



HAL
open science

Une approche sémantique pour représenter l'indice de végétation d'images Sentinel-2 et son évolution

Jordane Dorne, Nathalie Aussenac-Gilles, Catherine Comparot, Romain Hugues, Jean-Guy Planès, Cassia Trojahn dos Santos

► To cite this version:

Jordane Dorne, Nathalie Aussenac-Gilles, Catherine Comparot, Romain Hugues, Jean-Guy Planès, et al.. Une approche sémantique pour représenter l'indice de végétation d'images Sentinel-2 et son évolution. Spatial Analytics and GEomatics (SAGEO 2018), Nov 2018, Montpellier, France. pp.49-54. hal-02181917

HAL Id: hal-02181917

<https://hal.science/hal-02181917>

Submitted on 12 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Open Archive Toulouse Archive Ouverte

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

This is an author's version published in:

<http://oatao.univ-toulouse.fr/22586>

Official URL

<https://agritrop.cirad.fr/589351/>

To cite this version: Dorne, Jordane and Aussenac-Gilles, Nathalie and Comparot, Catherine and Hugues, Romain and Planès, Jean-Guy and Trojahn, Cassia *Une approche sémantique pour représenter l'indice de végétation d'images Sentinel-2 et son évolution*. (2018) In: Spatial Analytics and GEomatics (SAGEO 2018), 6 November 2018 - 9 November 2018 (Montpellier, France).

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

Une approche sémantique pour représenter l'indice de végétation d'images Sentinel-2 et son évolution

Jordane Dorne^{1,2}, Nathalie Aussenac-Gilles¹,
Catherine Comparot¹, Romain Hugues²,
Jean-Guy Planes², Cassia Trojahn¹

¹*Institut de Recherche en Informatique de Toulouse, Toulouse, France*

²*Thales Alenia Space, Toulouse, France*

{prenom.nom}@irit.fr, {prenom.nom}@thalesaleniaspace.com

RÉSUMÉ. Cet article propose une représentation sémantique (en RDF) d'indices de végétation calculés sur des images Sentinel-2. Cette représentation s'appuie sur une ontologie qui intègre la notion de tuile définie par l'ESA, i.e. une zone fixe de la surface terrestre. A chaque tuile, nous associons des indices de végétation calculés à partir des images dont l'emprise au sol correspond. Nous présentons cette ontologie et montrons comment ces représentations permettent d'étudier l'évolution de ces indices à partir de requêtes sur les données RDF.

ABSTRACT. In this paper we propose an RDF semantic representation of vegetation indices calculated from Sentinel-2 images. This representation relies on an ontology where the notion of tile, as defined by ESA, represents a fixed area on the Earth surface. Each tile is associated vegetation indices calculated from linked images. We present this ontology and illustrate how these representations can support the study of the evolution of these indices over time, by querying the RDF dataset.

MOTS-CLÉS : modèle sémantique, images satellites, indice de végétation, évolution

KEYWORDS: semantic model, satellite images, vegetation index, evolution

1. Introduction

Pour collecter des données civiles d'observation de la Terre, le programme Européen Copernicus lancé en 1998 par l'Agence Spatiale Européenne (ESA) a développé une série de satellites appelés Sentinel, notamment :

- Sentinel-1 : fourniture d'imagerie radar tout-temps, jour et nuit, à des fins d'observation du sol et des océans,

– Sentinel-2 : fourniture d'imagerie optique haute résolution pour l'observation des sols (utilisation des sols, végétation, zones côtières, fleuves, etc.).

La grande majorité des données du programme Copernicus sont disponibles et accessibles gratuitement par tous¹. Leur disponibilité ainsi que les nombreuses autres sources de données géolocalisées accessibles via le web ouvre de nombreuses perspectives économiques grâce à de nouvelles applications dans des domaines variés. Ces applications peuvent en effet bénéficier à la fois des méta-données des images (telles que la couverture nuageuse), d'une analyse automatique du contenu des images (par exemple pour y repérer la végétation, l'urbanisme ou d'autres objets d'intérêt) et de données ouvertes géo-localisées et datées pouvant être "situées" sur ces images (données gouvernementales, météorologiques, etc.). Pour associer aux images ces trois types de données (fournies et extraites des images ou collectées sur le web), 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 (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. Ce type d'approche s'appuie sur une ou plusieurs ontologies pour représenter les données, les relier entre elles et à l'image, et cela à un bon niveau d'abstraction. Deux problèmes se posent alors : définir une ontologie adaptée et gérer le grand volume de données ainsi générées.

Nous nous intéressons à la représentation sémantique de données extraites des images qui sont donc géo-localisées et qui évoluent dans le temps. Nous utilisons une ontologie réutilisant plusieurs vocabulaires du *Linked Open Data* reconnus comme des standards pour faciliter la réutilisation des données ainsi représentées, dont GeoSPARQL (Kolas *et al.*, 2013) pour la dimension spatiale, SOSA² pour représenter une partie des méta-données d'images satellites considérées ici comme données de capteurs, et OWL-Time³ pour la dimension temporelle. Nous supposons la surface terrestre quadrillée selon une grille formée de tuiles, telle que celle fournie par l'ESA pour les images Sentinel 2 Single-Title (S2ST).

Dans cet article, nous proposons une approche pour représenter sémantiquement des indices de végétation d'images Sentinel-2. Nous représentons la notion de tuile S2ST pour associer les données directement aux tuiles. Les indices de végétation calculés à partir de ces images peuvent donc être localisés sur une tuile et datés de la date de l'image. Nous montrons comment la repré-

1. <http://copernicus.eu/data-access>

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

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

sentation sémantique de ces données et leur interrogation permet d'étudier leur évolution, et ainsi l'évolution de la végétation au sol.

2. Ontologie proposée

L'ontologie proposée reprend celle présentée dans (Arenas *et al.*, 2018), définie pour intégrer des données géolocalisées (i.e. possédant une "géométrie") et datées autour d'images satellites. L'ontologie comporte plusieurs modules, chacun correspondant à un vocabulaire ayant un espace de nom spécifique. Le nom du module est exploité dans ce qui suit pour préfixer le vocabulaire qu'il définit. Nous présentons dans un premier temps les modules de l'ontologie réutilisée (Arenas *et al.*, 2018), c'est-à-dire les modules *sosa*, *geo*, *grid*, *eom*, puis dans un deuxième temps celui que nous avons ajouté, le module *ndvi*.

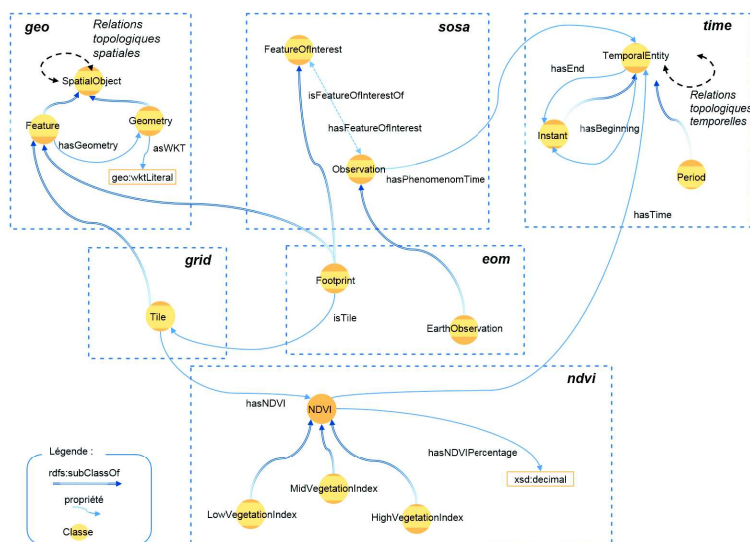


FIGURE 1. Ontologie modulaire proposée.

Le module *sosa* permet de représenter toute observation (*sosa:Observation*) réalisée par un capteur sur une "entité" (*sosa:FeatureOfInterest*), comme une activité datée (*sosa:phenomenonTime*). Le module *geo* sert à représenter la dimension spatiale des données à intégrer et réutilise GeoSPARQL. Il en reprend la classe *geo:Feature* pour représenter toute entité ou donnée géo-localisée à laquelle est associée une *geo:Geometry*, et des relations spatiales (*se trouve dans*, *recoupe*, etc.) entre les géométries. Nous utilisons ces relations pour lier deux ressources géo-localisées. Le module *grid* décrit les tuiles de l'ESA. Chaque tuile est représentée comme une instance de la classe *grid:Tile*, spécialisation de *geo:Feature* dont la propriété *geo:hasGeometry* fournit l'emprise au sol sous forme d'un polygone fermé - un carré - défini par des coordonnées GPS. Le module *eom* sert à représenter les méta-données d'images (module simplifié à deux classes sur la Figure 1). La propriété *eom:isTile* permet d'associer

une tuile au `eom:Footprint` de chacune des images S2ST lui correspondant : un footprint représente la zone géographique couverte par une image et est décrit par un polygone fermé (une géométrie). Ce footprint est considéré par ailleurs comme le point d'intérêt d'observations réalisées par des capteurs du satellite ayant pris l'image ; la classe `eom:Footprint` spécialise ainsi `geo:Feature` et `sosa:FeatureOfInterest`. Le `eom:Footprint` d'une image est ainsi daté via la propriété `sosa:phenomenonTime` qui identifie le moment où l'image a été prise.

Adapter l'ontologie à la représentation d'un nouveau type de données (comme les indices de végétation) revient à ajouter à ces modules un nouveau module pour ces données, le module `ndvi`, et à le lier aux modules existants. Chaque cadre de la Figure 1 présente un extrait du vocabulaire d'un module exploité pour associer les indices NDVI aux images S2ST. Au sein du module `ndvi`, l'indice de végétation est représenté par la classe `ndvi:NDVI`. Les propriétés spatiale et temporelle des indices de végétation sont fournies respectivement par la tuile à laquelle ils sont associés et la relation `time:hasTime` (OWL Time). Cette dernière permet d'exprimer des relations temporelles (*avant*, *après*, etc.) entre deux entités temporelles.

3. Cas d'étude

L'indice de végétation NDVI (*Normalized Difference Vegetation Index*) est calculé à l'aide des images S2ST de niveau L1C. Nous avons choisi les images comportant une couverture nuageuse inférieure à 3% afin de ne pas fausser les résultats. Cet indice est obtenu par un calcul sur les bandes "proche infrarouge" (PIR) et "rouge" (R) du capteur, qui permet de détecter le niveau de chlorophylle. Le résultat de cette opération donne une matrice de valeurs comprises entre -1 et 1 caractérisant le NDVI de chaque pixel de l'image. Les valeurs comprises entre -1 et 0 représentent les éléments composés d'eau, les valeurs comprises entre 0 et 0.25 représentent les éléments composés de terre. Nous nous intéressons ici uniquement aux valeurs concernant la végétation, c'est à dire aux valeurs comprises entre 0.25 et 1. Plus l'indice est élevé plus la végétation est dense. Pour associer un indice à chaque pixel d'une image, nous avons classé les indices en trois catégories : *LowVegetationIndex* (NDVI compris entre 0.25 et 0.5), *MidVegetationIndex* (NDVI compris entre 0.5 et 0.75), et *HighVegetationIndex* (NDVI compris entre 0.75 et 1), correspondant chacune approximativement à de l'herbe, de la culture, et de la forêt.

Pour obtenir une représentation de ces catégories en pourcentage d'image il faut calculer le nombre de pixels de chaque catégorie par rapport au nombre total de pixels de l'image. Nous avons représenté ces trois catégories dans le module `ndvi` sous forme de classes (Cf. Figure 1). Les indices de végétation calculés, instances d'une de ces classes, sont datés à partir de la date de prise de l'image et géo-localisés avec la tuile associée à l'image. Ci-après un exemple de graphe RDF représentant, selon notre modèle, l'indice de végétation Low (instance de `ndvi:LowVegetationIndex`) de la tuile 31TCJ au 21 avril 2018.

```

g-ndvi:ndvi_31TCJ_20180421T105031_low a ndvi:LowVegetationIndex .
g-grid:31TCJ a grid:Tile .
g-grid:31TCJ ndvi:hasNdvi g-ndvi:ndvi_31TCJ_20180421T105031_low .
g-ndvi:ndvi_31TCJ_20180421T105031_low
  ndvi:hasNdviPercentage "15.11"^^xsd:decimal .
g-ndvi:ndvi_31TCJ_20180421T105031_low
  time:hasTime ndvi:instant_1524300631 .
ndvi:instant_1524300631 a time:Instant .
ndvi:instant_1524300631
  time:inXSDDateTime "2018-04-21T10:50:31.026000"^^xsd:dateTime .
ndvi:instant_1524300631
  time:inXSDDateTimeStamp "1524300631"^^xsd:dateTimeStamp .

```

Un graphe de ce type est généré pour chaque catégorie d'indice d'une image à partir d'un script Python qui prend en entrée le fichier jpeg2000 de l'image, calcule les NDVI correspondants et peuple l'ontologie. Sur la Figure 2, le résultat d'une requête SPARQL indique l'évolution du NDVI de la tuile 31TCJ sur une année (filtrage sur la composante temporelle des NDVI). Grâce à ces valeurs, nous pouvons observer l'évolution de la végétation en fonction des saisons. Sur cet exemple, nous constatons au début du printemps (entre fin février et fin mai) une augmentation du pourcentage de la catégorie *HighVegetationIndex*, correspondant à la croissance du feuillage de forêts ; à l'inverse, nous notons une diminution de ce pourcentage pendant la période hivernale. Les catégories *MidVegetationIndex* et *LowVegetationIndex* sont aussi impactées par le changement de saisons.

time	tileId	ndviHighPercent	ndviMidPercent	ndviLowPercent
2017-05-26T10:50:31.026	http://melodi.irit.fr/lod/grid_S2ST/31TCJ	18.22	47.09	16.91
2017-07-05T10:50:31.026	http://melodi.irit.fr/lod/grid_S2ST/31TCJ	16.11	35.22	23.81
2017-08-14T10:50:31.026	http://melodi.irit.fr/lod/grid_S2ST/31TCJ	8.19	36.38	31.73
2017-10-08T10:50:09.027	http://melodi.irit.fr/lod/grid_S2ST/31TCJ	2.66	37.25	23.83
2017-10-13T10:50:31.026	http://melodi.irit.fr/lod/grid_S2ST/31TCJ	0.45	35	26.2
2017-10-28T10:51:29.027	http://melodi.irit.fr/lod/grid_S2ST/31TCJ	0.38	31.12	26.38
2017-11-07T10:52:29.027	http://melodi.irit.fr/lod/grid_S2ST/31TCJ	0.33	29.3	28.77
2017-11-22T10:53:41.026	http://melodi.irit.fr/lod/grid_S2ST/31TCJ	0.16	14.7	38.44
2017-11-27T10:53:59.027	http://melodi.irit.fr/lod/grid_S2ST/31TCJ	0.23	10.77	43.09
2018-02-10T10:52:01.026	http://melodi.irit.fr/lod/grid_S2ST/31TCJ	0.44	18.17	55.53
2018-02-25T10:50:19.027	http://melodi.irit.fr/lod/grid_S2ST/31TCJ	0.24	15.84	58.84
2018-04-21T10:50:31.026	http://melodi.irit.fr/lod/grid_S2ST/31TCJ	12.48	54.44	15.11

FIGURE 2. Résultat SPARQL du NDVI sur une année pour la tuile 31TCJ.

4. Conclusion

Dans cet article, nous avons proposé de représenter via une ontologie les indices de végétation des tuiles Sentinel-2 calculés à partir des images qui leur sont liées. L'ontologie sert aussi à établir des relations temporelles permettant l'analyse de l'évolution des indices. Cela peut être utile aux analyses en agriculture. Ce travail complète celui que nous avons mené pour l'intégration de données contextuelles aux métadonnées d'images via des relations spatio-temporelles (Arenas *et al.*, 2018). Il s'agit d'une première étape en vue d'identifier à terme un changement anormal. Nous envisageons de fournir des règles et des fonctionnalités de raisonnement pour faciliter les analyses.

Remerciements

Ce travail est co-financé par le projet CANDELA (convention de subvention num 776193 du programme de recherche et d'innovation Horizon 2020 de l'Union européenne), et par la convention Cifre ANR n° 2017/1399 entre TAS et le CNRS.

Bibliographie

- Arenas H., Aussenac-Gilles N., Comparot C., Trojahn C. (2018). Ontologie pour l'intégration de données d'observation de la Terre et contextuelles basée sur les relations topologiques. In *29es Journées Francophones d'Ingénierie des Connaissances*, p. 5-20. Nancy, France, AFIA.
- 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*, p. 496-499. Vienna, Austria.
- 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
- Lefrançois M., Zimmermann A., Bakerally N. (2017, mai). A SPARQL extension for generating RDF from heterogeneous formats. In E. Blomqvist, D. Maynard, A. Gangemi, R. Hoekstra, P. Hitzler, O. Hartig (Eds.), *Proc. of the 2017 Extended Semantic Web Conference, Part I*, p. 35-50. Portoroz, Slovenia, Springer.
- Lenzerini M. (2011). Ontology-based data management. In *Proceedings of the 20th acm international conference on information and knowledge management (cikm '11)*, p. 5-6. New York, NY, USA, ACM.
- Reitsma F., Albrecht J. (2005). Modeling with the semantic web in the geosciences. *IEEE Intelligent Systems*, vol. 20, n° 2, p. 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 N. Nikitina, D. Song, A. Fokoue, P. Haase (Eds.), *Iswc 2017 posters, demonstrations and industry tracks (iswc-pd-industry)*. Aachen. Consulté sur <http://ceur-ws.org/Vol-1963/#paper559>