geo:hasGeometry ?ws_geo .
  ?ws sosa:hosts ?sensor .
  ?sensor sosa:observes mfo:HUMIDITY .
  ?obs sosa:madeBySensor ?sensor .
  ?obs sosa:hasResult ?humidity .
  ?humidity mfo:hasQuantitativeValue ?humidityDataVal .
  ?humidityVal mfo:isClassifiedBy ?humidityUnits .
  ?obs sosa:phenomenonTime ?obs_time .
  ?obs_time time:hasEnd/time:inXSDDateTime ?obs_end .
  BIND (?img_temp-?obs_time as ?diffDateTime)
  FILTER (geof:sfContains(?img_geo,?ws_geo))
  FILTER (?diffDateTime<'PT168H'^^xsd:dayTimeDuration)
  FILTER (?diffDateTime>'PT0H'^^xsd:dayTimeDuration)
}
```

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

La Figure 6 montre l'information météorologique associée à une image faisant partie de la liste retournée par une requête de ce type. 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 choisie.

6 Conclusion

Nous avons proposé une ontologie pour l'intégration de données d'observation de la Terre et données contextuelles, basée sur des relations topologiques spatiales et temporelles. Cette ontologie spécialise des standards, notamment SOSA, GeoSPARQL et OWL-Time. Nous avons défini également un processus d'intégration qui sélectionne, transforme et intègre de données géospatiales hétérogènes (méta-données d'image satellite, données météorologiques, unités administratives, couverture terrestre, etc.). Ce processus s'appuie sur le tuilage des images pour traiter les données ayant une composante spatiale fixe, les relations temporelles, quant à elle, sont calculées à la volée à partir d'une topologie temporelle. Nous avons présenté un cas d'étude basé exploitant des méta-données d'image satellite Sentinel.

Dans la continuité de ces travaux, nous envisageons de considérer des sources propres à un domaine métier pour traiter un cas d'usage (l'agriculture et des rapports bulletins agricoles) et fournir des règles et des fonctionnalités de raisonnement pour faciliter les analyses. Il serait intéressant aussi 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

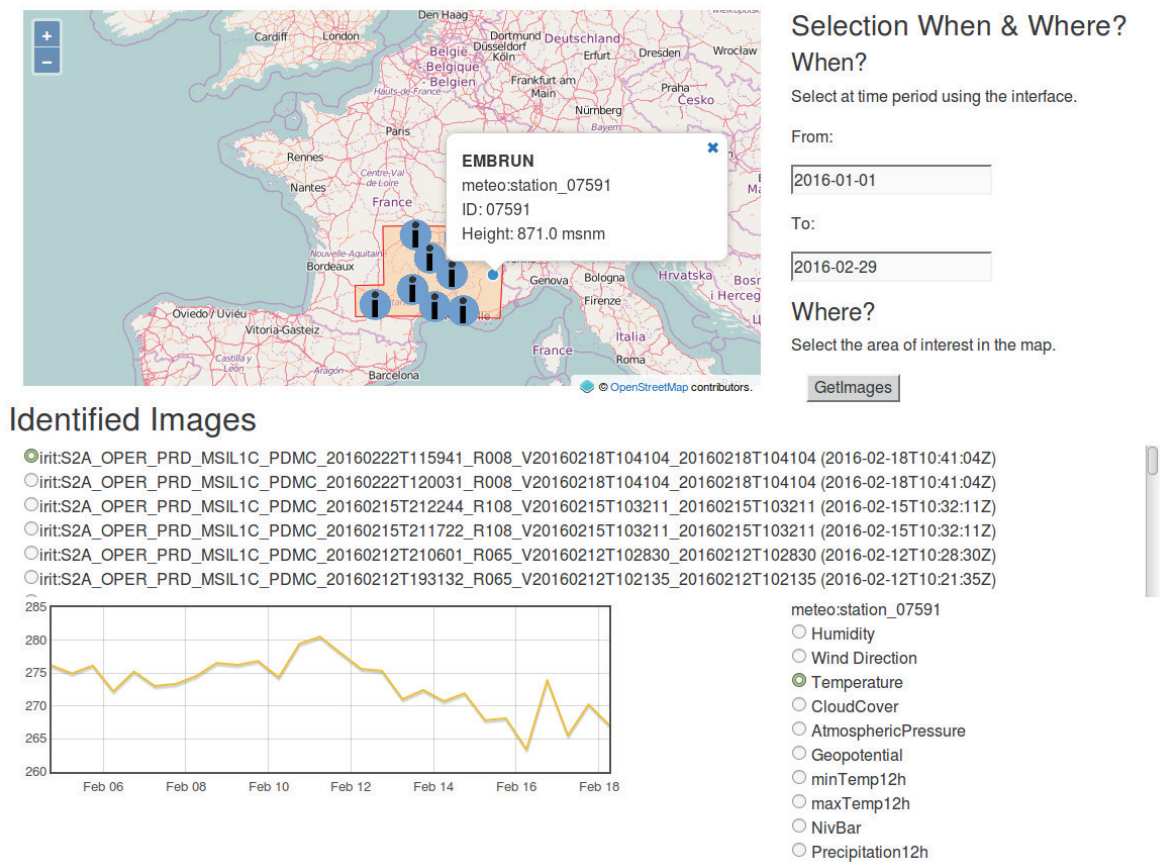


FIGURE 6 – Affichage des informations météorologiques associées à une image satellite en réponse à la requête.

temporelles de température et détecter une canicule. Nous envisageons également 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.

Références

- ARENAS H., AUSSÉNAC-GILLES N., COMPAROT C. & TROJAHN C. (2016a). Semantic integration of geospatial data from earth observations. In *Knowledge Engineering and Knowledge Management - EKAW 2016 Satellite Events*, p. 97–100.
- ARENAS H., AUSSÉNAC-GILLES N., COMPAROT C. & TROJAHN C. (2016b). Un modèle pour l'intégration sémantique de données géolocalisées liées à l'observation de la terre. In *Spatial Analysis and Geomatics - Atelier EXtraction de Connaissances à partir de données Spatialisées (SAGEO-EXCES)*, Nice.
- ATEMEZING G. A. (2015). *Publishing and consuming geo-spatial and government data on the semantic web*. PhD thesis, Thesis.
- BATTLE R. & KOLAS D. (2012). Enabling the Geospatial Semantic Web with Parliament and GeoSPARQL. *Semantic Web*, 3(October 2012), 355–370.
- BERETA K., SMEROS P. & KOUBARAKIS M. (2013). Representation and querying of valid time of triples in linked geospatial data. In *The Semantic Web: Semantics and Big Data*, p. 259–274, Berlin, Heidelberg: Springer Berlin Heidelberg.
- BLÁZQUEZ L. M. V., VILLAZÓN-TERRAZAS B., CORCHO Ó. & GÓMEZ-PÉREZ A. (2014). Integrating geographical information in the linked digital earth. *Int. Journal of Digital Earth*, 7(7), 554–575.

- BRIZHINEV D., TOYER S., TAYLOR K. & ZHANG Z. (2017). *Publishing and Using Earth Observation Data with the RDF Data Cube and the Discrete Global Grid System*. Rapport interne, W3C and OGC.
- CONSOLE M. & LENZERINI M. (2014). Reducing global consistency to local consistency in ontology-based data access - extended abstract. In *Informal Proceedings of the 27th International Workshop on Description Logics, Vienna, Austria, July 17-20, 2014.*, p. 496–499.
- COX S. (2017). Prov ontology supports alignment of observational data (models). In *Modeling and Simulation Society of Australia and New Zealand*, p. 403–409.
- ESPINOZA-MOLINA D. & DATCU M. (2013). Earth-observation image retrieval based on content, semantics, and metadata. *IEEE Trans. on Geoscience and Remote Sensing*, **51**(11), 5145–5159.
- ESPINOZA-MOLINA D., NIKOLAOU C., DUMITRU C. O., BERETA K., KOUBARAKIS M., SCHWARZ G. & DATCU M. (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*, **8**(4), 1696 – 1708.
- GASPERI J. (2014). Semantic Search Within Earth Observation Products Database Based on Automatic Tagging of Image Content. In *Proc. of the Conf. on Big Data from Space*, p. 4–6.
- GORE A. (1998). The digital earth. *Australian Surveyor*, **43**(2), 89–91.
- HARBELOT B., ARENAS H. & CRUZ C. (2014). un modèle sémantique spatio-temporel pour capturer la dynamique des environnements. In *14 ème conférence Extraction et Gestion des Connaissances*, p. 39 à 54, Rennes, France.
- HEATH T. & BIZER C. (2011). *Linked Data: Evolving the Web into a Global Data Space ; Lectures on the Semantic Web: Theory and Technology*. Morgan & Claypool.
- 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*, **35**, 25–34.
- KOLAS D., PERRY M. & HERRING J. (2013). *Getting started with GeoSPARQL*. Rapport interne, OGC.
- KOUBARAKIS M. & KYZIRAKOS K. (2010). Modeling and querying metadata in the semantic sensor web: The model strdf and the query language stsparql. In *The Semantic Web: Research and Applications*, p. 425–439, Berlin, Heidelberg: Springer Berlin Heidelberg.
- LEFORT L., BOBRUK J., HALLER A., TAYLOR K. & WOOLF A. (2012). A linked sensor data cube for a 100 year homogenised daily temperature dataset. In *Proc. of the 5th International Conference on Semantic Sensor Networks - Volume 904, SSN'12*, p. 1–16: CEUR-WS.org.
- LEFRANÇOIS M., ZIMMERMANN A. & BAKERALLY N. (2017). A SPARQL extension for generating RDF from heterogeneous formats. In *Proc. Extended Semantic Web Conference (ESWC'17)*, Portoroz, Slovenia.
- LENZERINI M. (2011). Ontology-based data management. In *Proc. of the 20th ACM Int. Conference on Information and Knowledge Management, CIKM '11*, p. 5–6, New York, USA: ACM.
- MAALI F. & ERICKSON J. (2014). Data Catalog Vocabulary (DCAT).
- PERRY M. & HERRING J. (2012). *OGC GeoSPARQL-A geographic query language for RDF data*. Rapport interne, Open Geospatial Consortium.
- RASKIN R. (2006). *Guide to SWEET Ontologies*. Rapport interne 9, NASA/Jet Propulsion Lab, Pasadena, CA, USA.
- REITSMA F. & ALBRECHT J. (2005). Modeling with the semantic web in the geosciences. *IEEE Intelligent Systems*, **20**(2), 86–88.
- SUKHOBOK D., SÁNCHEZ H., ESTRADA J. & ROMAN D. (2017). Linked data for common agriculture policy: Enabling semantic querying over sentinel-2 and lidar data. In *Proceedings of the ISWC 2017 Posters & Demonstrations and Industry Tracks co-located with 16th Int. Semantic Web Conference, Vienna, Austria, Oct. 23-25, 2017*.
- TANDY J., VAN DEN BRINK L. & BARNAGHI P. (2017). *Spatial Data on the Web Best Practices, W3C Working Group Note*. Rapport interne, W3C and OGC.
- TELEIOS (2016). Virtual Observatory Infrastructure for Earth Observation Data. <http://www.earthobservatory.eu/search/node/ontologies>. Accessed: 2016-03-01.
- 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.