A DCAP to Promote Easy-to-Use Data for Multiresolution and Multitemporal Satellite Imagery Analysis

Isabelle Mougenot
Espace Dev UM,
France
isabelle.mougenot
@univmontp2.fr

Jean-Christophe Desconnets
Espace Dev IRD,
France
jeanchristophe.desconnets
@ird.fr

Hatim Chahdi
Espace Dev IRD & LIPN
UMR CNRS,
France
hatim.chahdi@ird.fr

Abstract

Satellite imagery can be exploited for any number of thematic analyses for Earth observation purposes. Characterization activities using remotely acquired data are currently made complicated by different limitations relating to, as an example, the meaningful mapping between multi-sensor data or the adding of the geospatial context to satellite information. We argue that describing satellite images through a metadata application profile may leverage capabilities to promote easy-to-use data for further in-depth thematic analysis. Accordingly, an application profile conforming to the Dublin Core application profile (DCAP) guidelines and designed for Earth observations (EO) is being developed. More specifically, we discuss RDF-compliant machine-processable aspects of the EO application profile (EOAP) in terms of the DCMI Description Set Profile (DSP) model. Additionally, a methodological approach to represent a DSP model using UML profiling activities is proposed.

Keywords: metadata; Dublin Core application profile; Earth observation, satellite imagery; semantic web standards; UML metamodeling

1. Introduction and motivations

Earth observing satellite imagery provides various datasets at different spatial, spectral and temporal resolutions. Each of these datasets can provide a complementary view that can improve assessments on the observed objects. The technical diversity of satellite sensors as well as their increasing number allows images to be considered at an unprecedented volume of data, richer and precise enough to deliver novel insights, such as a novel understanding of ecosystems dynamic or the monitoring of environmental changes at a local scale. The main objective of our DCAP is to integrate data with different spatial, spectral or temporal characteristics in an appropriate way, to gain more information that can be obtained from each individual sensor.

The increasing number of remotely sensed images as well as their large-scale distribution are the first impediments for data integration. Additionally, images are the results of numerous parameters, from technical characteristics of imaging sensors to atmospheric effects that limit capacities for systematic observations at various levels. Moreover, image-based data and their associated metadata are recorded in numerous file formats, such as GeoTIFF or JPEG 2000 that all have specific ways of describing content-based images.

In this context it is critical to simplify efficient image-based data access and query processing to provide accessibility to a variety of expert and non-expert users in remote sensing. Consequently the main aim is to document image-based data as well as additional data using metadata standards. Produced documents directly allow answering the query without consulting the data itself. For this purpose we have developed a metadata application profile according to the Dublin Core application profile (DCAP) guidelines (Nilsson, 2009). This application profile is named EOAP (Earth Observation Application Profile) (Desconnets, 2014) and is designed to benefit from metadata standards interoperability and linked open data principles for data sharing.
on the web. Moreover, EOAP offers a descriptive framework that is flexible and extensible enough to adapt to numerous environmental uses cases as well as different viewpoints of users. The key objective is to ensure the use of the most comprehensive coverage of precise and accurate image-based data so that any environmental issues can be well addressed.

The current initiative is taken in the context of the GEOSUD research project\(^1\). GEOSUD aims to promote increased access and use of satellite imagery to the French public. In particular the main objective of GEOSUD is to provide various facilities to effectively access shared image based data and processing tools that enable data retrieval, visualization and higher level analysis. GEOSUD was motivated by the lack of use of spatial data to help control natural environments and sustainable resources within land policy-making and institutional settings. Therefore a national spatial data infrastructure (Kazmierski, 2014) was developed to improve access to Earth observation data, particularly that of high and very high resolution satellite data. A suite of geospatial web services offers interoperable access to images provided by different satellite data suppliers as well as some image processing facilities.

The application profile we have developed will be a core component of this infrastructure in the near future and will be used to enable more advanced search analyses of Earth observation data. We therefore describe how EOAP may be used by referencing appropriate user scenarios.

Furthermore, we consider the application profile model for Earth observation as a domain specific language (DSL) [Fowler, 2010] and the constraint language DSP as a metamodel, which is more likely to permit the building of such a language. We draw on RDF and RDFS languages, UML profile and RDF metamodels using Meta-Object Facility (MOF) to build the DSP model for Earth observation.

The manuscript is structured in five sections. Following the introduction, Section 2 describes the diversity of datasets across imaging sensors and shortly introduces appropriate metadata standards that meet the variety of requirements to describe Earth observation data. Section 3 describes the modeling activities at different abstraction levels to work towards building a description set model for Earth observation resources. This model is RDF-compliant and conforms to the constraint language DSP. Section 4 provides illustrations about realistic use cases with specific needs related to multi-temporal and multi-resolution remotely sensed data that cover most significant functional aspects of the application profile under development. Finally, the last section draws preliminary conclusions and provides prospects about decisions made to implement the Earth observation application profile.

2. Background

2.1. Datasets

The satellite images that we want become accessible, via the GEOSUD spatial data infrastructure (SDI), are from different satellite sensors which acquire high and very high resolution images. Each year, these datasets must provide a high-resolution coverage of the entire national territory (5 meters/pixel) and in a second step a very high-resolution coverage (1.5 meters/pixel). In addition, in order to analyze the seasonal functioning of ecosystems and territories, time series with high frequency acquisition from medium resolution sensors, should also be available. Finally, the on-off scheduling of the acquisition of very high-resolution images via a direct receiving antenna is also planned. To meet these goals, several satellite sensors are used: Landsat 8 for medium resolution images, SPOT5 and Rapid Eye for high-resolution images, PLEIADES and SPOT6 for those with very high resolution. Established in 2011, the GEOSUD SDI is expected to acquire, each year, about 600 images, estimated at a volume of 1 to 2 TB / year for high resolution products and about 12 Tbytes / year for the very high resolution products.

\(^1\) http://www.equipex-geosud.fr
\(^2\) Clearcut: refers to a mode of forestry development through cutting down of all trees of a parcel
Depending on the image processing chain, the images are distributed in different formats such as GeoTiff or JPG2000 and, at different processing levels: projected raw images, radiometrically and geometrically corrected images, according to different encoding levels: 8 bits, 12 bits or 16 bits. Depending on the resolution and, the encoding format, an image can reach up to 50 MB for a medium resolution image (Landsat8), 2 GB for a high resolution image in JPG2000 format, even 15 GB for an equivalent image in GeoTiff format. Finally, these images are made accessible, searchable and downloadable via a set of web services and user applications.

2.2. Metadata framework

Different standards may be, general or dedicated to a particular discipline structure metadata. Regarding spatial data, we first want to mention the ISO 19115 (ISO, 2003) standard for geographic information, ISO 19115-2 (ISO, 2009) dedicated to gridded data and the specialization of O & M (Observation and Measurement) specification which proposes, among other things, elements to describe sensors characteristics and acquisition conditions (Gaspéri, 2012).

To better describe the satellite images, we chose among the metadata elements proposed by the various standards mentioned above. Moreover, we used the Dublin Core elements as common core elements such as title, creator, or coverage. The ISO 19115 and 19115-2 standards provide specific descriptors to the inherent spatial dimension to our datasets, such as the description of the spatial reference system associated with the location of the image (ISO19115: MD_ReferenceSystem), or the description of intrinsic characteristics associated with matrix structure of the image (e.g. ISO19115: MD_SpatialRepresentation). ISO 19115 and ISO 19115-2 also provide elements that characterize the sensor used to acquire the image (MI_Platform, MI_Instrument). Finally, the O & M for image description brings elements relating to acquisition conditions which are essential to pre-process the images after their acquisition.

A potential reproach to metadata standards is that they have been designed independently of each other and thus are not able to meet all information needs. This is especially true in our context in which the applications planned around images should cover a broad spectrum of functionalities: discovery, location, consultation, processing and, archiving. In fact, their implementation requires the contribution of individual standard. Based on this background, the definition of an application profile is relevant and able to offer a description framework both constrained but interoperable. The application profile built for satellite imagery (Desconnets, 2014) is based on the Singapore framework and application profile methodology named DCAP (Dublin Core Application Profile). Specifically, our work has focused on the definition of a Description Set Profile, a structural model that complements the DCAM model to provide a prescriptive framework for the construction of an application profile (Nilsson, 2009). Thus, the application profile can be seen as a model that does not prescribe the data of interest, which are the satellite images, but the metadata elements which describe these datasets. The objective is both to reduce time and cost of datasets consultation and facilitate the management of big, heterogeneous and distributed data sources, such as the satellite images are. Their consultation is based on instances of DSP, i.e. metadata sets. In the next section, we will enlarge the implemented approach for the construction of the DSP model. The focus is given to modeling and metamodeling approaches.

2.3. RDF-compliant DSP to maximize reuse

We investigated the potential of using RDF language (Hayes, 2004) to build the DSP model. RDF is a W3C standard for encoding metadata, datasets and vocabularies on the web. Overall, we are giving top priority to release metadata instances in an open format on the web and we are considering this as a far better means of data exchange and sharing within the context of Linked Open Data (LOD) (Warren, 2014). DCMI provides some guidelines for encoding DSP specification in the RDF/XML concrete syntax (CWA15248, 2005). Additionally, many metadata standards including DCTERMS (DCES, 2012) are represented in a RDF serialization format,
e.g. RDF/XML or N3. Similarly controlled vocabularies, as for example Geonames (Ahlers, 2013) or TGN (Getty Thesaurus of Geographic Names), are also available in RDF formats to guarantee open use.

However RDF and RDF Schema (RDFS) formalisms are built on a number of language primitives that are not always in line with the requirements drawn up for the constraint language DSP. In particular, RDF only provides a construct for declaring binary properties. Consequently, representing non-binary relations is a well-known issue, since n-ary relations arise quite commonly during modeling activities. A DSP model is intended to represent the overall structure of a metadata description set by means of constraints that apply either on resources described, properties used or values that may be given with respect to the properties. In this direction a DSP model is built using the notions of description template and statement template that define the valid skeleton of a description and a statement, respectively. A DSP is then a collection of description templates (DescriptionTemplate), which in turn are collections of statement templates (StatementTemplate). At the same time this notion of collection involves three entities, namely DescriptionTemplate, Property and Constraint and reveals a complex constraint that requires a ternary relation. A UML (Unified Modeling Language) class diagram (Rumbaugh, 1991) for specifying such collections is introduced (FIG. 1) and describes the class DescriptionTemplate associated with the classes StatementTemplate, Property and Constraint by means of an n-ary relationship. StatementTemplate is represented as an association class and appears as a class linked to the association with a dashed line. DescriptionTemplate contains a reference to StatementTemplate, which in turn contains a first reference to the class Property and a second reference to the class Constraint. Consequently UML could represent StatementTemplate as an association class, whilst the RDF language may define StatementTemplate as an auxiliary node. In the DSP model, StatementTemplate is represented as an auxiliary node that does not signify a named resource, i.e. a blank node or anonymous resource. In addition, following the same reasoning, StatementTemplate contains a reference to the class Constraint. Constraint is a nested structure that contains a collection of constraints and is also represented as an anonymous resource.

Blank nodes come with a significant overhead and additionally add unnecessary complexity to the DSP model.

As UML leverages the power of modeling effectively, we propose an extra approach to separate the modeling of the DSP into a three level hierarchy: a model at level 1 (Earth observation domain model), a meta-model at level 2 (DSP meta-representation) and a meta-metamodel at level 3 (RDF and RDFS meta-representations). We defend the claim that UML profiling mechanisms could help increase the usefulness of a RDF application profile particularly
in a linked open data context. An UML profile [(D’Souza, 1999)] represents a lightweight extension mechanism to the UML language by defining custom stereotypes in particular. Stereotypes are applied to UML elements to refine their semantics, either as classes or associations.

On that point, we take advantage of the work carried out on ontology metamodeling (Brockmans, 2006) with a corresponding UML profile and a collection of stereotypes that convey the meaning of the semantic web languages primitives (RDF, RDFS and OWL). Some of these stereotypes are illustrated in the simplified diagram of the DSP model depicted in FIG. 2.

A meta-class, as an example RDFSC lass, BlankNode or RDFProperty, refines the semantics of each class of the DSP model. For instance, the generic class BlankNode marks appropriate dependencies on the classes StatementTemplate and Constraint that result from the translation of n-ary properties, respectively.

![FIG. 2: Simplified DSP diagram qualified with the RDF/RDFS UML profile](image)

3. Towards a RDF-compliant DSP model for Earth observations

We consider the constraint language DSP as a metamodel, associated with a dedicated UML profile that allows expressing the DSP semantic specifications by means of stereotypes. These stereotypes specialize the stereotypes described within the RDF/RDFS UML profile and give a specific way of defining typical constraints between DSP languages elements.

The main advantage of such an approach is to significantly raise the level of abstraction by providing nested models and different possibilities of zooms on the same elements of interest accordingly. We can therefore build a DSP model tailored to our specific domain by instantiating the DSP profile. The interest is two fold: we just have to concentrate on the description of the Earth observation resources and the approach is generic and may be reused for other thematic domains. FIG. 3 gives an excerpt of the DSP instantiation model for EO resources. The classes EarthImage_T and Temporal_Extent_T are annotated by the meta-class dsp:DescriptionTemplate that is a kind of RDFSClass. EarthImage_T refers to the class EarthImage through an association annotated with the meta-property dsp:resourceClass. Additionally EarthImage_T is supported by a number of unnamed classes marked by the meta-class dsp:NonLiteralStatementTemplate. One of these classes links the Earth Image description template to the second description template entitled Temporal_Extent_T and entirely dedicated to the temporal aspects. An object Temporal_Extent_T is connected through an object typed by dsp:StatementTemplate to an object...
iso19108:Instant which corresponds to the date when the image is collected. The temporal metadata standards, in our case ISO 19108 are involved and will greatly facilitate the use of images in the context of multitemporal studies (see subsection 4.2). A similar outcome was achieved related to temporal aspects.

The same excerpt of the EOAP model is presented in Listing 1 on the following page using N3 syntax. The EOAP model pays particular attention to carefully describing the required level of temporal and spatial coverage in relation to real needs for Earth observing and uses relevant metadata standards. We illustrate in the section 4, the importance of the spatial and temporal dimensions by portraying various use cases of interest.

The contribution of the EOAP model together with the provision of multiresolution and multitemporal satellite imagery eases identification of environmental patterns over time and space.

4. Potential implementation of DCAP for image discovery and consultation purposes

4.1. Targeted users

The GEOSUD SDI is intended for French public stakeholders, whose missions contribute to environmental monitoring of French territories. The term “public stakeholder” covers a wide range of actors: those of research and higher education (research laboratory, university, high school) conducting studies, for instance, on the structure, functioning or dynamics of ecosystems.

Others actors include such decentralized or central governmental units, whose missions require, for instance the building of cartographic products to meet the monitoring or control of the implementation of governmental policies. Finally, local authorities are also public stakeholders whose missions have been recently extended to the management of the environment (waste management, biodiversity, air quality...) and for which high-resolution satellite imagery provides reference spatial datasets among others.
The variety of actors and the missions assigned to them point out the diversity of expertise and point of view, which our distribution platform must meet for image discovery and consultation purposes. The majority of these actors have little to no skills in remote sensing. Logically, their discovery and consultation requests will in the first stage be based on spatial and temporal properties of the satellite image. Others who have strong skills are able to have an expert approach and evaluate the appropriateness of an image from its characteristics (pixel resolution, format, encoding), or even those of the sensor or the conditions of acquisition of the image (incidence angle, cloud cover...)

4.2. Planned use cases

As described above, the target users are from very different domains and skills. The intended uses are equally variable, as they are designed to meet a wide range of environmental issues (see previous section). Among all these uses and to illustrate the relevance of our application profile in the GEOSUD SDI, we have chosen to describe three of them. The first two cases are about regulatory control missions for the management of renewable resources, the monitoring of land use planning by local governmental unit. The third case is an experiment from the scientific community with the aim to identify wetlands in tropical area.

Listing 1: EOAP RDF Excerpt in N3

```
ooap:TemporalExtent_T
  a          dsp:DescriptionTemplate;
  dsp:maxOccur "infinity"^^xsd:nonNegativeInteger;
  dsp:minOccur "0"^^xsd:nonNegativeInteger;
  dsp:resourceClass iso19108:Period;
  dsp:standalone false;
  dsp:statementTemplate [ a                dsp:NonLiteralStatementTemplate;
                          dsp:maxOccur "1"^^xsd:nonNegativeInteger;
                          dsp:minOccur "0"^^xsd:nonNegativeInteger;
                          dsp:nonliteralConstraint [ a       dsp:NonLiteralConstraint;
                                          dsp:valueURI occurrence "mandatory"^^dsp:occurrence;
                                          dsp:_vocab:EncodingScheme occurrence
                                          "mandatory"^^dsp:occurrence;
                                          dsp:property iso19108:instant ];
  dsp:property iso19108:begin ].

ooap:EarthImage_T
  a          dsp:DescriptionTemplate;
  dsp:maxOccur "1"^^xsd:nonNegativeInteger;
  dsp:minOccur "0"^^xsd:nonNegativeInteger;
  dsp:resourceClass ooap:EarthImage;
  dsp:standalone true;
  dsp:statementTemplate [ a                dsp:NonLiteralStatementTemplate;
                          dsp:maxOccur "1"^^xsd:nonNegativeInteger;
                          dsp:minOccur "0"^^xsd:nonNegativeInteger;
                          dsp:nonliteralConstraint [ a       dsp:NonLiteralConstraint;
                                          dsp:descriptionTemplateRef ooap:TemporalExtent_T;
                                          dsp:valueURI occurrence "disallowed"^^dsp:occurrence;
                                          dsp:property dct:terms:temporal ].
```
4.2.1 Non expert use case: mapping and temporal monitoring of clearcuts in Landes massif in the Southwest of France by decentralized governmental unit (DRAAF³)

In line with their missions and to enforce the regulations on exploitation of forest resources, DRAAF and DDT⁴ must implement control and monitoring tools, firstly, to establish control plans of clearcuts and secondly to ensure sustainable management of the forest resource. The production of a map identifying at time \( t \) the clearcuts of a forest is based on a methodology defined by (Ose, 2015). First, the mapping of clearcuts requires having two sets of vector data, one used to restrict the study area and the other to take into account the land use. The determination of clearcuts is based on two high-resolution satellite images acquired in the same season (preferably in spring) between an interval of two years or more. This is to calculate the difference of NDVI (Normalized Difference Vegetation Index) during the given interval and quantify the evolution of clear cuts. Taking the example of the DRAAF, who wants to establish the mapping in the Landes forest during the last two years, the GIS engineer knowing the clearcuts mapping methodology (but not a specialist in remote sensing) will want to discover the required images for his study with these words: "I am searching for high-resolution images that were acquired during the period from April to May for the years 2014 and 2015, covering the area from latitudes 43,97°; 45,06° - longitudes -1,56°; -0,133°".

4.2.2 Non expert use case: mapping of artificial sprawl in the peri-urban zone of Montpellier by the national observatory of agricultural space consuming (ONCEA)

In France, urban sprawl dynamics are particularly strong. The increase and spreading of built-up areas towards the periphery takes place to the detriment of natural and agricultural spaces. The conversion of land with agricultural potential is cause of serious concerns as it is usually irreversible. Thus, for the land use planning services, the mapping of artificial sprawl dynamics is an essential tool for the quantification of lost agricultural space.

Based on the method of Dupuy et al. (2012), the monitoring of artificial sprawl is based on a very high resolution image provided by satellite such as Pleiades or SPOT6. It is also necessary to use French large-scale data repositories that provide both the state of the human impact of an area (roads, buildings) and the land use (forest, crops...). An object-oriented analysis of very high resolution image ensures the recognition of new buildings and transport infrastructure elements and thus quantifies the agricultural areas that were urbanized. The

ONCEA GIS engineer wants to build the mapping of artificial sprawl for the cities of Lattes and Pérols (southern of the city of Montpellier) to assess the evolution since 2012. To provide this, the engineer must have very high resolution images covering the area in question. These images should have a sub-metric resolution and must be acquired during the spring or summer. The object-oriented analysis is more efficient for this period. Also, we could formulate his request in our discovery application as: "I am searching for images with sub-metric resolution, acquired during the period from March to September, covering the cities of Lattes and Pérols”.

The use of toponyms to select the images overlapping the study area takes advantage of semantic external resources. In this case, we use the places ontology called Geonames (http://sws.geonames.org/). It allows us to match the spatial footprint of an image, expressed in geographic coordinates, with those of cities supplied by Geonames ontology. Finally, we can annotate the images with the names of the cities that are included or that intersect the spatial extent of an image.

---

² Clearcut: refers to a mode of forestry development through cutting down of all trees of a parcel
³ DRAAF: Regional Headquarter for food, agriculture and forest
⁴ DDT: Sub-regional headquarter for territory management
4.2.3 Expert use case: Discrimination of wetlands in Madagascan forest by a remote sensing specialist

Proposed by [Hajalalaina, 2013] the identification of wetlands in the Madagascan forest meets agronomic and environmental issues. The wetlands are potential areas for rice crops and also biodiversity reserve. The proposed method is based on the use of multi-source and multi-resolution images. The LandSat7ETM+ product, at 30 meters of resolution, will as a first step allow drawing up a map of wetlands at regional level through a classification of the image pixel. In a second step, an object-oriented classification is applied to a high resolution image (2.5 meters), namely SPOT5 image. This second step results in a mapping of wetlands at the local level. In order to have these two kinds of images available, the remote sensing specialist will formulate his query specifying the name of the sensors required which provide the expected spatial resolution for the determination of wetlands. The specialist will also build his request on the characteristics of the image in order to define the temporal and spatial extents over which he wishes to conduct his study. Thus, we could formulate the entire request as: "I am searching for images acquired by Landsat 7 platform and the images acquired by the SPOT5 in panchromatic mode whose spatial footprints are between the latitudes -20,58°; -22,35° and longitudes 47,85°; 46,44°, which were acquired between the month May and June."

5. Conclusions and future work

The work carried out results in the definition of a Dublin Core application profile that is intended to meet the needs of different actors with regard to both satellite imagery uses and environmental issues. The use cases that we have examined reveal high requirements in capabilities to access, query and analyze a significant number of series of satellite images. The application profile EOAP is hence logically supported by metadata standards that are specifically dedicated either to spatial and temporal dimensions or to descriptions of observations and measures. An image is above all a digital resource and EOAP is also drawing on Dublin Core Metadata Element Set.

Additionally we initiate work towards meta-modeling activities to complete the RDF-based DSP model with higher levels of abstraction to efficiently drive the building of a thematic model that conforms to the DSP model. We will continue our modeling efforts focusing on two main directions. First UML profiling activities could constitute an efficient way to design an application before committing to implementation. We will therefore develop a generic RDF-based editor to build DSP models from the defined UML profiles. Secondly, we will add some constraints based on OCL (Object Constraint Language) (Clark, 2002) to the DSP metamodel. These constraints will be used for the validation of instantiations.

Acknowledgements

This work was supported by public funds received in the framework of GEOSUD, a project (ANR-10-EQPX-20) of the program "Investissements d'Avenir" managed by the French National Research Agency.

References


CWA 15248 Guidelines for machine-processable representation of Dublin Core Application Profiles, April 2005


Fowler, Martin. 2010. Domain Specific Languages (1st ed.). Addison-Wesley Professional.


