GI_Forum 2021, Issue 1
Page: 53 - 59
Research Paper
Corresponding Author:
martin.sudmanns@sbg.ac.at
DOI: 10.1553/giscience2021_01_s53

One GUI to Rule Them All: Accessing Multiple Semantic EO Data Cubes in One Graphical User Interface

Martin Sudmanns[1,2], Hannah Augustin[1], Lucas van der Meer[1], Christian Werner[1], Andrea Baraldi[3] and Dirk Tiede[1]

[1]University of Salzburg, Austria
[2]Spatial Services GmbH, Salzburg, Austria
[3]Baraldi Consultancy in Remote Sensing (BACRES), Modena, Italy

Abstract

Spatio-temporal analysis capabilities of big Earth observation (EO) data are possible now on various infrastructures, but the transferability and interoperability of analyses remain challenging. This contribution describes an approach for interacting with multiple semantic EO data cubes, where for each observation, at least one nominal (i.e., categorical) interpretation is available and can be queried in the same instance. Our in-house developed Web-based graphical user interface (GUI) provides technical access to multiple semantic EO data cubes, regardless of what infrastructure they are implemented on. It is designed to create semantic models using a graphical language, and an inference engine is able to evaluate these models against existing semantic EO data cubes based on a user's defined area and timespan of interest. Querying on a semantic level allows the transferability of semantic models across EO data cubes. Our contribution shows an approach towards solving this open research gap and discusses relevant challenges such as transferability of semantic models, on-demand instantiation, and federated EO data cubes. We believe that this approach offers new opportunities for improved semantic and syntactic interoperability in EO analyses and is better positioned to allow semantically-enabled queries possible in a federated EO data cube context.

Keywords: Big Earth observation data, interoperability, spatio-temporal querying, semantic EO data cubes

1 Introduction

Infrastructures for accessing and processing big Earth observation (EO) data are becoming increasingly mature and reliable. A prominent example is Google Earth Engine (Gorelick et al., 2017), but especially technologies based on the "data cube" idea, like the Earthserver (Baumann et al., 2016), the Euro Data Cube (https://eurodatacube.com/), or implementations of the Open Data Cube (Killough, 2018). Almost all of them employ some sort of spatio-temporal analysis capabilities. There is still no community-agreed definition for EO data cubes, but several works exist to better understand. The data cube manifesto (Baumann, 2017) defines a data cube as "a massive multi-dimensional array; 'massive' entails that we talk about sizes significantly beyond the main memory resources of the server hardware. Data values of the same data type sit at grid points as defined by the d axes of the d-dimensional data cube. Coordinates along these axes allow addressing data values unambiguously.". These new technological advances offer users access to EO data via spatio-temporal coordinates rather than archive-specific file-based access. New challenges surround how best to allow flexible and transferable analyses, potentially across multiple data cubes and technical infrastructures.

One of the most notable concepts for improving the transferability of analyses and algorithms is to populate EO data cubes with analysis-ready-data (ARD) (Lewis et al., 2018, Dwyer et al., 2018, Giuliani et al., 2017). Imagery calibrated to bottom-of-atmosphere (surface reflectance) together with a set of mandatory and optional quality information (e.g. cloud contamination) is one example of ARD for optical EO data. CEOS defines ARD as "satellite data that have been processed to a minimum set of requirements and organised into a form that allows immediate analysis with a minimum of additional user effort and interoperability both through time and with other datasets" (Lewis et al., 2018). In theory, this allows an algorithm to be executed in different EO data cubes as long as they provide ARD. An example is the Water Observation from Space (WOFS) algorithm (Mueller et al., 2016), which has been successfully applied in the Digital Earth Australia and Digital Earth Africa data cubes.

Challenges in interoperability can be broken down into syntactic and semantic aspects of communication, while transferability is used in the context of robustness to changes of inputs. Syntactic interoperability can be achieved by technical standards for communication between a software client and a server (Schaeffer et al., 2012). Semantic interoperability refers to "the ability of services and systems to exchange data in a meaningful/useful way" (Research Data Alliance 2015). Transferability can refer to many things, but here we refer to the ability of an algorithm or analytical workflow to be used with different sets of input variables, ideally with minimal to no customisation required. These differences include but are not limited to different spatio-temporal areas of interest (e.g. geographic locations, time-spans, spatial extent or shape), different imagery from the same sensor, different sensor data (e.g. resolution, revisit time, spectral and radiometric characteristics), and even different application scenarios.

An approach beyond providing ARD is the semantic EO data cube, which provides additional semantic enrichment and data (Augustin et al., 2019). A semantic EO data cube is defined as "a data cube, where for each observation at least one nominal (i.e., categorical) interpretation is available and can be queried in the same instance". This allows executing analyses and data combination on a semantic level towards improved semantic interoperability; as long as the interpretation (semantic enrichment) of the data is the same, an algorithm is semantically interoperable and can be transferred across multiple data cubes and multiple sensors. However, such an implementation requires image understanding routines within an expert system (e.g. a factbase storing the facts (data and information), knowledgebase storing rules, inference engine applying the rules to the facts) in which semantic EO data cubes take over the role as factbases (Tiede et. al, 2017, Laurini & Thompson, 1992, p. 641).

The semantic EO data cube is a method mainly developed at the Department of Geoinformatics – Z_GIS; the infrastructure was built around the Sen2Cube.at semantic EO data cube for Austria. The semantically-enabled approach allows semantic querying, facilitated by our in-house developed Web-based Graphical User Interface (GUI) designed to allow users to switch between semantic EO data cubes of very different locations worldwide. The main purpose of the GUI is to develop, share and execute models based on the same semantic querying language, establishing a growing, common knowledge base.

2 Semantic EO data cubes and GUI-based access

A semantic EO data cube is typically embedded in a larger infrastructure that allows regular updates and semantic enrichment (and potentially automated instantiation of new EO data cubes) and convenient Web-based access directly in the browser with a graphical query language. Semantic models defined in the graphical querying language are translated into data cube queries and evaluated by an inference engine. Additional functionality includes a quick preview of query results, processing metadata (e.g. model, time frame, processing time) and access to query results either as a download or direct integration in other applications as a standardised WMS.

Semantic enrichment refers to interpreted content of EO imagery, i.e., mapping data to an interpretation that represents stable concepts. It is a necessary pre-processing step to create a semantic EO data cube. These interpreted concepts are generally non-ordinal, categorical variables; however, subsets of these variables may be ordinal (e.g., vegetation categorised by increasing greenness or intensity). The relative level of semantic enrichment can vary in terms of complexity and the "symbolic" level of the concepts/variables. The concept of semantic enrichment itself is independent of the technology and can be potentially achieved with other approaches, including any artificial-intelligence-based approach.

The definition of a semantic EO data cube as having at least one interpretation together with every observation requires not only a spatio-temporal data model that considers thematic information layers but also a metadata model. The metadata model must: (1) define the type of semantic enrichment; (2) allow displaying the type of the content in the GUI; and (3) allow automated evaluation of semantic models in the inference engine. To achieve this, we invented a 'layout' of a semantic EO data cube to describe the thematic information layers of the semantic EO data cubes. All of the three components are developed generically and consider the layout, thus allowing the creation of different 'flavors' of semantic EO data cubes.

Multiple semantic EO data cubes and the knowledgebase containing semantic models are accessible via the same GUI. This means that the GUI serves as a unified access point for multiple semantic EO data cubes (See Figure 1). Users do not need to use different access points for different semantic EO data cubes, while the semantic EO data cubes can even be hosted on different infrastructures.

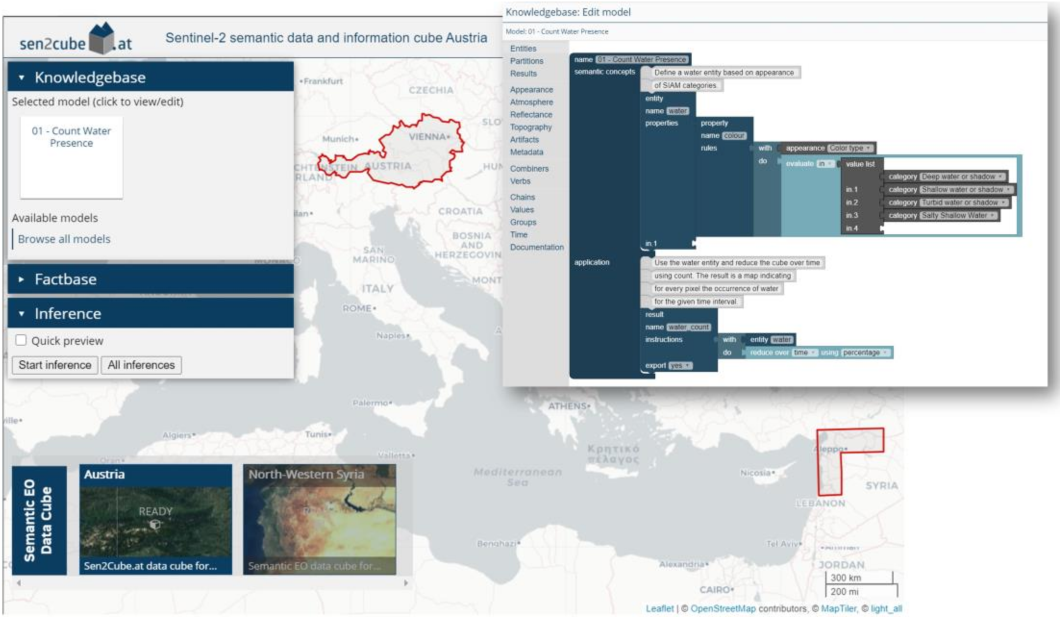

Figure 1: The GUI is designed to provide direct access to different semantic EO data cubes as factbases of the expert system (here: Sen2Cube.at covering Austria and a semantic EO data cube covering the north-western part of Syria). Since the factbases are defined using the same layout, the same semantic models are directly transferable and applicable between the different cubes.

Achieving semantic and syntactic interoperability between two or more semantic EO data cubes requires considering the transferability and re-usability of the semantic models and its dependency on the type of semantic enrichment and additional information (Sudmanns et al., 2018). The formulation of the semantic models is tied to the semantic EO data cube content, and we have identified three cases for achieving interoperability. First, if the layout of the semantic EO data cubes is defined differently, the models are directly transferable without any further adjustments. This is the easiest case in which users can switch between the semantic EO data cubes in the GUI and apply their semantic models. Second, if the layouts are different, it depends on which thematic information layers the models use. Usually, not all interpretation categories are used by a model. Therefore, the first sub-case is that the subset of categories used in a model is available in the layout, even if the rest is different. The second sub-case is that a model uses categories that are not available in the layout of another data cube, resulting in a situation in which the model cannot be evaluated. An example would be that one semantic EO data cube uses a different semantic enrichment or a digital information model, which is not available within an EO data cube at a different geographic location.

In our current setup, two semantic EO data cubes are instantiated and accessible within the Web-based GUI. Although the semantic EO data cube covering Austria is deployed on the EODC GmbH infrastructure and the one covering the north-western part of Syria is deployed on the University of Salzburg infrastructure. The GUI is designed to allow users to switch between them in the selection menu. Users create a model and can apply it to both of them, which is possible since both semantic EO data cubes are instantiated based on the same layout

(Figure 2). Therefore, the first case to achieve semantic and syntactic interoperability is already covered, while the second case with the two sub-cases remains a research gap.

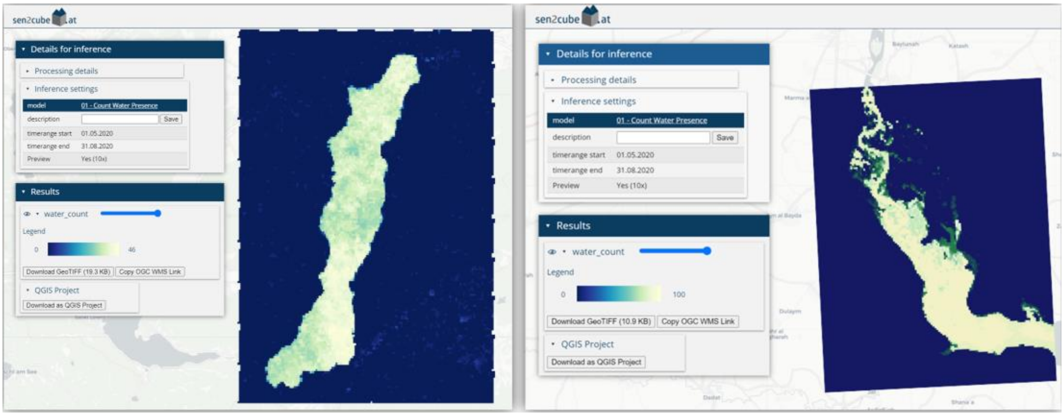

Figure 2: A semantic model can be transferred to multiple semantic EO data cubes and create comparable results. Here: Surface water extraction in Austria (left) and Syria (right) using the same semantic model, bright colours indicate a higher number of water observations in the selected time span.

3 Discussion

The in-house inference engine is programmed not to make any assumptions about the content, i.e., data and information layers of the semantic EO data cube. Every semantic model is evaluated based on the layout that defines the loading and processing of the required datasets.

Compared to accessing an EO data cube containing ARD using a Jupyter-Notebook, our approach creates an additional overhead in the development phase but removes several burdens from the users to create transferable analyses. It is easier to transfer semantic models instead of Jupyter Notebooks because the semantic enrichment and the inference engine abstract loading correct products and datasets from the users. The semantic model definition is separated from the selected AOI or time interval. Since the model creation and application are separated, and the model development approach is free of coding, it can contribute to increasing user uptake and allows the inclusion of new users, e.g., in an educational context, or allows different clients (mobile, desktop,...) to access the same knowledgebase and factbase. Once several semantic EO data cubes are available and accessible in the GUI, users do not need multiple access points or even have to learn different interfaces to conduct analyses on different parts of the world or using different sensors.

The main challenges with this approach are how to exchange models between semantic EO data cubes that have different layouts and how to deal with spatio-temporally dependent models. The key is to identify whether a model uses semantic categories that are a proper subset of the semantic EO data cube against which it will be queried. Further, some semantic models may not be fully independent of the geographic areas that are covered by the semantic EO data cube. This includes the formalisation of temporal sequences of agricultural practices that may be shifted due to different altitudinal belts or completely different due to the location (e.g. different latitude or hemisphere). Different climate zones or atmospheric conditions may also limit the transferability of some semantic models. A possible solution could be to calculate advanced measurements of fitness-for-use of available imagery and associated semantic EO data cubes or evaluate the spatio-temporal applicability of semantic models.

4 Conclusion and path forward

The concept and implementation of semantic EO data cubes are relatively new, yet they show promising performance and are suitable to be investigated further. They implement an image interpretation strategy such as computer vision to perform semantic enrichment, thus containing EO data together with at least one (categorical) interpretation. This approach allows querying using semantic models that are evaluated by an inference engine. Querying on the semantic level is the next level of abstraction that allows for semantic and syntactic interoperability. However, defining and instantiating semantic EO data cubes and evaluating the transferability of semantic models still requires extensive human expert intervention, similar to other EO based algorithms developed in a specific context.

In this contribution, we show our approach of making multiple semantic EO data cubes technically accessible to users within a single GUI, which is designed to allow users to choose between the single data cubes in a selection menu. A knowledgebase stores semantic models that can be used to query any of the semantic EO data cubes. We also show the challenges that may arise when semantic EO data cubes have different layouts and, therefore, may require different models.

Next investigations and outlook will be aligned along four lines of research with research gaps that have not yet been solved or tackled: (1) extending the layout to describe semantic EO data cubes, thus allowing machine-readable requests about the content and capabilities to allow automated evaluation whether a model is able to be evaluated or not; (2) automating the instantiation of multiple semantic EO data cubes based on a selected layout, their deployment in state-of-the-art cloud infrastructures and accessibility within one GUI; thus, users should not necessarily be concerned with the hosting provider of the semantic EO data cube; (3) enabling a single semantic query to be conducted across multiple semantic EO data cubes, e.g., in a federated context; (4) conducting user studies for testing and evaluating the efficiency of our approach compared to other approaches. These are pre-requisites for the successful implementation of (on-demand) semantic EO data cubes. Once they are operational, additional questions will concern application- and domain-related suitability of semantic models as well as the maintenance of the semantic EO data cubes, e.g. specifying user roles (admin, user, maintainer, ...) to define which user is allowed to submit a semantic query and how much resources will be available and allocated.

References

Augustin, H., Sudmanns, M., Tiede, D., Lang, S., & Baraldi, A. (2019). Semantic Earth observation data cubes. *Data*, 4(3), 102. https://www.mdpi.com/2306-5729/4/3/102

Baumann, P., Mazzetti, P., Ungar, J., Barbera, R., Barboni, D., Beccati, A., ... & Wagner, S. (2016). Big data analytics for earth sciences: The EarthServer approach. *International Journal of Digital Earth*, 9(1), 3-29. https://doi.org/10.1080/17538947.2014.1003106

Baumann, P (2017). Data Cube manifesto. https://earthserver.eu/tech/datacube-manifesto/The-Datacube-Manifesto.pdf

Dwyer, J. L., Roy, D. P., Sauer, B., Jenkerson, C. B., Zhang, H. K., & Lymburner, L. (2018). Analysis ready data: enabling analysis of the Landsat archive. *Remote Sensing*, 10(9), 1363. https://doi.org/10.3390/rs10091363

Giuliani, G., Chatenoux, B., De Bono, A., Rodila, D., Richard, J. P., Allenbach, K., ... & Peduzzi, P. (2017). Building an earth observations data cube: lessons learned from the swiss data cube (sdc) on generating analysis ready data (ard). *Big Earth Data*, 1(1-2), 100-117. https://doi.org/10.1080/20964471.2017.1398903

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, 202, 18-27. https://doi.org/10.1016/j.rse.2017.06.031

Killough, B. (2018). Overview of the open data cube initiative. In: IGARSS 2018 IEEE *International Geoscience and Remote Sensing Symposiu*m (pp. 8629-8632). https://doi.org/10.1109/IGARSS.2018.8517694

Laurini, R., & Thompson, D. (1992). Fundamentals of spatial information systems (Vol. 37). Academic press. San Diego.

Lewis, A., Lacey, J., Mecklenburg, S., Ross, J., Siqueira, A., Killough, B., ... & Hosford, S. (2018,). CEOS analysis ready data for Land (CARD4L) overview. In IGARSS 2018 IEEE *International Geoscience and Remote Sensing Symposium* (pp. 7407-7410). https://doi.org/10.1109/IGARSS.2018.8519255

Mueller, N., Lewis, A., Roberts, D., Ring, S., Melrose, R., Sixsmith, J., ... & Ip, A. (2016). Water observations from space: Mapping surface water from 25 years of Landsat imagery across Australia. *Remote Sensing of Environment*, 174, 341-352. https://doi.org/10.1016/j.rse.2015.11.003

Research Data Alliance. (2016). Agrisemantics Working Group. https://www.rd-alliance.org/group/agrisemantics-wg/case-statement/agrisemantics-working-group. Accessed March 2021.

Schaeffer, B., Baranski, B., Foerster, T., & Brauner, J. (2012). A Service-oriented Framework for Real-time and Distributed Geoprocessing. In E. Bocher and M. Neteler (Eds.), *Geospatial Free and Open Source Software in the 21st Century*, Springer Heidelberg. doi:10.1007/978-3-642-10595-1_1

Sudmanns, M., Tiede, D., Lang, S., & Baraldi, A. (2018). Semantic and syntactic interoperability in online processing of big Earth observation data. *International Journal of Digital Earth*, 11(1), 95-112. https://doi.org/10.1080/17538947.2017.1332112

Tiede, D., Baraldi, A., Sudmanns, M., Belgiu, M., & Lang, S. (2017). Architecture and prototypical implementation of a semantic querying system for big Earth observation image bases. *European journal of remote sensing*, 50(1), 452-463. https://doi.org/10.1080/22797254.2017.1357432