

Harmonized Multiresolution Geodata Cube for Efficient Raster Data Analysis and Visualization

Lassi Lehto, Jaakko Kähkönen, Juha Oksanen and Tapani Sarjakoski

Geoinformatics and Cartography
Finnish Geospatial Research Institute (FGI)
National Land Survey of Finland
Masala, Finland

email: lassi.lehto@nls.fi, jaakko.kahkonen@nls.fi, juha.oksanen@nls.fi, tapani.sarjakoski@nls.fi

Abstract— The Data Cube concept provides a useful metaphor for management of raster geodata resources in the cloud. An initiative, called GeoCubes Finland, has been launched with the aim to facilitate access to geospatial raster data for academic research. The work is carried out in the context of a major research infrastructure development program in Finland. In the ingestion process, data sets are pre-processed into a harmonized, multiresolution cloud-based repository and brought into a common georeferencing frame, resolution levels, encoding format and tiling scheme. A custom Application Programming Interface (API) has been developed for flexible query, download and analysis of the repository's content layers. Cloud-optimized access to pre-stored resolution levels supports efficient interactive visual exploration of analysis results computed on-the-fly.

Keywords— raster data; multi-resolution; harmonisation; cloud service; visualization.

I. INTRODUCTION

Geospatial data sets are growing rapidly in number and volume. In particular, the volume of raster data sets is becoming difficult to manage. The resolution of image sensors has steadily improved and new imaging technologies have been taken into use. In addition, integrating geospatial raster data sets for analysis has become a tedious task, as the data sets typically differ in critical parameters like origin, resolution, coordinate reference system, encoding, format, etc. It is important to develop mechanisms that facilitate access to relevant data resources.

The data cube concept has emerged as a solution for organizing a repository of harmonized geospatial raster datasets [1] [2]. In computer technology, a data cube is a multi-dimensional array of values [3]. Those values represent certain facts that can be organized along various aspects. For instance, results of a vote can be considered along political party, voting districts, age or gender of candidates, year of election, etc. These aspects correspond to the axes of the multi-dimensional array and the values live in cells inside this array.

In the geospatial domain, the data cube approach was first used in the Earth Observation (EO) community for organizing vast amounts of satellite images [4]. In this application area time is a very important dimension, as satellite imagery is typically available in extensive time series. The four most

typical data cube dimensions in an EO application thus are the two geospatial coordinate axes, time and the imagery type.

In the case of a geospatial data cube, the predominant axes naturally are latitude, longitude and the possible height. A voxel-based approach for organizing truly 3D geodata is also possible [5]. The cell value represents the fact about the physical environment that the data set happens to describe. Thus, content theme can be seen as a predominant axis of a geospatial data cube. The most important benefit of organizing data sets as a data cube repository is immediate availability of the contents for integrated analysis. The approach thus aims at fulfilling the goals of the concept Analysis Ready Data (ARD) [6].

The most important positive aspects that a multidimensional data cube approach provides for geospatial application can be listed as

- A data cube forms an integrated data repository, where harmonized content layers are ready for use without troublesome data preparation steps
- Ingested datasets are aligned on pixel level and thus can be readily integrated for visualization and analysis
- A data cube can be accessed along any of its dimensions, allowing for new kinds of knowledge retrieval and spatial analysis processes
- A joint data cube repository enables data providers to deliver content through a new, easy-to-use channel, thus expanding their user base
- The simplicity of raster data processing can be more effectively and widely exploited
- By organizing a data cube as a cloud-based service with Web-friendly APIs, the potential use scenarios can be further widened

The importance of the data cube concept for geospatial application domain is also demonstrated by the fact that the Open Geospatial Consortium has initiated standardization work on the subject [7]. Another significant development is the Open Data Cube (ODC) initiative. ODC is an open source software library for organizing vast amounts of satellite imagery according to principles of data cube [8]. The software includes components for cataloguing and indexing EO resources and for ingesting those resources into a harmonized and optimized storage format for easy data retrieval. Furthermore, the ODC library provides tools for accessing and downloading content and for performing analysis operations

on it. The ODC community has also developed tools for visual exploration and statistical analysis of the data cube contents. Concrete national and regional data cube implementations based on ODC include the Australian data cube called Digital Earth Australia [9], the Swiss Data Cube [10], Colombian Cube, Vietnam Open Data Cube, and the Africa Regional Data Cube covering Kenya, Senegal, Sierra Leone, Ghana and Tanzania [11].

ODC can also be applied to data resources other than EO data. The examples mentioned in the ODC documentation include gridded data sets like elevation models, geophysical grids and other interpolated surfaces. However, most of the activities around geospatial data cubes focus on satellite image processing, in general, and on their time series applications, in particular [12].

An initiative has been launched in Finland to build a multi-resolution, cloud service-based geodata cube, called GeoCubes Finland, containing some of the most important national geodatasets [13]. The contents of the GeoCubes include data layers like Digital Elevation Model (DEM), administrative areas, land use, surface deposits and various attributes of the national forest inventory. The content layers are provided by the governmental agencies that collect and continuously maintain them. GeoCubes Finland thus differs from other geodata-related data cube implementations by not focusing on satellite imagery, but rather on other traditional geodata sets. The GeoCubes development is being carried out in the context of a large research infrastructure development program in Finland, called Open Geospatial Information Infrastructure for Research (oGIIR) [14].

In Section II, the basic principles of the GeoCubes Finland are presented. Section III discusses the custom content access API of the GeoCubes Finland data repository in more detail. In Sections IV and V, some visualization related considerations and example applications are discussed. Section VI concludes the paper.

II. GEOCUBES FINLAND

A. General

A particular feature that makes GeoCubes Finland different from other geospatial data cube implementations is its multi-resolution nature. A set of fixed resolution levels (1, 2, 5, 10, 20, 50, 100, 200, 500 and 1000 m) have been selected to store the contained data sets. The resolutions are selected to facilitate processing of analysis operations, typically run on round resolution values, and to enable easy integration with statistical and other auxiliary data sets. Opposite to resolution levels typical for web mapping schemas, such as the powers of two as in Google Maps, in GeoCubes Finland the resolution levels are selected to serve human user. The actually available set of resolution levels depend on the original accuracy of the source data set. Thus, the finest resolution of most of the GeoCubes Finland's content layers is 10 m. Only some national data sets can be reasonably be represented in 1 m resolution. These include the most accurate digital elevation model and the administrative division data sets.

Resolution levels could possibly be seen as one dimension of a multi-dimensional geodata cube, but each resolution level has an individual range for the cube's coordinate-related axes. Therefore, the resolution levels of the cube must actually be seen as a set of separate cube instances. Hence, the plural form of the cube's name: GeoCubes Finland.

In addition to industry-standard access interface to coverage data, the Web Coverage Service (WCS) [15], the GeoCubes contents can be accessed via direct file URLs and through the custom-built GeoCubes API.

B. Technical Implementation of the Repository

GeoCubes Finland's data storage is implemented as a set of Cloud Optimized GeoTIFF (COG) files [16] [17]. These files are stored on a cloud service platform, organized in directories by data provider, data set and the edition of data set. The area of the country is divided into tessellation of sixty 100 km * 100 km sized blocks, each maintained as a separate GeoTIFF file to ease processing tasks. The tessellation also facilitates parallelization of various processing steps and works as a rudimentary spatial index. All the files can always be accessed in a straightforward manner via http (HyperText Transfer Protocol) by their URLs (Uniform Resource Locator).

A GeoTIFF file arranged according to COG specification contains overviews and has its raster content organized as tiles. The overviews are ordered in sections from the coarsest to the most detailed, followed by the full resolution raster. The metadata describing the offset of each section has to be at the beginning of the file. With this file structure, a calling COG-aware client can make use of the http 'GET Range' query, which enables an indicated subsection of a file to be requested. By first requesting the metadata portion, the client can then select an appropriate range of bytes to only download the needed geospatial area of an appropriate resolution, i.e. overview level. This mechanism significantly improves the efficiency of raster data retrieval for cloud-service based applications.

The multiple resolutions are implemented in two ways, both as internal GeoTIFF overview layers and as individual external resolution-specific GeoTIFF files. Internal overviews facilitate easy content delivery, as all resolutions are contained inside as single file. Separate resolution-specific GeoTIFF files in turn enable better control of resolutions for computations run in external applications. The overview functionality, like most other computing tasks in GeoCubes Finland, is implemented using Geospatial Data Abstraction Library (GDAL) [18].

To collect the individual files as a complete view covering the whole country, the so-called Virtual Format mechanism (VRT) of GDAL is used [19]. In this approach, an XML-formatted text file is used to refer to the set of image files that constitute the integrated view. As such, a single VRT file can for instance refer to all external GeoTIFF overview files on certain resolution level, thus facilitating analysis and visualization of larger areas. In the same way, a VRT file can refer to various resolution-specific files in a certain tessellation block, providing a light-weight representation of the multi-resolution data set of that block. VRT files can be

easily transported across network connections. Because the file references are stored as absolute addresses, the VRT file can be opened in a third-party application. This approach limits downloading of raster content on the spatial area and on the resolution level the user actually needs.

III. CONTENT ACCESS API

An API has been developed for exploring, downloading and analyzing the GeoCubes Finland contents [20]. The API is designed according to the principles of RESTful Web API [21], in which the request is defined by the path components of the query URL. The API provides requests for querying the basic metadata of the repository, querying cell values on a given location, downloading content by bounding box, polygon or administrative unit and analyzing the content in terms of cell value distribution, change detection, etc. The analysis operations are run on the server platform, without downloading any content to the client side.

In all content queries, the query can be run on a desired resolution level and year. The multi-resolution structure of the GeoCubes repository gives to the user a flexible choice on speed vs. accuracy of the operation. For instance, an analysis procedure under development can be first tested on a coarse resolution level, and the real, time-consuming run on the finest resolution be carried out only when seemed appropriate. In case of analysis based on visual exploration, the resolution level, on which the analysis is run, can always be matched with the zoom level of the visualization. This way the analysis can be run in roughly constant time and the interactivity level of the application can be kept stable.

The general scheme of the access API can be described as follows:

```
what to do /
  on which resolution level /
    with which content layer /
      where /
        when /
          how
```

As an example, the query for the cell value on a given location becomes as:

```
legend/500/mvmi-paatyyppi/340500,6695000/2009
```

where mvmi-paatyyppi (in Finnish) is one of the themes (forest inventory data) of GeoCubes.

Another query would request download of a content layer inside the given list of municipalities ('kuntajako' in Finnish), from the given resolution level and year:

```
clip/100/corine/kuntajako:734,761,834,433,224,444,927/2000
```

The resulting data set, requested using the GeoCubes Web client, is shown in Figure 1.

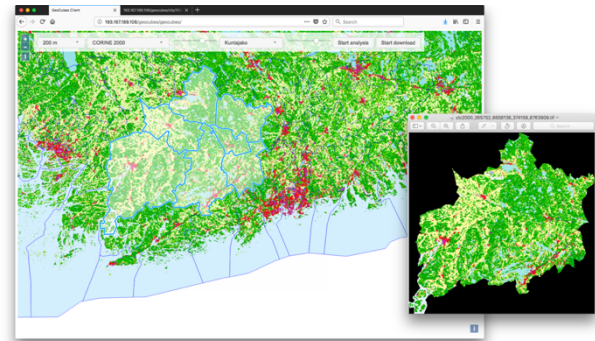


Figure 1. A CORINE data set downloaded via GeoCubes API using administrative units' boundaries as selection area.

Downloading of a DEM from two blocks in VRT format with multiresolution content would go as:

```
clip/20/km10/blocks:300000,6900000,300000,6800000/2018/vrt/mr
```

An example of an analysis would be change detection between CORINE versions 2000 and 2012 in land use type fields ('pellot' in Finnish) inside the given bounding box:

```
analyse:changedetect/10/corine:Pellot/bbox:225700,6660000,494300,6740000/2000,2012
```

The result of the above analysis is an image depicting the changes in red (removals) and green (additions).

Another analysis would determine the distribution of cell values inside the two indicated counties ('maakuntajako' in Finnish):

```
analyse:distribution/500/mvmi-maaluokka/maakuntajako:04,06/2015
```

The result is a list of existing cell values together with their frequencies. The analysis results visualized in the GeoCubes Web client using the D3.js library [22] is shown in Figure 2.

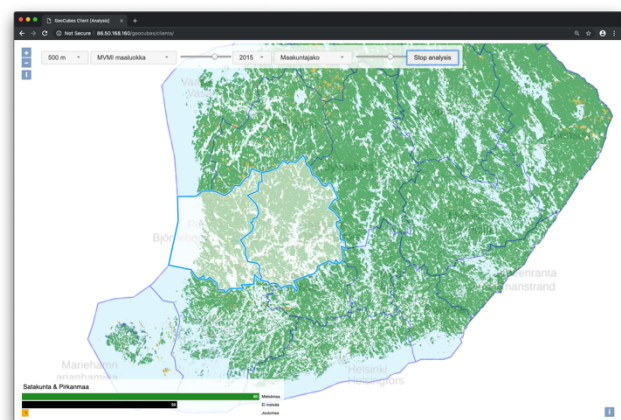


Figure 2. GeoCubes analysis results showing distribution of data values inside the selected area.

As an auxiliary data source for analysis purposes, the Finnish administrative areas in four different levels are also available in vector form in the GeoCubes repository.

The GeoCubes API has been implemented as a Web service using Django Web framework [23], together with Python-based service side scripts making an extensive use of the Python API of GDAL. In addition to the custom API, the most relevant service interfaces standardized by the Open Geospatial Consortium are also supported. The main components of the GeoCubes Finland platform are shown in Figure 3.

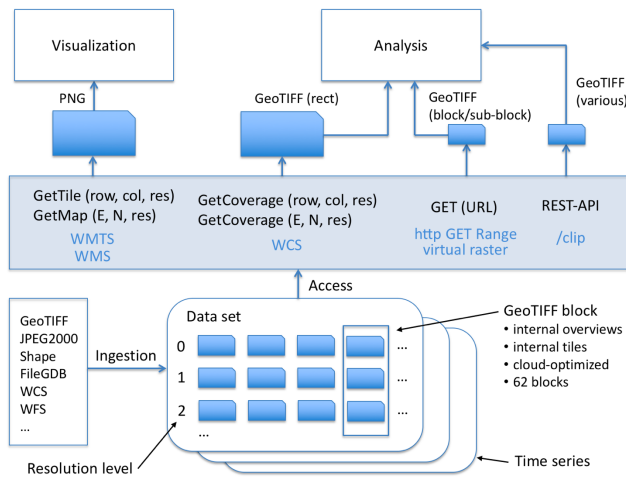


Figure 3. The main system components of the GeoCubes Finland platform.

IV. VISUALIZATION CONSIDERATIONS

For basic visualization of the content layers, a Web Map Tile Service (WMTS) is available on the GeoCubes platform. The service makes use of a Web Map Service (WMS) that uses a VRT file combining the original GeoTIFF files as its source. This arrangement enables both ad hoc visualizations through WMS and cache-based static image delivery via WMTS. In the case of the WMTS service, the pre-rendered tiles are available on the fixed resolution levels of the GeoCubes. For best visual presentation, the client should apply scale levels that correspond to those resolutions. The WMS implementation of GeoCubes platform is based on MapServer and the WMTS service on MapProxy.

Visualization of administrative areas poses a specific challenge. Because areas are represented as raster data to facilitate analysis with other content layers, boundary lines cannot be used to separate an area from the neighboring areas. The only way to reliably display the administrative division of the country in a legible manner is to ensure that neighboring areas are always presented with clearly distinguishable colors. The Finnish lowest level administrative division has more than 300 municipalities. Thus, to achieve an appropriate data set with cell values corresponding to the real municipality codes, a raster with 16-bit cell values was created. However, for Web browser-based visualization, a 256-colour PNG

image is required. A custom software module was developed that creates random color components for a 256-colour RGB color table and applies the colors to municipalities so that the three color components of a municipality always differ more than 30 units from all of the neighboring areas' color components. The result is a bright and colorful municipality map, in which all areas can be reliably distinguished (Figure 4).

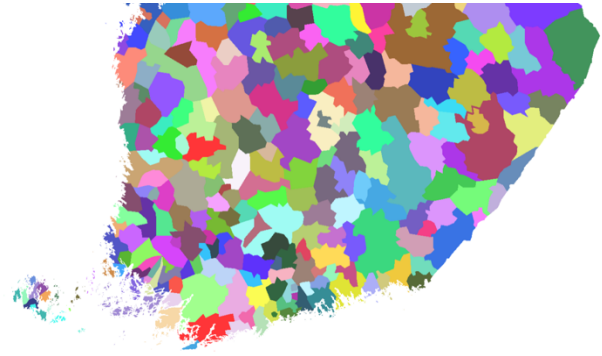


Figure 4. A municipality map with well distinguishable colors.

Another example of enhanced visualization tested in GeoCubes Finland is the locally stretched DEM visualization. The terrain of the country is mostly flat, higher elevations existing only in the northern Lapland. A consistent, stable country-wide DEM visualization thus depicts flat areas with very limited color scale, making it impossible to recognize subtle height differences. A dynamic visualization module was developed for GeoCubes that stretches every individual view requested by the client to full color scale. As the GeoCubes data repository has multiple resolution levels, the visualization processing can be performed efficiently in constant time, independently of the presentation scale of the client. The two images in Figure 5 depict the difference between fixed and dynamic DEM visualization on a particularly flat area in the Finnish Ostrobothnia.

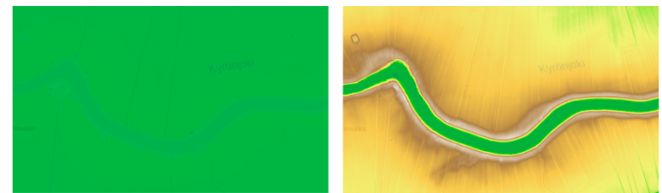


Figure 5. A static DEM visualization compared with a dynamic, locally stretched graphic scale.

A 3D example of visualization of GeoCubes' contents is depicted in Figure 6. Here, DEM data is extracted from the GeoCubes API around possible crater locations, where a meteorite impact is deemed to be behind the peculiar rounded land form. The multiresolution contents of the GeoCubes repository supports effective 3D visualization of crater areas of various sizes.

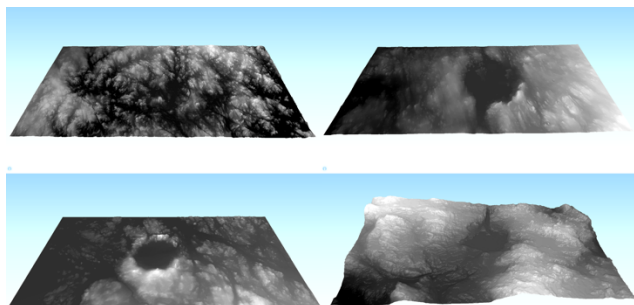


Figure 6. Possible meteorite impact craters visualized, based on GeoCubes DEM. Big scale differences in visualizations are supported by GeoCubes' multiresolution contents.

V. APPLICATION EXAMPLES

The analysis processes can be effectively run on different resolution levels and their results are readily available for visual exploration over the whole range of scales supported by the geodatacube's resolutions. The GeoCubes API has been designed to support access, querying and analysis of the geodatacube's contents on the resolution level appropriate for the actual use situation.

A. Field Inspection

As an example of using GeoCubes API, a field inspection of repository cell values has been tested. In this approach, the inspector moves around on the terrain and constantly receives values of the GeoCubes layer of interest into his cell phone. This way, the inspector can compare the environment around him with the stored category values, for instance, he can determine the correctness of forest type information – and do that on all available resolution levels of the repository.

In the developed pilot implementation, the position of the inspector is recorded with an application called Owntracks. The location is reported back to a server (Eclipse Mosquitto) [24] implementing the ISO-standardized Message Queuing Telemetry Transport (MQTT) -protocol. The user is presented with a map-based application (Leaflet with Realtime extension). When the map application gets notice from the MQTT server that the user has moved to a new location, it then send a request to the 'legend' operation of the GeoCubes API together with the location information. As a result, it will get cell values from the requested layer on all available resolution levels and can compare them with the reality around him. The pilot's architecture is presented in Figure 7.

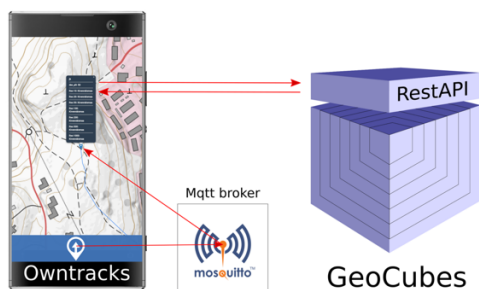


Figure 7. A field inspector accessing GeoCubes cell values using Mosquitto server.

B. Route Finding

In another analysis example, the GeoCubes DEM layer is used for route finding for a forest vehicle that has certain limits for movement in steep slopes. DEM data for study area was acquired by accessing the GeoCubes API's 'clip' operation from within the QGIS application [25]. Then, analysis functions available in QGIS were used to compute slope values for the area and these were used as the cost surface for route finding. Areas too steep for the vehicle were excluded from the computation. The case study demonstrates the use of the GeoCubes' content for analysis using tools outside the GeoCubes platform. The same analysis could also be run in different resolution levels, depending on the needs of the application. The resulting route alternatives are depicted in Figure 8 using three.js, a 3D JavaScript library [26].

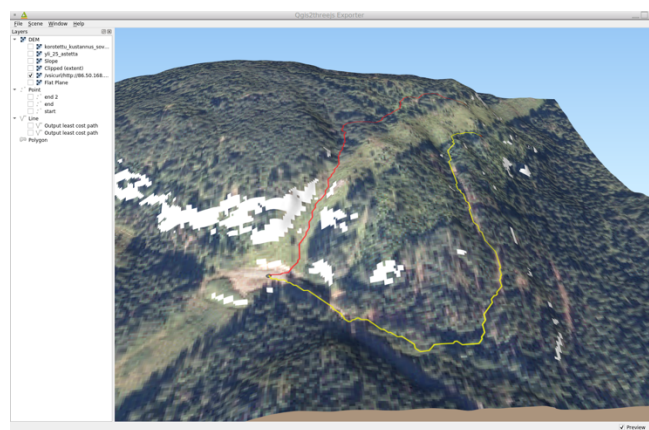


Figure 8. Route finding in QGIS with DEM from GeoCubes using slope as cost surface. White areas denote terrain that has been classified to be too steep and is excluded. The red and yellow lines denote found alternative routes.

VI. CONCLUSIONS

One of the major obstacles for wider use of geospatial raster data sets in different research and analysis scenarios is the work required for pre-processing the available data sets. This often involves coordinate reference system transformations, resampling procedures, coding system translations, integration of map sheet-based data files, etc. To facilitate the introduction of geospatial data sets into research processes in a multidisciplinary setting, a harmonized, easy-to-access data storage would be really beneficial. In the GeoCubes Finland initiative, this kind of approach has been taken.

A representative set of geospatial data sets with national coverage has been ingested into the cloud service-based data repository. The pre-processing phase involves operations like rasterizing vector-formatted source data sets, resampling of source data into the set of standardized GeoCubes resolution levels, harmonization of value coding systems between different data set editions, encoding of the resulting raster content into the common cloud-optimized image format and storing it to the cloud repository, both as binary images files and as textual VRT representations.

By storing the raster data sets in multiple resolution levels, the GeoCubes repository can support certain use cases very efficiently. These include interactive visual exploration of on-the-fly analysis results and testing of an analysis procedure on coarse resolution levels before launching an accurate analysis on detailed levels. The result layer of an analysis can be configured as a new GeoCubes content layer and thus be run dynamically by the calling visualization client. The analysis procedure will always select the resolution level closest to the visual scale, thus enabling nearly constant processing times to be maintained.

Further work on the GeoCubes Finland platform will focus on adding new layers to the repository's data collection, developing modular analysis functions for server-side processing, and enhancing the GeoCubes Web client. A QGIS plugin module for accessing the GeoCubes API is also going to be developed. More user testing will be carried out to gather feedback and guidance for further development of the platform.

ACKNOWLEDGMENT

We made use of computing resources provided by the Open Geospatial Information Infrastructure for Research (oGIIR, urn:nbn:fi:research-infras-2016072513) funded by the Academy of Finland, and CSC – The IT Center for Science Ltd.

REFERENCES

- [1] P. Baumann, The Datacube Manifesto, 2017. http://earthserver.eu/sites/default/files/upload_by_users/The-Datacube-Manifesto.pdf [retrieved: Jan 2020].
- [2] L. Lehto, J. Kähkönen, J. Oksanen, and T. Sarjakoski. GeoCubes Finland - A Unified Approach for Managing Multi-resolution Raster Geodata in a National Geospatial Research Infrastructure. In: *GEOProcessing 2018, the Tenth International Conference on Advanced Geographic Information Systems, Applications and Services*, March 25-29, 2018, Rome, Italy. ISBN: 978-1-61208-617-0, pp. 18-22.
- [3] Wikipedia, Data cube. https://en.wikipedia.org/wiki/Data_cube, 2020 [retrieved: Jan 2020].
- [4] A. Lewis et al., "Rapid, high-resolution detection of environmental change over continental scales from satellite data – the Earth Observation Data Cube". *International Journal of Digital Earth*, vol. 9, Iss. 1, 2016, pp. 106-111.
- [5] U. Pyysalo and T. Sarjakoski, "Voxel approach to landscape modelling". *The International Archives of the Photogrammetry and Remote Sensing*, July 2–11, 2008, Beijing, China, XXXVII(B4/1), pp. 563–568.
- [6] G. Giuliani et al. Building an Earth Observations Data Cube: lessons learned from the Swiss Data Cube (SDC) on generating Analysis Ready Data (ARD), Big Earth Data, 2017, 1:1-2, pp. 100-117, DOI: 10.1080/20964471.2017.1398903.
- [7] OGC, Datacube Domain Working Group Charter, https://external.opengeospatial.org/twiki_public/pub/CoveragesDWG/Datacubes/17-071_Datacube-DWG_Charter.pdf, 2017 [retrieved: Jan 2020].
- [8] ODC, Open Data Cube Home Page, <https://www.opendatacube.org>, 2019 [retrieved: Jan 2020].
- [9] A. Lewis et al., "The Australian Geoscience Data Cube – Foundations and lessons learned". *Remote Sensing of Environment*, 2017, ISSN 0034-4257, <https://doi.org/10.1016/j.rse.2017.03.015>, pp. 276-292.
- [10] Swiss Data Cube (SDC) Home Page, <https://www.swissdatacube.org>, 2017 [retrieved: Jan 2020].
- [11] Africa Regional Data Cube Home Page, <http://www.data4sdgs.org/index.php/initiatives/africa-regional-data-cube>, 2016 [retrieved: Jan 2020].
- [12] A. Lewis et al., 2016. Rapid, high-resolution detection of environmental change over continental scales from satellite data - the Earth Observation Data Cube. *Int. J. Digital Earth* 9 (1), pp. 106–111.
- [13] L. Lehto, J. Kähkönen, J. Oksanen, and T. Sarjakoski. Supporting Wide User-Base in Raster Analysis - GeoCubes Finland. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4, pp. 329-334. <https://doi.org/10.5194/isprs-archives-XLII-4-329-2018>.
- [14] oGIIR, Open Geospatial Information Infrastructure for Research Home Page, <http://ogiir.fi>, 2020 [retrieved: Jan 2020].
- [15] OGC, Web Coverage Service. <http://www.opengeospatial.org/standards/wcs> [retrieved: Jan 2020].
- [16] GeoTIFF, GeoTIFF home page, <http://trac.osgeo.org/geotiff/>, 2019 [retrieved: Jan 2020].
- [17] Cloud Optimized GeoTIFF (COG) Home Page, <https://www.cogeo.org>, 2019 [retrieved: Jan 2020].
- [18] GDAL, Geospatial Data Abstraction Library Home Page, <http://gdal.org>, 2020 [retrieved: Jan 2020].
- [19] GDAL, Virtual File Tutorial, https://www.gdal.org/gdal_vrtut.html, 2019 [retrieved: Jan 2020].
- [20] L. Lehto, J. Kähkönen, J. Oksanen, and T. Sarjakoski, 2019. Flexible Access to a Harmonised Multi-resolution Raster Geodata Storage in the Cloud. *GEOProcessing 2019, the Eleventh International Conference on Advanced Geographic Information Systems, Applications and Services*, Feb 24-28, 2019, Athens, Greece. ISBN: 978-1-61208-687-3, pp. 26-28.
- [21] RESTful API Tutorial, <https://searchmicroservices.techtarget.com/definition/RESTful-API>, 2020 [retrieved: Jan 2020].
- [22] M. Bostock, 2018. Data-Driven Documents, D3.js Home Page. <https://d3js.org> [retrieved: Jan 2020].
- [23] Django Software Foundation, 2018. Django Home Page. <https://www.djangoproject.com> [retrieved: Jan 2020].
- [24] R. A. Light, "Mosquitto: server and client implementation of the MQTT protocol," *The Journal of Open Source Software*, vol. 2, no. 13, May 2017, DOI: 10.21105/joss.00265.
- [25] QGIS Home Page, <https://qgis.org/en/site/>, 2020 [retrieved: Jan 2020].
- [26] three.js Home Page, <https://threejs.org>, 2020 [retrieved: Jan 2020].