



Semantic Earth Observation Data Cubes

Hannah Augustin ^{1,*}, Martin Sudmanns ¹, Dirk Tiede ¹, Stefan Lang ¹ and Andrea Baraldi ²

¹ Interfaculty Department of Geoinformatics—Z_GIS, University of Salzburg, 5020 Salzburg, Austria

² Italian Space Agency (ASI), 00133 Rome, Italy

* Correspondence: hannah.augustin@sbg.ac.at; Tel.: +43-662-8044-7574

Received: 15 June 2019; Accepted: 15 July 2019; Published: 17 July 2019



Abstract: There is an increasing amount of free and open Earth observation (EO) data, yet more information is not necessarily being generated from them at the same rate despite high information potential. The main challenge in the big EO analysis domain is producing information from EO data, because numerical, sensory data have no semantic meaning; they lack semantics. We are introducing the concept of a semantic EO data cube as an advancement of state-of-the-art EO data cubes. We define a semantic EO data cube as a spatio-temporal data cube containing EO data, where for each observation at least one nominal (i.e., categorical) interpretation is available and can be queried in the same instance. Here we clarify and share our definition of semantic EO data cubes, demonstrating how they enable different possibilities for data retrieval, semantic queries based on EO data content and semantically enabled analysis. Semantic EO data cubes are the foundation for EO data expert systems, where new information can be inferred automatically in a machine-based way using semantic queries that humans understand. We argue that semantic EO data cubes are better positioned to handle current and upcoming big EO data challenges than non-semantic EO data cubes, while facilitating an ever-diversifying user-base to produce their own information and harness the immense potential of big EO data.

Keywords: remote sensing; big Earth data; big EO data; information extraction; semantic enrichment; time-series

1. Introduction

The current Earth observation (EO) data pool is vastly different than a mere decade ago, but the main challenge remains: to produce information from data to generate knowledge [1,2]. We are surrounded by a growing ocean of EO data, but sensory data are not information and have no inherent meaning (i.e., lacking semantics) without some form of interpretation. At a minimum, this data pool is characterised by a rapidly growing data volume, accelerating data velocity (i.e., increasing data acquisition and processing speeds) and an increasingly diverse variety of sensors and products [3]. The term “data cube” is broadly understood as a multi-dimensional array organising data in a way that simplifies data storage, access and analysis compared to file-based storage and access [4]. Applying data cube technology to EO datasets attempts to address some of the challenges and opportunities rooted in these big data characteristics.

There is a growing number of implementations currently referred to as EO data cubes with the goal of lowering the barrier to store, manage, provide access to and analyse EO data in a more convenient manner. Data cubes of EO imagery typically are organised in three dimensions: latitude, longitude and time. The definitions or specifications of EO data cubes will not be discussed here but can be understood as a way of organising EO data using a logical view on them, either based on an existing archive (i.e., “indexing”) or a specific, application-optimised, multi-dimensional data structure (i.e., “ingestion”). The logical view refers to the way of accessing EO data by using spatio-temporal coordinates either in an application programming interface (API) or a query language instead of file

names. The main advantage of ingesting data is that the data can be stored in a query-optimised way, and specific access patterns can be realised more efficiently, such as time series analysis or spatial analysis.

Various technical solutions to create these logical views on EO data have rapidly gained traction over the past few years. The first national scale EO data cube was established in Australia [5], whose technology is now the basis of Digital Earth Australia [6] and the Open Data Cube (ODC) [7]. The free and open source ODC technology is also behind other operational EO data cubes, such as in Switzerland [8], Colombia [9], Vietnam [10], the Africa Regional Data Cube [11] and at least nine other national or regional initiatives under development [7]. Rasdaman [12], an array database system that has been around since the mid-1990s, is another leading technology behind initiatives such as EarthServer [13] and the Copernicus Data and Exploitation platform for Germany (CODE-DE, [14]). Other software implementations exist, such as the Earth System Data Cube from the European Space Agency [15] and SciDB [16].

State-of-the-art EO data cubes simplify data provision to users by facilitating data uptake and aiming to provide analysis-ready data (ARD) [4]. While there is still an ongoing discussion about how ARD are defined and specified, it is usually understood as calibrated data, and in the case of CARD4L (Committee on EO Satellites ARD for Land), even contains masks as a target requirement specification, such as for cloud and water [17,18]. The intention is to shift the burden of pre-processing from users to data providers, who are often better equipped to consistently and reliably process large volumes of high-velocity data [6,17,19]. Processing steps with a high potential level of automation can be conducted centrally where they only must be conducted once and are then available to all users. This contrasts with requiring every user to pre-process the data they would like to use on their own and improves comparability of initial data conditions between users and applications.

Web-based access to these EO data cube implementations brings users closer to the data and implements a computation platform at the data location [20]. This is a different strategy than providing EO data to users as individual, downloadable images of a pre-determined spatial extent. Data cubes make data access much more efficient and effective by providing users with data tailored more specifically to their needs, reducing unnecessary data transfer [20]. Pairing data access from EO data cubes with computational environments, (e.g., processing resources accessible using Jupyter notebooks) moves in the direction of other existing Web-based geospatial computation platforms, such as Google Earth Engine [21]. While these platforms are powerful, analyses sometimes have limited transferability to different geographic locations or points in time, or a low level of results or inferential reproducibility [22].

Even with tailored ARD access and Web-based processing capabilities, users of EO are still confronted with tons of data rather than information and the ill-posed challenge of reconstructing a scene from one or more images [23]. In this case, a scene is understood as the content of an image, whereby the result of this challenge is some form of interpretation or classification map of an image. Images suffer from data dimensionality reduction and a semantic information gap. An image is a 2D snapshot of the 4D world (i.e., three spatial and one temporal dimension), whereby all the information required to reconstruct a comprehensive and complete descriptive scene is not available from one or multiple images over time [24].

Information production from EO images still generally relies on unstructured, application-specific algorithms or increasingly popular machine learning procedures. This often results in low to no semantic interoperability between workflows, sensors or images based on the findable, accessible, interoperable and reusable (FAIR) principles [25]. The FAIR principles refer to data, and the algorithms, tools and workflows that produce them. If data-derived information is linked to the images used to generate them, provenance is maintained and accessible to users. Combining EO images with symbolic image-derived information in a collaborative, analytics environment effectively facilitates increased semantic interoperability between workflows and analyses while extending machine-actionability [26].

If the image-derived information is semantically interoperable and consistent between locations and acquisitions, semantic interoperability is established at least at the starting point of further analysis.

We are introducing the concept of a semantic EO data cube as an advancement of state-of-the-art EO data cubes. Semantic EO data cubes move beyond data storage and provision by offering basic, interoperable building blocks of image-derived information within a data cube. This enables semantic analyses that can be incorporated into simple rule-sets in domain language, and users are able to develop increasingly expressive, comprehensive rule-sets and queries. Given semantic enrichment that includes clouds, vegetation, water and “other” categories, certain semantic content-based queries covering a user-defined area of interest (AOI) in a given temporal extent are possible, such as for the most recent observations excluding clouds (e.g., user-defined cloud-free mosaic), or an observed moment in time with the maximum vegetation extent. These queries of the interpreted content of available images are independent of imposed spatial image extents and are made possible by including semantic enrichment. However, since the information is still tied to the EO images it is based on, it is also possible to search for and retrieve images based on their semantic content rather than only metadata (e.g., where and when each image was acquired). We argue that EO data cubes have the potential to offer much more than data and information product storage and access. They move towards reproducible analytical environments for user-driven information production based on EO images and allow non-expert users to use EO data in their specific context.

This paper focuses on the concept of a semantic EO data cube, assuming the basis is an EO data cube containing EO data together with a nominal interpretation for each observation. Multiple discussions and standardisation processes are currently underway to clarify what constitutes ARD and what minimum requirements constitute a data cube. However, this has little bearing on the base concepts presented here, which have implications for data access, data retrieval, semantic queries of data, semantic interoperability of different methods and results and more. We argue that semantic EO data cubes are better positioned to handle current and upcoming big EO data challenges than non-semantic EO data cubes, while facilitating an ever-diversifying user-base to produce their own information and harness the immense potential of big EO data.

2. Theoretical Framework

Concepts under the same name sometimes differ between domains. The concepts essential for our understanding of semantic EO data cubes are described for clarity, and our definition of what constitutes a semantic EO data cube is explained.

2.1. Clarifying Concepts

Data are not the same as information, and we find ourselves increasingly collecting data, yet not producing more information from them at the same rate. Information can be understood at least in two different ways: as a quantifiable measure in the sense of the information content of a message or an image (e.g., bits and bytes representing something informative [27]), or as a subjective concept, an interpretation (i.e., knowledge produced from a process) [28,29]. Information is used to generate knowledge and understanding, which might lead to wisdom [1,2].

Two terms ought to be clarified before moving forward because they are not interchangeable from our perspective, nor in the domain of computer vision: images and scenes. An EO image is broadly understood as a pixel-discretised field representing measurements of reflected radiations from Earth in different wavelengths (e.g., temperature, visible light, microwave). EO data are delivered as images or single measurements, depending on the design of a sensor. Here we refer to numerical observations represented by pixels and delivered as images. A scene, however, refers to the represented content of an image, meaning that which was observed [30].

The goal of most EO analysis is to produce actionable information to support decision-making processes. This requires transforming EO data into information, or digital numbers into subjective concepts that describe a scene. An EO image is a 2D representation of a 3D scene on Earth at a fixed

moment in time, and multiple 2D images of the same 3D scene acquired over time move towards representing a snapshot of the 4D world (i.e., 3D space through time). In this context, what an image or set of images can tell you about a scene is information.

The challenge of reconstructing scenes or generating information about a scene from a mono-temporal 2D image or set of images through time underpins any classification of remotely-sensed imagery and is inherently ill-posed. It is ill-posed in the Hadamard sense because a single, unique solution may not exist, or the solution does not depend continuously on data [31,32]. The last criterion of data-dependence refers to stability, where small changes in the equation or conditions result in small changes in the solution. An ill-posed problem does not meet one or more of these criteria (e.g., there are a huge number of possible solutions when classifying imagery).

The ill-posed problem of reconstructing scenes from images stems primarily from what is known as the sensory gap [33]. For optical EO images, this gap exists between the 2D image that has been sensed (e.g., digital numbers) and the 4D world (e.g., objects, states, events, processes). This gap introduces uncertainty that inherently complicates the interpretation of images and reliable, consistent information production. One aspect of the sensory gap is the sensor transfer function, which relates to the resolvability of phenomena by the given sensor (e.g., spatial, temporal, spectral, radiometric resolution). Another aspect relates to the reduction of dimensionality inherent to images (i.e., 4D to 2D; reducing a flood event to a snapshot in time). These aspects together allow for multiple interpretations of the same or similar representations (e.g., a green pixel in a true-colour image might represent a vegetated rooftop, forest, pasture, football field or something else entirely).

In the context of EO image classification, multiple classifications are possible for any given EO image or collection of images, and many current classification methods are very sensitive to changes in input parameters or starting conditions. Certain methods even produce similar but non-identical results each time they are run on the same initial data. In the case of well-established approaches of supervised classification, different users generally use different sets of samples even if using the same data and being interested in the same categories, which consequently produces different results.

What is known as the semantic gap also contributes to difficulties in producing information from images, and it refers to the gap between something that exists and what it means, regardless of how it is observed or represented [33]. Semantics more broadly refers to a multi-domain study of meaning but influences research in many domains, such as philosophy, linguistics, technology (e.g., the semantic Web [34,35], ontology-based data access [36]) and interoperability (e.g., sharing geographic information [37], processing EO data [26]).

When we speak of semantics in EO, this refers to what an EO image represents in terms of how it is interpreted, usually by an expert. An image can be described using an unbelievable number of words and concepts, yet images do not have intrinsic meaning. Each person has their own definition or understanding of different concepts or symbols, not to mention what they find to be important in a given image or scene [38]. Images gain meaning through relations to other images and the interpretation by a viewer, which is influenced by cultural and social conventions, not to mention the viewer's intention. In the context of image databases, how users search for and interact with images creates additional meaning, especially if given an exploratory user interface [39].

Using the term semantic in relation to EO data cubes refers to how an existing EO data cube is semantically-enabled, meaning a user can interact with it using semantic concepts rather than digital numbers or reflectance values. The ability to search for and retrieve EO data using spatially-explicit semantic content-based information rather than metadata, keywords, tags, or other linked data has strong implications for changing the way EO data is queried, accessed and analysed. However, to semantically-enable an EO data cube, some level of semantics needs to be available for every observation. In the case of EO imagery, this means semantics need to be available for each representation in space and time (i.e., pixel).

2.2. Our Definition of a Semantic EO Data Cube

A semantic EO data cube or a semantics-enabled EO data cube is 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. Interpreting an EO image (i.e., mapping data to symbols that represent stable concepts) results in semantic enrichment [23]. This data interpretation used in creating a semantic EO data cube may differ depending on the user and the intended purpose. Semantic variables are non-ordinal, categorical variables, but subsets of these variables may be ordinal (e.g., vegetation with sub-categories of increasing greenness or intensity) [40]. See Figure 1 (left) for a schematic illustration of a semantic EO data cube.

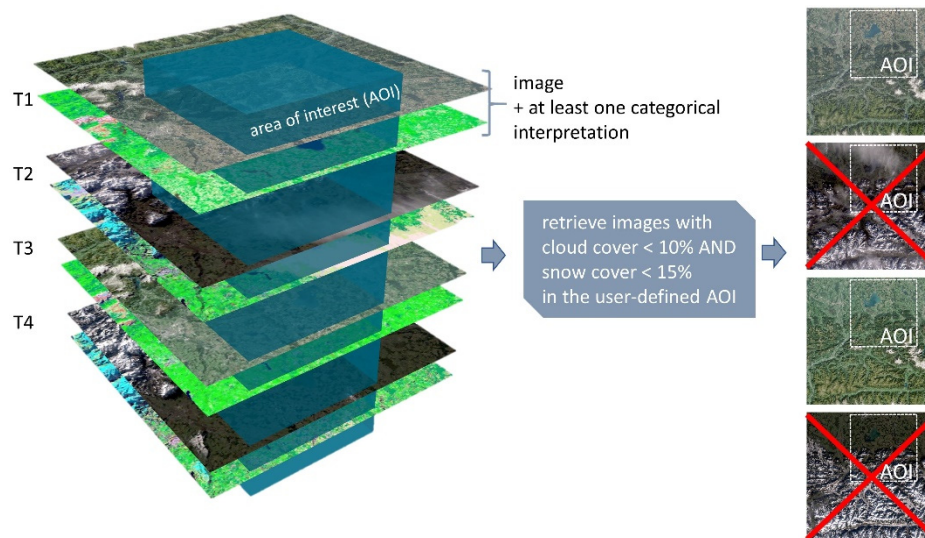


Figure 1. Schematic illustration of a semantic Earth observation (EO) data cube (left) used for an exemplary semantic content-based image retrieval (SCBIR) query. Here, a query searches for images with low cloud and low snow cover within a user-defined area of interest (AOI)-based on the associated semantic information. It retrieves images that match the semantic content-based criteria for the AOI instead of the entire image's extent. In a classic image wide query such AOI specific semantic queries are not possible.

Semantic enrichment included in a semantic EO data cube may be at a relatively low or higher semantic level. A lower semantic level means that symbols may be associated with or represent multiple semantic concepts requiring further analysis or interpretation to align with more specific concepts. The concepts in a lower level semantic enrichment can be considered semi-symbolic in that they are a first step to connecting sensory data to symbolic, semantic classes [41]. This could include information such as colour, or other ways of characterising the spatio-temporal context of each observation. A relatively high semantic level refers to explicit expert knowledge or existing ontologies. In the context of optical EO, one example of relatively high level semantic information would be land cover, such as the land cover classification system (LCCS) developed by the Food and Agriculture Organisation of the United Nations [42].

Other data and information may be combined with a semantic EO data cube to extend possible analysis, but what makes it semantically-enabled is that each observation in space over time has an interpretation. An interpretation that can be generated in an automated way with no user interaction is ideal for handling big EO data. It is also extremely beneficial if the resulting interpreted categories are transferable between different geographic locations, moments in time, images or sensors.

Only including well-known, data-derived indices for each observation (e.g., normalised difference vegetation index (NDVI)) is not sufficient to semantically-enable an EO data cube. Most of these indices are not inherently semantic, in that they still need to be interpreted to have symbolic meaning (e.g., at what NDVI is a pixel considered to contain vegetation or some other interpreted category?).

Indices can, however, contribute different quantitative insights to existing interpretations of an image in a stratified analysis (e.g., this collection of pixels is interpreted as being vegetation, but what was the average NDVI in June 2018 compared to June 2019 within this area?). While the indices can be calculated on the fly since the EO data are also present in a semantic EO data cube, it is up to the user as to whether calculating and incorporating such data-derived indices in a data cube reduces computational resources or has other benefits for further analyses.

Including additional data or information that is not directly derived from EO data does not semantically enable an EO data cube but might enable new query possibilities of EO data in space and time. Such data or information could concern the geographic area (e.g., digital elevation model (DEM)), socio-economic data, or masks of various kinds (e.g., urban area or forest mask). All of these data and information sources are not derived directly from the EO data such that they: (1) do not add information about each EO image's content, but rather the scene content or other characteristics pertaining to the time they were acquired; and (2) may no longer be true for the moment in time an EO image was captured (e.g., a DEM acquired before an earthquake). A DEM, for example, could be used as a spatial selection criterion, even if not specifically related to the semantic content of each image (e.g., selecting observations above a given elevation for alpine areas). Another example would be including an annual forest mask used by environmental regulatory bodies, but that annual mask may not be true even for the EO data available for that given year contained in the data cube.

In semantic EO data cubes it is crucial that EO data be stored with data-derived information for each acquisition. A data cube containing only data-derived interpretations could be considered semantic, but EO data have too much potential to be constrained to a single interpretation, especially since there is no single correct interpretation of image content. World ontologies are infinite. Multiple different perspectives and interpretations need to be possible to close the semantic gap [38], and users should be allowed to generate their own interpretations within a semantic EO data cube should those available not be suitable for their needs. The loss of connection to original EO data constrains semantics to the available interpretation, eliminates access to the source of the data-derived information important for provenance and limits further analysis. Some users might benefit from incorporating reflectance values from specific bands (e.g., calculating an index), using the semantic information to generate composite images through time, or generating different information based on the data to augment existing semantic enrichment.

The focus of semantic EO data cubes is to facilitate ad hoc, flexible information generation from data, that might have potential to lead to knowledge. Semantic EO data cubes combine concepts from EO, image processing, geoinformatics, computer vision, image retrieval and understanding, semantics, ontologies and more. Similar to how the semantic Web can be considered an extension of the Web [34], semantic EO data cubes offer a solution to combining EO data with meaning. This ultimately better enables people and computers to work together to access, retrieve and analyse EO data and data-derived information in a semantically-enabled and machine-readable way.

3. Examples from Existing Semantic EO Data Cubes

Three applied examples of semantic EO data cubes are presented, and each of them uses the same relatively low-level, generic, data-derived semantic enrichment as the basis for each of the semantic EO data cubes. This general-purpose semantic enrichment is application- and user-independent and thus can support multiple application domains. The semantic enrichment used in the following examples is automatically generated (i.e., without any user-defined parameterisation or training data) by the Satellite Image Automatic Mapper™ (SIAM™). This software is an expert system that employs a per-pixel physical spectral model-based decision-tree to images calibrated to at least top-of-atmosphere reflectance in order to accomplish automatic, near real-time multi-spectral discretisation based on a priori knowledge [43]. The decision tree maps each observation located within a multi-spectral reflectance hypercube to one multi-spectral colour name, which is stable and sensor agnostic. It is sensor-agnostic in that data calibrated to at least top-of-atmosphere reflectance by optical sensors

can be used to generate semantic enrichment comparable between sensors (e.g., Sentinel-2, Landsat). SIAM™'s output has been independently validated at a continental scale by [44].

This colour naming results in a discrete and finite vocabulary referring to hyper-polyhedra within a multi-spectral feature space, whereby the colour names create a vocabulary that is a mutually exclusive and totally exhaustive partitioning of the multi-spectral reflectance hypercube. These colour names have semantic associations using a knowledge-based approach and thus are considered semi-symbolic (i.e., semi-concepts). More broadly, this vocabulary of colour names can be thought of as stable, sensor-agnostic visual “letters” that can be used to build “words” (i.e., symbolic concepts) that have a higher semantic level using knowledge-based rules. The output may be considered sufficient for generating CARD4L masks as specified in the product family specification [18], but also offers building blocks for a complete scene classification map.

In the following examples, these data-derived information building blocks (i.e., semi-concepts) are based on Landsat 8 or Sentinel-2 images and are stored using either Open Data Cube or rasdaman technology to create semantic EO data cubes. While the semi-concepts themselves are inferior in semantics to land cover classes, they are reproducible, transferable between images and geographic locations, and each colour has a semantic association. These implementations serve as the foundation for semantic content-based image retrieval (SCBIR) (Section 3.1) or other semantic queries (Sections 3.2 and 3.3). Spectral-based semi-concepts can serve as the basis for more expressive, automated scene classification, queries and analysis within each of these prototypical semantic EO data cubes using knowledge-based rules (see Section 4.4).

3.1. Semantic Content-Based Image Retrieval

The example of operational SCBIR has been prototypically implemented within a semantic EO data cube based on Landsat 8 data and the rasdaman array database system as an underlying data cube technology [45]. While this prototypical implementation (see Figure 1) did not cover a large database, it is designed for scalability by relying on parameter free, fully automated and multisensory enabled semantic enrichment, as well as on a data cube technology proven to be scalable to PB sizes [13].

Unlike a traditional content-based image retrieval system, a SCBIR system is expected to cope with spatially-explicit (i.e., area of interest (AOI)-based), temporal, semantic queries (e.g., “retrieve all images in the database where the AOI does not contain clouds or snow”). Very few SCBIR system prototypes targeting EO images have been presented in the literature [46,47]. None of them is available in operating mode to date.

The implementation of SCBIR is urgently needed in today’s big EO archives to overcome the limitations of currently implemented image data retrieval methods using image metadata (e.g., acquisition date, sensor, pre-processing level) and image wide statistics like average cloud cover. The latter is especially a problem because the average cloud cover statistic is one of the most used pre-selection criteria for image retrieval of big EO data but is an average over an entire image. Spatially-explicit AOI-based querying that makes use of the semantic information of each pixel in a data cube could help in making use of hidden or “dark” data in big EO databases. This could, for example, lead to retrieving more cloud-free time series or improving cloud free mosaic composition, utilising data contained in images with low average cloud cover.

A SCBIR query is visualised in Figure 1 based on the prototypical implementation. A query based on the semantic information for low cloud cover combined with low snow cover in the selected AOI would only retrieve 2 of the 4 sample images in this example, making query results better posed for following analyses. While our definition of a semantic EO data cube does not prescribe any particular level of semantic enrichment, SCBIR queries beyond cloud/snow cover are possible depending on the available image interpretation, e.g., searches for images where flooding occurred, containing a low tidal range, or where a peak in vegetation coverage occurs.

3.2. Flood Extent in Somalia Based on Landsat 8 Imagery

One of the first implementations of a semantic data cube was a study to extract surface water dynamics and the maximum flood extent as an indicator for flood risk using a dense temporal stack of 78 Landsat 8 images [48]. By using water observations of three years, areas are delineated that are prone to being flooded, as illustrated in Figure 2. In this study, the array database system rasdaman was used to instantiate a semantic EO data cube with pre-processed Landsat imagery and semantic enrichment generated with SIAM™, which can be accessed by using a self-programmed Web frontend, visually supporting the design of semantic queries. In this system the analyses are automatically translated database queries, which increase reproducibility, readability and comprehensibility for a human operator and can be conducted within a few minutes. The study showed how a generic semantic EO data cube can be used for on-the-fly information production using a very simple ruleset.

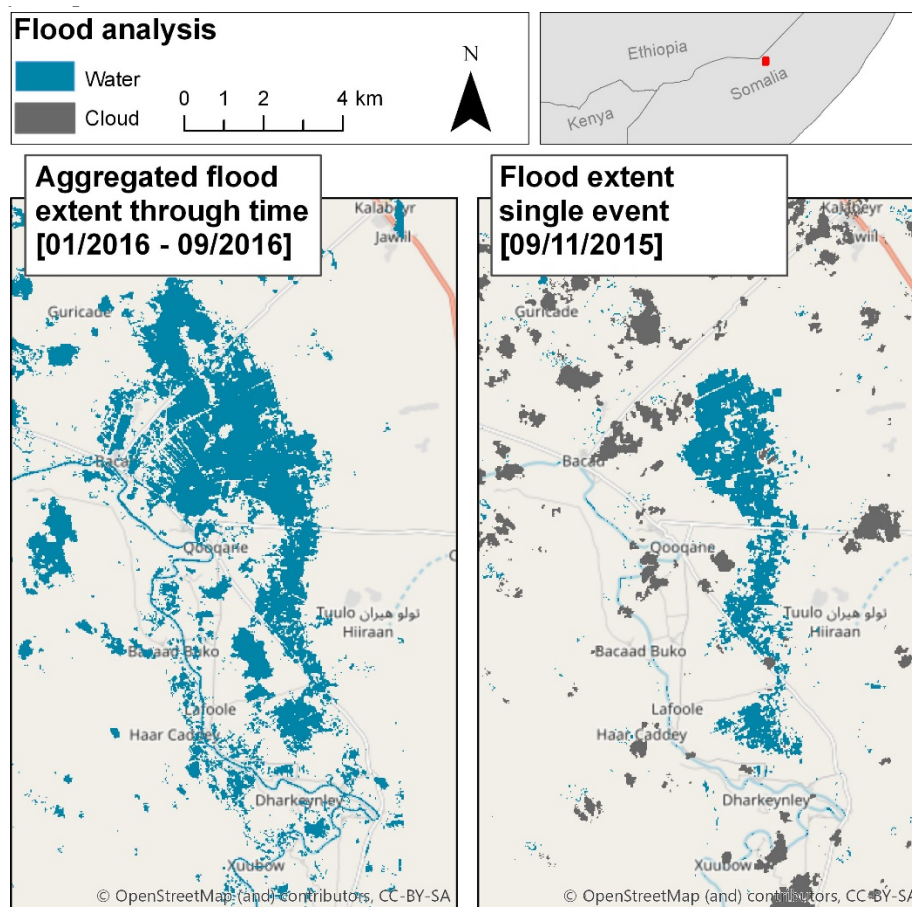


Figure 2. A flood mask generated from 78 semantically enriched Landsat 8 images over 9 months in Somalia (left) as an indicator for flood risk is compared to a single event analysis following a reported flood event in the year before (right). Both maps are the result of basic user queries using the semantic information only, without the use of additional parameters or calculations on the original data sets. Originally published as CC-BY-ND by [48], modified.

3.3. Semantic EO Data Cube along the Turkish/Syrian Border

The potential of semantic EO data cubes is demonstrated here using a proof-of-concept implementation based on ODC technology, described in detail by [49]. In this case EO refers to satellite-based remote sensing data produced by the Copernicus programme's Sentinel-2 satellites. All available Sentinel-2 data (i.e., ca. 1000 images to date) covering over 30,000 km² along the north-western Syrian border to Turkey (latitudes 36.01°–37.05°N; longitudes 35.67°–39.11°E) are continuously incorporated in an automated way including being mapped into semi-symbolic

colour names by SIAM™. The example output generated here demonstrates that traditional statistical model-based algorithms may be replaced by querying symbolic information, starting from semi-symbolic colour names with semantic associations that are not bound to a specific theme or application within a semantic EO data cube.

In March 2019 flash flooding was reported in various parts of Syria [50]. The worst flooding was reported in Idlib province, which is just south of the western most part of the study area (see Figure 3). While optical imagery is often hindered by cloud cover in rain events, a query for water-like pixels around the time of intense precipitation shows that certain flooded areas have been observed by Sentinel-2 satellites. A normalised observed surface water occurrence (SWO) over time is calculated for two spatio-temporal extents of interest, namely 15 March to 15 April for the entire study area in 2018 and 2019 (see Figure 4). The calculation of the result for each spatio-temporal extent took around 10 minutes to complete using the same hardware and software as described by [49]. The algorithm, described in pseudocode in Figure 5, can be applied to any semantic concept that exists in the semantic EO data cube. This is demonstrated in Figure 6 where the same algorithm was applied to the semantic EO data cube, but for vegetation-like pixels rather than water-like pixels for the same two spatio-temporal extents.

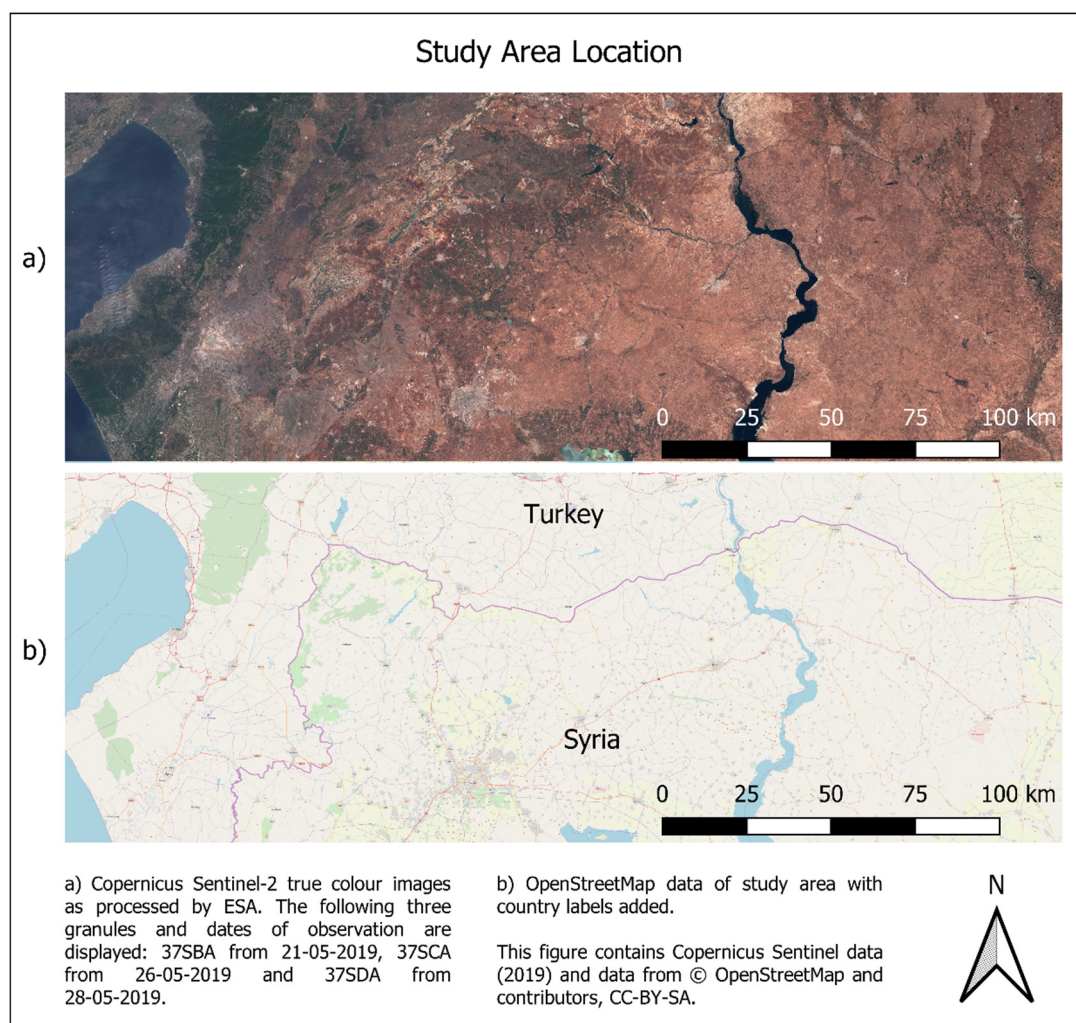


Figure 3. The spatial extent of the semantic EO data cube comprises three Sentinel-2 granules. (a) displays the true colour Sentinel-2 images as processed by the European Space Agency (ESA); (b) shows the area as represented in OpenStreetMap.

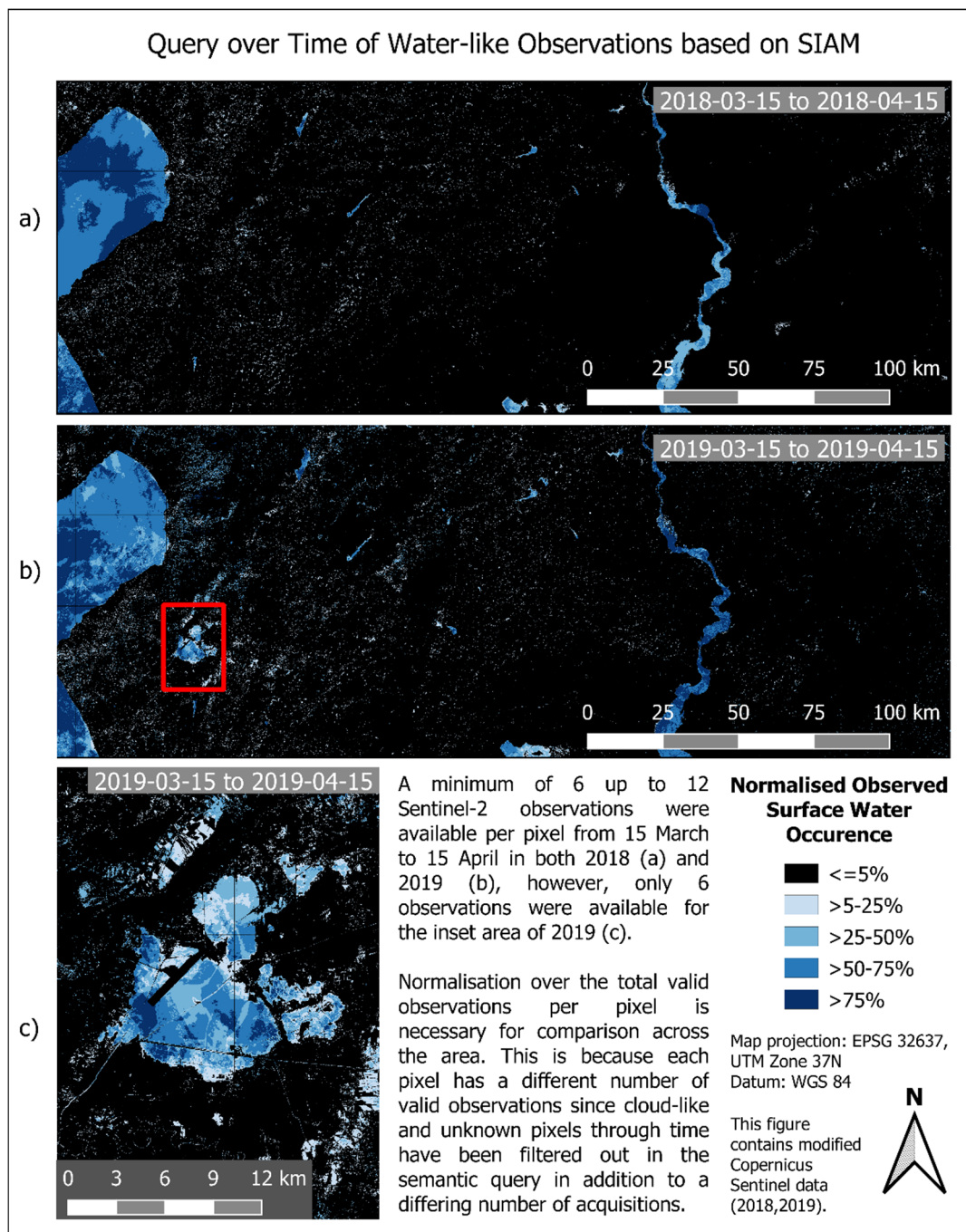


Figure 4. This figure displays the results of the semantic query for water-like observations for two spatio-temporal extents of interest. (a) Query for water-like observations from 15 March to 15 April 2018. (b) Query for water-like observations from 15 March to 15 April 2019. (c) Close-up of an area where water-like observations were present in 2019 but not in 2018.

```

1  connect to data cube
2  query data cube:
3      what: ingested tiles of SIAM's 33 semantic semi-concept granularity for Sentinel-2 data
4      where: covering [area]
5      when: acquired from [start] to [end]
6      return 3D array of per-pixel SIAM categories over time
7  split SIAM categories over time into smaller cells along spatial axes for processing
8  for each cell:
9      load into memory
10     create "water" array of same shape with all values set to 0
11     ask semantic query:
12         if water-like (>=21 and <=24), change 0 to 1
13     if invalid, ice/snow, clouds or unknown (NaN or ==25 or ==29 or ==33), change 0 to 255
14     perform time-series analysis (input: 3D "water" array from semantic query):
15         change values 255 to NaN
16         calculate "total water observations" array:
17             sum values over time (water == 1; else == 0)
18         calculate "total clean observations" array:
19             create Boolean array where values not NaN are True
20             sum values over time (True == 1; False == 0)
21         calculate "normalised occurrence" array:
22             divide "total water observations" by "total clean observations"
23         change NaN to 0 in all arrays
24         return three resulting 2D arrays:
25             "normalised occurrence"
26             "total water observations"
27             "total clean observations"
28         create GeoTiff of each returned array
29     mosaic all cell-based GeoTiffs back together for each respective analysis

```

Figure 5. Pseudocode describing how the normalised observed surface water occurrence (SWO) over time is calculated based on semi-concepts, in addition to two other outputs necessary for its calculation. The array of “total clean observations” provides the number of observations over time per-pixel after excluding cloud-like, snow-like and unknown pixels in the spatio-temporal extent of interest. Snow-like are excluded in this case based on the knowledge that there is generally no snow within the spatio-temporal extent of interest. “Total water observations” refers to the number of observations over time per-pixel that water-like spectral profiles were observed. It is the ratio between these two outputs (i.e., total divided by clean observations per-pixel) that results in the normalised observed SWO.

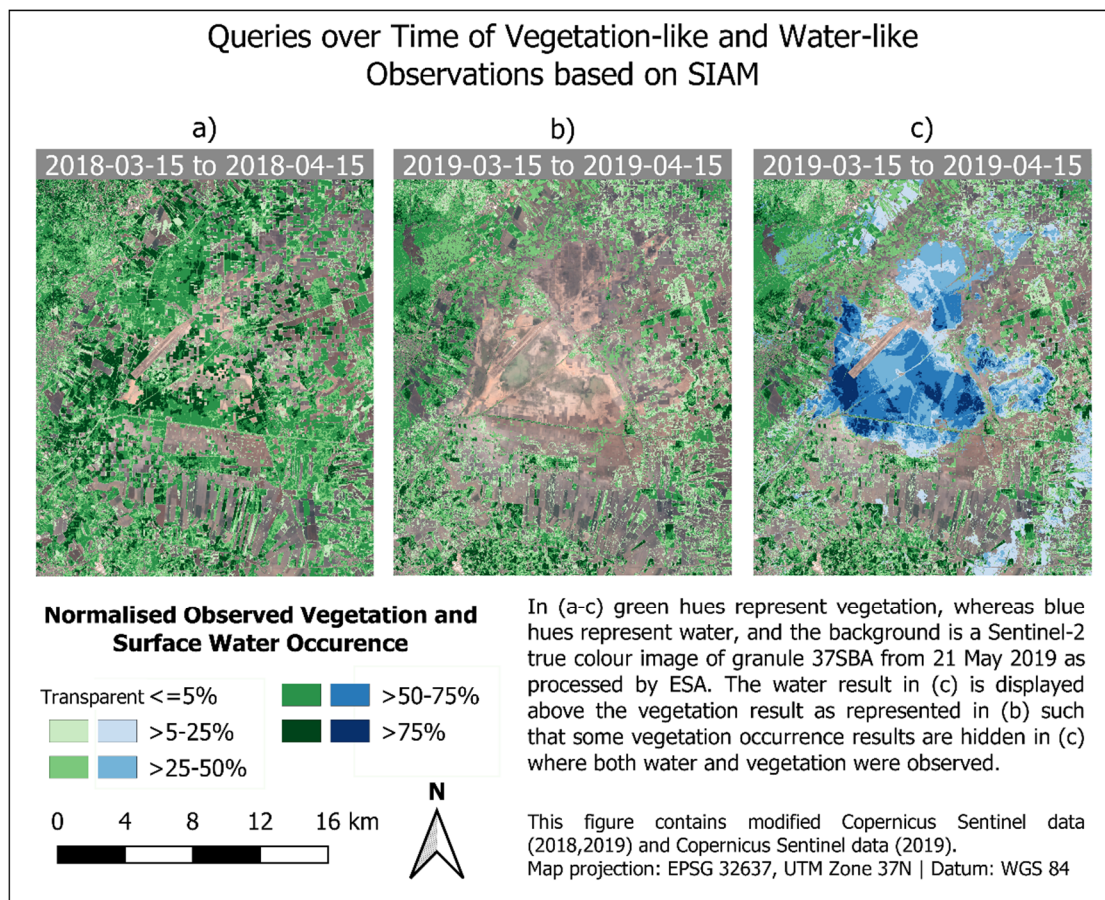


Figure 6. This figure displays the results of a different semantic query for the same two spatio-temporal extents of interest used in the query of water-like observations seen in Figure 5. (a) Normalised observed vegetation occurrence from 15 March to 15 April 2018. (b) Normalised observed vegetation occurrence from 15 March to 15 April 2019. (c) Normalised observed SWO from 15 March to 15 April 2019 overlaid above normalised observed vegetation occurrence as represented in (b).

4. Discussion and Outlook

Semantic EO data cubes are interdisciplinary in their conceptualisation, combining concepts related to image retrieval, computer vision, human cognition, semantics, world ontologies, remote sensing and more. The applied examples presented in Section 3 are brought into context of semantic EO data cubes, according to the definition and concepts provided in Section 2. Semantic EO data cubes also have the potential to be a foundational element in image understanding systems, which is discussed briefly in Section 4.4 and is a focus of on-going research.

4.1. Improvements to Data and Image Retrieval

Combining semantic enrichment with EO images has implications for EO archives, databases and the ways in which users can search for and select images [45,51]. EO data cubes already enable users to retrieve data independent of the image's spatial extent. The best-case scenario is when images processed to ARD specifications are used as the basis of an EO data cube and not just any images or quality indicators. Semantic EO data cubes enable users to search for and retrieve EO data in their spatio-temporal extent of interest based on their content, rather than image-wide statistics.

Since data and semantic enrichment are both available, SCBIR can improve ARD provision to users by expanding the possibilities that users have to retrieve images that meet their requirements. Currently a user may be interested in an area that occupies only 10% of an image. If this section of the image is cloud-free but the rest of the image is not, this image will not be returned when searching for

low average image-wide cloud coverage statistics (Figure 1). Semantic EO data cubes can provide average cloud coverage information about a user-defined AOI that could be used for data retrieval instead of aggregated image-wide metadata or statistics. Not only can users retrieve data for a given spatio-temporal extent that has low cloud cover, but based on any other category available. This means that such queries theoretically could also include searches for images containing a certain percentage of water, snow or vegetation given reliable semantic enrichment at that semantic level.

Including semantic enrichment with EO data can also serve to improve automated user-defined image composites or mosaics. The classic example is creating a cloud-free composite for a given spatio-temporal extent. As long as the semantic enrichment offers some information about cloud cover, users can retrieve cloud-free pixels for their spatio-temporal extent of interest without having to run a complex algorithm or rely on pixel-based statistics over time. A user could search for the most recent cloud-free pixels within a given spatio-temporal extent (e.g., May 2019) based on semantics instead of statistics, whereby the result could look something like Figure 3a.

SCBIR and semantically-enabled best pixel selection is even more important in the big EO era so that the data best suited for the analysis can be efficiently and effectively retrieved from huge archives in an automated way. An overview of different capabilities between file-based hubs or archives (e.g., Copernicus Open Access Hub), non-semantic EO data cubes and semantic EO data cubes is provided in Table 1.

Table 1. Feature matrix for different approaches of storing and analysing EO images.

Feature	File-Based EO Image Hubs	Non-Semantic Data Cubes	Semantic EO Data Cubes
• Image download	X	X	X
• Metadata-based search	X	X	X
• Image-wide processing	X	X	X
• AOI-based processing	-	X	X
• Fast access to imagery	-	X	X
• Time series analysis (statistical)	-	X	X
• Time series analysis (semantic)	-	-	X
• SCBIR	-	-	X
• Content-based best pixel selection for cloud-free composites	-	-	X ¹
• Generic approach with re-usable and sharable tools	-	-	X ¹

¹ Depending on the implementation level.

4.2. Semantic Content-Based Queries

The presence of a categorical interpretation for each observation allows users to pose semantic queries in EO data cubes. Semantic queries are queries about the world that exist and “make sense” regardless of whether images or data exist. They move beyond answering questions related to image retrieval (e.g., “Which data in my area of interest have less than 10% cloud cover?”) towards queries about the world (e.g., “Where and when have glaciers in the Alps grown over the last decade?”). These queries may or may not be able to be answered based on available EO data. The query space is only limited by the semantic level of enrichment and any additional information or knowledge that is available (e.g., DEM, image-derived indices).

Semantic EO data cubes enable information retrieval and semantically-enabled analysis while allowing users to better explore what is possible with available EO data in an ad hoc way beyond the confines of specific applications. There is a difference between requiring a user to know what application-specific information they want to produce from EO data, and trying to answer the question, “what is possible with these data?” [52]. For example, flooding in Turkey and Syria was known to have occurred in the spatio-temporal extent in 2019 used in queries shown in Figures 5 and 6,

but it was unclear whether optical Sentinel-2 imagery was able to capture any of it, and if so, where. Applying a query for water or water-like pixels aggregated over time, such as shown in Figure 2, is a spatially-explicit way to help answer that question. Additionally, such a query might be even more powerful if the user has limited spatially-explicit precipitation or temperature information and is unaware of any flooding that may or may not have occurred in a given area at any point in time.

It is important to emphasise that any analysis of EO data is only relevant for the snapshots in time that are available. Information derived from them may also not necessarily be valid for much of the time between acquisitions. For example, just because flooding is not observed or detected using Sentinel-2 data does not mean that flooding did not occur in a given spatio-temporal extent. Even big EO data with a high temporal sampling rate must always be interpreted keeping this in mind and is best when combined with additional information or domain knowledge.

Including semantic enrichment for each image enables semantic queries to be applied to EO data and derived information without requiring complex algorithms to process all data for a geographic area or given timespan. Even though the semantic level of the interpretations may vary amongst implementations, algorithms can access the reflectance values already associated with an interpretation that can be referenced later in the workflow, if necessary. Data-derived content-based information is available for each existing observation and can be read in a machine-based way using categories that users understand.

Working with symbolic categories instead of reflectance values means that users can work with queries that are readily understandable if the vocabulary of a community is being used, or a standard set of classes such as LCCS or similar. However, using categories means an unfortunately non-reversible data reduction, or reduction of the feature space in comparison to a multitude of bands with a higher bit depth (e.g., 48 categories stored as 8-bit data in comparison to 13 bands of 12-bit data, such as for Sentinel-2). This data reduction benefits query performance, in particular, but needs to be taken into account for every analysis. Based on our definition of a semantic EO data cube, the original data is available and accessible should users require them.

Having the original data available with categories also creates new possibilities for other applications, such as stratifying data analysis based on semantic enrichment. This could be relevant for improving sampling for machine-learning algorithms based on the frequency and distribution of certain categories through space and time. For example, samples could be stratified based on the occurrence of spectrally similar pixels by category within a study area in an attempt to mitigate sampling bias. Other analysis can also benefit from stratification based on category, such as topographic correction (e.g., certain categories will be darker in terrain shadow than others, and clouds are unaffected), or calculating indices (e.g., first querying for vegetation before calculating NDVI to avoid having to set a threshold to distinguish vegetation with the index alone).

4.3. Automated, Generic Semantic Enrichment for Big EO Data

Semantic EO data cubes are most powerful when combined with semantically rich yet generic interpretations because semantics differ between domains, applications, users and the targeted purpose of analysis. Closing the semantic gap when generating information from EO data is very difficult and goes beyond the focus of this paper (refer to [44] as a starting point on this topic), but even the simplest semantic enrichment better positions EO data cubes for analysis than ones containing no semantics at all. Any data-derived semantic information can be used as the basis of a semantic EO data cube, but generic semantic enrichment is highly extendible. It allows multiple domains to simultaneously benefit from EO data and derived information without having to reprocess huge amounts of data for every analysis. Workflows utilising the same generic, data-derived building blocks for analysis also supports increased semantic interoperability. However, big EO data necessitates data-derived interpretations that can be generated without user parameterisation (i.e., automated), are reliable and acceptable in quality and with reasonable processing times [20].

The semantic enrichment generated by SIAM™ and used in the applied examples was chosen because it is fast, fully automated, scalable to handle big EO data, sensor-agnostic and comparable between images captured at different locations and times. The limited semantic depth can be partly compensated through the availability of the temporal dimension in dense time series because the concepts are particularly stable (i.e., robust to changes in input data and imaging sensor specifications). Semantic categories that are sensor agnostic means that users can compare the semantic content of different images acquired by different sensors using the same semantic concepts. Higher level semantics can improve information generation but are generally limited to a specific theme or application. This may be beneficial in some cases and those interested in generating information can decide what is necessary for them before processing massive amounts of EO data to create a semantic EO data cube.

The examples presented in Sections 3.2 and 3.3 both queried water-like pixels based on the low-level generic semantic enrichment available over time. Even with a semi-symbolic level of semantic enrichment, queries for water-like observations could be conducted for a single acquisition or aggregated over multiple acquisitions (Figures 2 and 5). Query results shown in Section 3.3 took an additional step of excluding cloud-like and snow-like pixels and normalising the results over time for increased comparability given spatio-temporal heterogeneity of available data. The same query for two different spatio-temporal extents as shown in Figure 5 were generated within 10 minutes on relatively limited computing resources as documented by [49]. Especially in situations where timely information generation is critical, such generic implementations may be particularly useful. They can also serve in finding spatio-temporal locations interesting for further analysis using available data. Figure 6 demonstrates two semantic queries on two spatio-temporal extents based on the same semantic EO data cube, showcasing the benefit of being able to conduct various semantic queries using generic semantic enrichment.

Many other surface water occurrence algorithms and analysis for EO data exist but cannot necessarily be conducted ad hoc for user-defined spatio-temporal AOIs, are more computationally expensive, and results are not necessarily able to be queried. For example, work conducted at the European Commission's Joint Research Centre by [53] has generated various high-resolution global surface water information layers. These results provide valuable information based on EO data, but cannot be queried for content, are separate from the data that they were derived from and are limited to pre-defined temporal extents (e.g., annually). The surface water information generated by [54] or [55] for each EO observation and used in their surface water dynamics analysis could be the basis for a semantic EO data cube, but it would be semantically limited to the concept of water and does not seem to be continuously updated with newly available data (i.e., images acquired up to now) in an automated way. These implementations provide static layers, and are not currently posed to provide more dynamic, near-real-time or continuously updated results such as information about the maximum observed water extent in 2019 as it happens based on cloud-free/clean pixels.

In Figure 5 it is visible that large, permanent water bodies sometimes returned less than 100% of normalised observed surface water occurrence. This has to do with the semantic query not taking pixels associated with haze or very thin clouds into consideration, which are not necessarily water-like nor cloud-like. Queries can be improved, and more complex knowledge-based rules implemented. These proof-of-concept results demonstrate that even queries low in complexity based on low-level semantic enrichment can produce higher-level information that might be useful in certain scenarios.

4.4. Towards an Image Understanding System

While our definition does not specify applications and implementations of semantic EO data cubes, a prominent use-case is as part of an application-independent expert system, where the semantic EO data cube serves as a fact base. In an expert system, users connect rules stored within a knowledge base to a fact base to infer new information. In such a set-up, the knowledge base is continuously-augmented with rules based in domain knowledge. This allows using already existing

encoded expert knowledge or having users contribute their own knowledge. An overall architecture such as proposed by [45] consists of an image understanding sub-system in addition to the semantic EO data cube, which both makes use of already existing interpretations and feeds the fact base with newly derived, true information.

A prototype of an expert-system-based architecture is currently under development for Austria, where a semantic EO data cube serves as a backbone for user-generated semantic queries [56]. This system combines a fully automated semantic enrichment of Sentinel-2 images up to basic land cover types with a semantic EO data cube and Web interface for human-like queries based on semantic models of the spatio-temporal 4D physical-world domain. Although still under development, first results are promising and show that users are able to formulate even complex queries using the semantic pre-processing as simple building blocks to derive information at a higher semantic level than the initial building blocks.

5. Conclusions

The aim of this paper was to define what a semantic EO data cube is and what they make possible in terms of image retrieval, analysis and information production potential. Lots of EO data are being collected, yet proportionally less are being used to produce information, many domains are underserved in relation to what EO could offer, and users of EO data need to have a high level of technical competence to produce information from EO data.

By combining EO data with an interpretation for each observation of a scene, semantic EO data cubes allow users to run queries on big EO data and time-series that were not previously possible and provide imaged-derived information building blocks for analysis that are more meaningful than measured surface reflectance. Semantic enrichment enables semantic content-based image retrieval, allowing users to retrieve specific observations based on what they contain rather than image-wide statistics. Semantic queries (i.e., queries that exist independent of EO images) can be run on EO data that are at least at the semantic level of enrichment or higher without having to necessarily run complex, application-specific algorithms for each analysis. Including semantics in an EO data cube also establishes a minimal level of semantic interoperability for different analyses conducted within the same semantic EO data cube or a different implementation using the same semantic enrichment. This has implications for improving reproducibility of methods and results, especially when applying the same methods based on the same semantic enrichment to different spatio-temporal extents.

Semantic EO data cubes go beyond state-of-the-art EO data cubes by managing image-derived information together with data accessible for querying, and thus serve as initial building blocks for semantic queries. Instead of attempting to answer a specific question using EO data, semantic EO data cubes move towards exploring what questions can possibly be answered using the EO data available for a given spatio-temporal extent of interest. Analysis is only limited by the semantic enrichment included and can be extended using transparently coded rule-sets or additional information and knowledge to produce information with a higher semantic level.

We believe that semantic EO data cubes are better positioned to serve big EO data than existing EO data cube implementations, especially when containing ARD and generic, sensor-agnostic semantic enrichment that can be automatically generated in a scalable way. The potential of semantic EO data cubes is just beginning to be explored, but hopefully it is evident that there is plenty of potential yet to be discovered. Semantic EO data cubes are the foundation for big EO data expert systems, where new information can be inferred automatically in a machine-based way using semantic queries that humans understand.

Author Contributions: All authors were involved in the conceptualisation of this paper. Software used for generating semantic enrichment in the applied examples, SIAM™, was developed by A.B. Example 3.1 was provided by D.T., 3.2 by M.S. and D.T. and example 3.3 was provided by H.A. Original draft preparation was predominantly conducted by H.A. with review and editing prior to submission by D.T., M.S., S.L. and H.A.

Funding: This research has received funding from the Austrian Research Promotion Agency (FFG) under the Austrian Space Application Programme (ASAP) within the project Sen2Cube.at (project no. 866016) and from the Austrian Science Fund (FWF) through the Doctoral College GIScience (DK W1237-N23).

Acknowledgments: We would like to thank Christian Werner for his contributions in discussions about the various concepts included in this paper.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Rowley, J. The Wisdom Hierarchy: Representations of the DIKW Hierarchy. *J. Inf. Sci.* **2007**, *33*, 163–180. [[CrossRef](#)]
2. Ackoff, R.L. From data to wisdom. *J. Appl. Syst. Anal.* **1989**, *16*, 3–9.
3. Laney, D. 3-D data management: Controlling data volume, velocity and variety. In *Application Delivery Strategies*; META Group Inc.: Stamford, CT, USA, 2001.
4. Baumann, P. The Datacube Manifesto. Available online: <http://www.earthserver.eu/tech/datacube-manifesto> (accessed on 30 January 2018).
5. Lewis, A.; Oliver, S.; Lymburner, L.; Evans, B.; Wyborn, L.; Mueller, N.; Raevksi, G.; Hooke, J.; Woodcock, R.; Sixsmith, J.; et al. The Australian Geoscience Data Cube—Foundations and lessons learned. *Remote Sens. Environ.* **2017**, *202*, 276–292. [[CrossRef](#)]
6. Dhu, T.; Dunn, B.; Lewis, B.; Lymburner, L.; Mueller, N.; Telfer, E.; Lewis, A.; McIntyre, A.; Minchin, S.; Phillips, C. Digital earth Australia—Unlocking new value from earth observation data. *Big Earth Data* **2017**, *1*, 64–74. [[CrossRef](#)]
7. Killough, B. Overview of the Open Data Cube Initiative. In Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Valencia, Spain, 22–27 July 2018; pp. 8629–8632.
8. Giuliani, G.; Chatenoux, B.; Bono, A.D.; Rodila, D.; Richard, J.-P.; Allenbach, K.; Dao, H.; Peduzzi, P. Building an Earth Observations Data Cube: Lessons learned from the Swiss Data Cube (SDC) on generating Analysis Ready Data (ARD). *Big Earth Data* **2017**, *1*, 100–117. [[CrossRef](#)]
9. Ariza-Porras, C.; Bravo, G.; Villamizar, M.; Moreno, A.; Castro, H.; Galindo, G.; Cabera, E.; Valbuena, S.; Lozano, P. CDCol: A Geoscience Data Cube that Meets Colombian Needs. In Proceedings of the Advances in Computing, Cali, Colombia, 19–22 September 2017; Springer: Cham, Switzerland, 2017; pp. 87–99.
10. Cottom, T.S. An Examination of Vietnam and Space. *Space Policy* **2019**, *47*, 78–84. [[CrossRef](#)]
11. Group on Earth Observations (GEO). Digital Earth Africa: Project Overview. Available online: https://www.ga.gov.au/_data/assets/pdf_file/0008/73376/Digital-Earth-Africa.pdf (accessed on 28 May 2019).
12. Baumann, P.; Furtado, P.; Ritsch, R.; Widmann, N. The RasDaMan approach to multidimensional database management. In Proceedings of the 1997 ACM symposium on Applied computing—SAC '97, San Jose, CA, USA, 28 February–2 March 1997; ACM Press: San Jose, CA, USA, 1997; pp. 166–173.
13. Baumann, P.; Mazzetti, P.; Ungar, J.; Barbera, R.; Barboni, D.; Beccati, A.; Bigagli, L.; Boldrini, E.; Bruno, R.; Calanducci, A.; et al. Big Data Analytics for Earth Sciences: The EarthServer approach. *Int. J. Digit. Earth* **2016**, *9*, 3–29. [[CrossRef](#)]
14. Storch, T.; Reck, C.; Holzwarth, S.; Keuck, V. Code-De—the German Operational Environment for Accessing and Processing Copernicus Sentinel Products. In Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Valencia, Spain, 22–27 July 2018; pp. 6520–6523.
15. Gans, F.; Mahecha, M.D.; Reichstein, M.; Brandt, G.; Fomferra, N.; Permana, H.; Brockmann, C. The Earth in a Box: A light-weight data cube approach to empower the study of land-surface processes and interactions. *EGU Gen. Assem. Conf. Abstr.* **2018**, *20*, 9841.
16. Appel, M.; Lahn, F.; Buytaert, W.; Pebesma, E. Open and scalable analytics of large Earth observation datasets: From scenes to multidimensional arrays using SciDB and GDAL. *ISPRS J. Photogramm. Remote Sens.* **2018**, *138*, 47–56. [[CrossRef](#)]
17. Lewis, A.; Lacey, J.; Mecklenburg, S.; Ross, J.; Siqueira, A.; Killough, B.; Szantoi, Z.; Tadono, T.; Rosenavist, A.; Goryl, P.; et al. CEOS Analysis Ready Data for Land (CARD4L) Overview. In Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Valencia, Spain, 22–27 July 2018; pp. 7407–7410.

18. Committee on Earth Observation Satellites CEOS Analysis Ready Data for Land (CARD4L): Product Family Specification Optical Surface Reflectance (CARD4L-OSR) Version 4.0. Available online: http://ceos.org/ard/files/CARD4L_Product_Specification_Surface_Reflectance_v4.0.pdf (accessed on 15 June 2019).
19. Dwyer, J.L.; Roy, D.P.; Sauer, B.; Jenkerson, C.B.; Zhang, H.K.; Lymburner, L. Analysis Ready Data: Enabling Analysis of the Landsat Archive. *Remote Sens.* **2018**, *10*, 1363.
20. Sudmanns, M.; Tiede, D.; Lang, S.; Bergstedt, H.; Trost, G.; Augustin, H.; Baraldi, A.; Blaschke, T. Big Earth data: Disruptive changes in Earth observation data management and analysis? *Int. J. Digit. Earth* **2019**, 1–19. [[CrossRef](#)]
21. Pagani, G.A.; Trani, L. Data Cube and Cloud Resources as Platform for Seamless Geospatial Computation. In Proceedings of the 15th ACM International Conference on Computing Frontiers, Ischia, Italy, 8–10 May 2018; ACM: New York, NY, USA, 2018; pp. 293–298.
22. Goodman, S.N.; Fanelli, D.; Ioannidis, J.P.A. What does research reproducibility mean? *Sci. Transl. Med.* **2016**, *8*, 341ps12. [[CrossRef](#)]
23. Baraldi, A.; Tiede, D. AutoCloud+, a “Universal” Physical and Statistical Model-Based 2D Spatial Topology-Preserving Software for Cloud/Cloud-Shadow Detection in Multi-Sensor Single-Date Earth Observation Multi-Spectral Imagery—Part 1: Systematic ESA EO Level 2 Product Generation at the Ground Segment as Broad Context. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 457.
24. Matsuyama, T.; Hwang, V.S.-S. *SIGMA: A Knowledge-Based Aerial Image Understanding System*; Plenum Press: New York, NY, USA; London, UK, 1990.
25. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [[CrossRef](#)] [[PubMed](#)]
26. Sudmanns, M.; Tiede, D.; Lang, S.; Baraldi, A. Semantic and syntactic interoperability in online processing of big Earth observation data. *Int. J. Digit. Earth* **2018**, *11*, 95–112. [[CrossRef](#)] [[PubMed](#)]
27. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
28. Buckland, M.K. Information as thing. *J. Am. Soc. Inf. Sci.* **1991**, *42*, 351–360. [[CrossRef](#)]
29. Capurro, R.; Hjørland, B. The concept of information. *Annu. Rev. Inf. Sci. Technol.* **2003**, *37*, 343–411. [[CrossRef](#)]
30. Nazif, A.M.; Levine, M.D. Low Level Image Segmentation: An Expert System. *IEEE Trans. Pattern Anal. Mach. Intell.* **1984**, *PAMI-6*, 555–577. [[CrossRef](#)]
31. Hadamard, J. Sur les problemes aux derivees partielles et leur signification physique. *Princet. Univ. Bull.* **1902**, *13*, 49–52.
32. Bertero, M.; Poggio, T.A.; Torre, V. Ill-posed problems in early vision. *Proc. IEEE* **1988**, *76*, 869–889. [[CrossRef](#)]
33. Smeulders, A.W.M.; Worring, M.; Santini, S.; Gupta, A.; Jain, R. Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1349–1380. [[CrossRef](#)]
34. Berners-Lee, T.; Hendler, J.; Lassila, O. The Semantic Web. *Sci. Am.* **2001**, *284*, 28–37. [[CrossRef](#)]
35. Heflin, J.; Hendler, J. A portrait of the Semantic Web in action. *IEEE Intell. Syst.* **2001**, *16*, 54–59. [[CrossRef](#)]
36. Poggi, A.; Lembo, D.; Calvanese, D.; De Giacomo, G.; Lenzerini, M.; Rosati, R. Linking Data to Ontologies. In *Journal on Data Semantics X*; Spaccapietra, S., Ed.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 133–173.
37. Harvey, F.; Kuhn, W.; Pundt, H.; Bishr, Y.; Riedemann, C. Semantic interoperability: A central issue for sharing geographic information. *Ann. Reg. Sci.* **1999**, *33*, 213–232. [[CrossRef](#)]
38. Bahmanyar, R.; de Oca, A.M.M.; Datcu, M. The Semantic Gap: An Exploration of User and Computer Perspectives in Earth Observation Images. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2046–2050. [[CrossRef](#)]
39. Santini, S.; Gupta, A.; Jain, R. Emergent semantics through interaction in image databases. *IEEE Trans. Knowl. Data Eng.* **2001**, *13*, 337–351. [[CrossRef](#)]
40. Baraldi, A. Vision Goes Symbolic Without Loss of Information Within the Preattentive Vision Phase: The Need to Shift the Learning Paradigm from Machine-Learning (from Examples) to Machine-Teaching (by Rules) at the First Stage of a Two-Stage Hybrid Remote Sensing. In *Earth Observation*; Rustamov, R., Ed.; IntechOpen: London, UK, 2012.
41. Baraldi, A.; Boschetti, L. Operational Automatic Remote Sensing Image Understanding Systems: Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOIA). Part 1: Introduction. *Remote Sens.* **2012**, *4*, 2694–2735. [[CrossRef](#)]

42. Di Gregorio, A.; Henry, M.; Donegan, E.; Finegold, Y.; Latham, J.; Jonckheere, I.; Cumani, R. *Land Cover Classification System: Advanced Database Gateway*; Software Version 3; FAO: Rome, Italy, 2016.
43. Baraldi, A.; Durieux, L.; Simonetti, D.; Conchedda, G.; Holecz, F.; Blonda, P. Automatic Spectral-Rule-Based Preliminary Classification of Radiometrically Calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/OrbView/GeoEye, and DMC/SPOT-1/-2 Imagery—Part I: System Design and Implementation. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 1299–1325. [[CrossRef](#)]
44. Baraldi, A.; Humber, M.L.; Tiede, D.; Lang, S. GEO-CEOS stage 4 validation of the Satellite Image Automatic Mapper lightweight computer program for ESA Earth observation level 2 product generation—Part 1: Theory. *Cogent Geosci.* **2018**, *4*, 1–46. [[CrossRef](#)]
45. Tiede, D.; Baraldi, A.; Sudmanns, M.; Belgiu, M.; Lang, S. Architecture and prototypical implementation of a semantic querying system for big Earth observation image bases. *Eur. J. Remote Sens.* **2017**, *50*, 452–463. [[CrossRef](#)]
46. Dumitru, C.O.; Cui, S.; Schwarz, G.; Datcu, M. Information Content of Very-High-Resolution SAR Images: Semantics, Geospatial Context, and Ontologies. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 1635–1650. [[CrossRef](#)]
47. Li, Y.; Bretschneider, T. Semantics-based satellite image retrieval using low-level features. In Proceedings of the 2004 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Anchorage, AK, USA, 20–24 September 2004; Volume 7, pp. 4406–4409.
48. Sudmanns, M.; Tiede, D.; Wendt, L.; Baraldi, A. Automatic Ex-post Flood Assessment Using Long Time Series of Optical Earth Observation Images. *Gi_Forum* **2017**, *1*, 217–227. [[CrossRef](#)]
49. Augustin, H.; Sudmanns, M.; Tiede, D.; Baraldi, A. A Semantic Earth Observation Data Cube for Monitoring Environmental Changes during the Syrian Conflict. *Gi_Forum* **2018**, *1*, 214–227. [[CrossRef](#)]
50. Flood List News Iraq and Syria—Flooding Hits Syria Refugee Camps, Displaces Thousands Near Tigris River in Iraq—FloodList. Available online: <http://floodlist.com/asia/iraq-and-syria-floods-march-april-2019> (accessed on 30 May 2019).
51. Tiede, D.; Baraldi, A.; Sudmanns, M.; Belgiu, M.; Lang, S. ImageQuerying—Earth Observation Image Content Extraction & Querying across Time and Space. In Proceedings of the 2016 Conference on Big Data from Space (BiDS'16), Santa Cruz de Tenerife, Spain, 15–17 March 2016; pp. 192–195.
52. Willcocks, L.P.; Mingers, J. *Social Theory and Philosophy for Information Systems*; John Wiley & Sons Ltd.: Chichester, UK, 2004; ISBN 978-0-470-85117-3.
53. Pekel, J.-F.; Cottam, A.; Gorelick, N.; Belward, A.S. High-resolution mapping of global surface water and its long-term changes. *Nature* **2016**, *540*, 418. [[CrossRef](#)]
54. Tulbure, M.G.; Broich, M.; Stehman, S.V.; Kommareddy, A. Surface water extent dynamics from three decades of seasonally continuous Landsat time series at subcontinental scale in a semi-arid region. *Remote Sens. Environ.* **2016**, *178*, 142–157. [[CrossRef](#)]
55. Mueller, N.; Lewis, A.; Roberts, D.; Ring, S.; Melrose, R.; Sixsmith, J.; Lymburner, L.; McIntyre, A.; Tan, P.; Curnow, S.; et al. Water observations from space: Mapping surface water from 25 years of Landsat imagery across Australia. *Remote Sens. Environ.* **2016**, *174*, 341–352. [[CrossRef](#)]
56. Tiede, D.; Sudmanns, M.; Augustin, H.; Lang, S.; Baraldi, A. Sentinel-2 Semantic Data Information Cube Austria. In Proceedings of the 2019 Big Data from Space (BiDS'19), Munich, Germany, 19–21 February 2019; Soille, P., Loekken, S., Albani, S., Eds.; Publications Office of the European Union: Brussels, Belgium, 2019; pp. 65–68.

