IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium | 979-8-3503-6032-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/IGARSS53475.2024.10641643

AN EARTH OBSERVATION DATA ECOSYSTEM TO ENHANCE ENVIRONMENTAL MONITORING AND SOCIETY'S RESILIENCE IN CYPRUS AND THE EMMENA REGION

Stelios P. Neophytides[a], Michalis Mavrovouniotis[a], Nikos Christoforou[a], Thanassis Drivas[d], Marinos Eliades[a],Christiana Papoutsa[a], Rodanthi-Elisavet Mamouri[a], Konstantinos Fragkos[a] , Dragos Ene[a], Felix Bachofer[b], Thomas Krauss[b], Egbert Schwarz[b], Johannes Bühl[c], Patric Seifert[c], Gunter Schreier[b], Albert Ansmann[c], Charalambos Kontoes[d], Diofantos G. Hadjimitsis[a]

[a]ERATOSTHENES Centre of Excellence, Limassol, Cyprus.
[b]German Aerospace Center (DLR), Cologne, Germany.
[c]Leibniz Institute for Tropospheirc Research (TROPOS), Leibniz, Germany.
[d]National Observatory of Athens, Beyond Centre of EO Research and Satellite Remote Sensing, Athens, Greece.

ABSTRACT

The rapid growth of Earth Observation (EO) and Remote Sensing (RS) data has underscored the critical need for identifying optimal solutions to effectively manage EO Big Data. This entails simplifying data sharing and facilitating adaptation across multidisciplinary applications to better serve the research community. Various architectures and structures have been developed to manage and deploy these data in an analysis-ready format. In this study, we provide a concise overview of an advanced EO Big Data infrastructure located in Limassol, Cyprus, comprising diverse data sources acquired from an acquisition station, an atmospheric ground base station, and various living labs. Additionally, we present the EO data ecosystem of Cyprus that is specifically designed to efficiently store the aforementioned data.

Index Terms— Earth Observation, Remote Sensing, Big Earth Data, Analysis Ready Data, Data Cube

1. INTRODUCTION

Over the last decades, there is an increasing number of Earth observation (EO) satellite missions aiming to enhance the environmental monitoring of the planet's surface [1, 2]. Satellites have proven their effectiveness in space-based measurements [3]. The evolution of space technologies generates enormous volume of data, known as Big Earth Data. For instance, until 2018 the generated data from the first three Sentinel missions (1, 2 & 3), Moderate Resolution Imaging Spectroradiometer's (MODIS) Aqua and Terra Units, and Landsat 7 and 8 equals to ten terabytes of data every day, with a total quantity of data acquired covering our planet expected to be around five petabytes [4]. According to European space agency (ESA)'s annual report[1] for 2021, 413 petabytes were downloaded from Open Access Hub users, demonstrating the increased demand and interest from the scientific community.

As a consequence, infrastructures are essential to cope with the complexities of Big Data which are characterized by the 5-Vs (volume, velocity, value, variety, and veracity) [5]. For example, the development of novel raster files types like cloud optimized GeoTIFFs (COG) and cloud optimized raster encoding (CORE), which are compressed raster file types, enables efficient storage of large amounts of EO big data and consequently faster analysis on cloud infrastructures [6].

Traditional and modern techniques, such as the extract-load- transform, extract-transform-load, five-layer-fifteen-level, and reverse extract-transform-load techniques, have been used as big EO data pipelines aiming to: 1) discover important insights; and 2) collect EO big data in near real-time (NRT) [7].

Different frameworks and technological advances have been developed and accomplished, respectively, in order to organize the Big Data obtained from different sources (i.e., space, airborne and terrestrial). For example, EO Data Cubes - a multi-dimensional structure - which maintains data in a grid structure sorted in time and grouped by product type. There are several paradigms of operational Data Cubes around the world such as the Swiss DC [8], Austrian Semantic DC [9], Brazilian DC [10] and Australian Data Cube [11]. Other solutions such as the Google Earth Engine [12] are expected to become the panacea in the upcoming years because of their flexibility.

EO satellites are generating raw data that require preprocessing in order to be appropriate for analysis. For ex-

[1] https://www.esa.int/About_Us/ESA_Annual_Report_2021

Authorized licensed use limited to: Cyprus University of Technology. Downloaded on May 20,2025 at 08:04:34 UTC from IEEE Xplore. Restrictions apply.

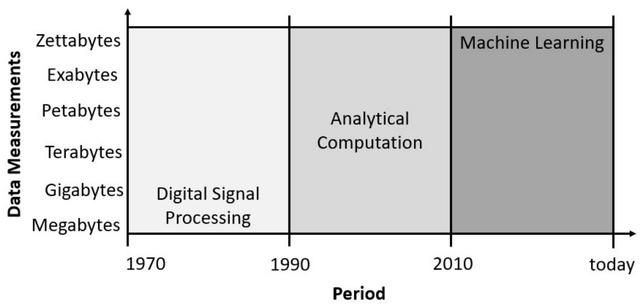

Fig. 1. The evolution of information extraction methods.

ample, Sentinel-2 data in some cases suffer from high levels of cloud coverage that require masking, whereas for NASA's MODIS Level-3 data, no pre-processing is required. Information extraction and data cleaning are essential steps in the pre-processing stage when acquiring Big EO Data to transform the data in an analysis-ready data format.

Over the last decades [13], different methodologies emerged for information extraction in Big EO Data as presented in Fig. 1. For example, initially, in which the available data was measured in megabytes, digital signal processing methods were used on statistical models. Later on, analytical computation methods on physical models have been used. Nowadays, due to the enormous amount of available data which is measured in zettabytes (or even higher), Artificial Intelligence techniques such as traditional shallow and deep learning methods, are used in data models and they are capable of extracting meaningful information [14].

The main objective of this work is to describe the designed EO Big Data infrastructure in Limassol, Cyprus , located in the Eastern Mediterranean, Middle East, and North Africa (EMMENA) region [15].

The rest of the paper is structured as follows. Section 2 describes the current EO Big Data infrastructure of Cyprus, data acquisition pre-processing chains, and preliminary metrics on the efficiency of Cyprus ecosystem. Section 3 demonstrates two case studies using the Cyprus EO data cube. Finally, Section 4 concludes this work.

2. EO BIG DATA INFRASTRUCTURE OF CYPRUS

In this section, a description of the EO big data infrastructure of Cyprus is given, including the handling of NRT satellite-derived data, EO archived data, and data obtained from several other sensors. The complete infrastructure is presented in Fig. 2. Cyprus is a country with a unique geostrategic position at the crossroads of three continents and with excellent weather conditions that can contribute significantly to various sectors of EO in the EMMENA region. The designed EO Big Data infrastructure consists of different components such as the EO data acquisition station (EODAS) and two living labs

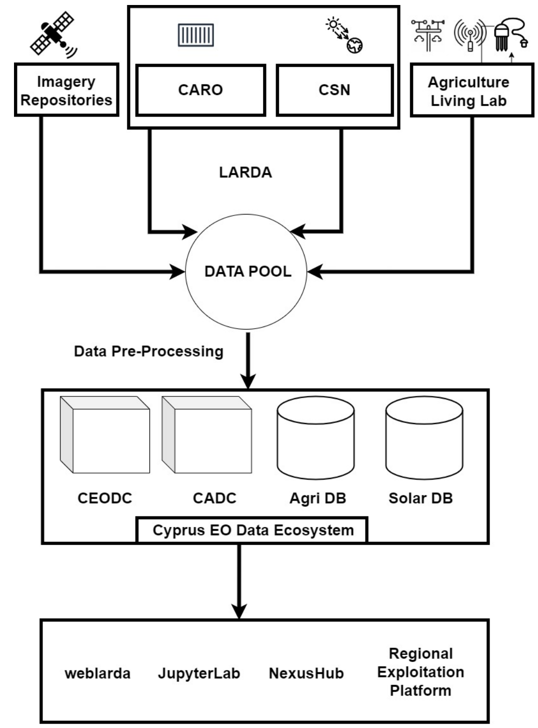

Fig. 2. The structure of EO Big Data Infrastructure in Cyprus.

(LLs). The agriculture LL consists of different in-situ sensors such as soil moisture, evapotranspiration, and carbon dioxide. Atmosphere LL consists of a Ground Based Station (GBS). More details are given in subsection 2.1.

Well-established Deep Learning methodologies such as convolutional neural networks are responsible for data-driven information extraction in order to reduce the size of data by disregarding unnecessary features [14]. These data models require almost no prior knowledge of RS to extract features but require big data for training them [13]. More details of the data pool are given in subsection 2.2.

Regarding the storage, management, and analysis of the resulting ARD, two different data cube types are implemented: 1) the Cyprus EO Data Cube (CEODC), and 2) the Cyprus atmosphere Data Cube (CADC). In addition, an agriculture LL database and a solar database are maintained. These four data storage structures together form the Cyprus EO Data Ecosystem (CEODE) which is described in subsection 2.3 later on. Data gathered in the CEODE differ in format and structure, thus there is a need for developing different data storage structures.

2.1. Infrastructure Components

The infrastructure consists of: 1) EODAS (expected to be installed in Nicosia, Cyprus), 2) GBS (established in Limassol, Cyprus), and Cyprus Solar Network (CSN) (consisting of 5 stations across Cyprus).

Authorized licensed use limited to: Cyprus University of Technology. Downloaded on May 20,2025 at 08:04:34 UTC from IEEE Xplore. Restrictions apply.

The EODAS is equipped with a three-band feed and capabilities for simultaneous reception of all three bands (S, X and Ka) making it suitable to support current and future missions. The size of the reflector allows the station to successfully meet the stringent criteria set up by international and national bodies in terms of service availability and reception margins. The electromechanical structure of the antenna allows tracking and acquisition of data from satellites orbiting as low as 400 kilometers above the surface of the Earth. These NRT-acquired satellite EO data from EODAS will be first pipe-lined in a batch and stream architecture to a data pool and then stored to our CEODC.

The GBS consists of a precipitation disdrometer, a microwave radiometer,a PollyXT Raman LiDAR (Raman Light Detection and Ranging - RL), a Ka-band cloud radar, a Doppler lidar, and a sun-lunar photometer. These instruments are gathering real-time profiling data over Limassol and they are stored in 6-hour files. The GBS enables the observation of the state and evolution of the troposphere and lower atmosphere, such as the distribution of particles from dust events, wildfires in Cyprus or neighboring countries, their impact on clouds and precipitation, the observation of local wind fields and turbulence, and other phenomena affecting the atmosphere, including insect migration. The LACROS research data analyser (LARDA) [16], developed by TROPOS, processing chain is responsible to handle the data derived from the GBS. LARDA contains processing tasks like remote reading of profiling data from complex 3D data structures and data processing methods like interpolation.

A variety of sensors are used by the CSN, a ground-based network of stations for tracking solar energy and solar ultraviolet (UV) radiation. Pyranometers are used to measure total solar irradiance, while UV radiometers are used to measure solar UV irradiance. CSN's central station has a spectral radiometer which acts as the network's reference instrument. In addition, an optical laboratory for calibration of the instruments will be established at the central station[17].

2.2. Data Pre-Processing

Data cleaning is performed as soon as the data arrives in the data pool from the different sources. The raw satellite data are examined concerning their adequacy to be used for analysis in the following processing steps.

Specifically, pre-processing analysis for Sentinel-2 data is applied, such as resampling of all the bands to 10 meters and a reprojection to EPSG:32636. In contrast, the pre-processing analysis for Sentinel-1 GRD consists of radiometric calibration, speckle filtering, and geometric correction to produce the backscatter product. Sentinel-1 pre-processing chain produces VV and VH polarizations. MODIS data are retrieved directly from archives in an ARD format. Thus, the only pre-processing applied is the reprojection from the sinusoidal grid to EPSG:32636. All the aforementioned data types at the end of the pre-processing are converted into COG.

2.3. Cyprus Earth Observation Data Ecosystem

The CEODE consists of the CEODC, the CADC, and the NoSQL-based databases of agriculture and solar.

For the CEODC, the open data cube framework is adapted where a PostGIS database system is designed to index different raster datasets from the local storage. CEODC contains long-time series satellite imagery data from ESA's Sentinel-1 and Sentinel-2 (all the 13 bands are indexed), as well as 14 different NASA's MODIS products such as vegetation indices 16-day L3 global 250 m., vegetation continuous fields yearly L3 global 250 m., leaf area index/FPAR 8-Day L4 global 500 m., burned area monthly L3 global 500 m., and Net Evapotranspiration 8-Day L4 Global 500 m.

For both Sentinel missions, a fallback connection to three main EO data providers' repositories (i.e., Copernicus Data Space Ecosystem, ONDA, and Theia) is established through the Earth Observation Data Access Gateway. This assures the redundancy of the system. MODIS data are downloaded via a connection with NASA's earthdata.

The satellite data related to atmospheric conditions from satellites like NASA's Calipso and ESA's AEOLUS are stored in the atmospheric online analytical processing (OLAP) cube. OLAP cubes – known as hypercubes when 3 or more dimensions exist – are data structures with the capability of online processing as well as for organizing data by different criteria [18]. For example, PollyXT lidar measures different aspects of aerosol (e.g., dust and non-dust particles), by separating the day in 6-hour time slices in a 7.5m height resolution, thus, OLAP structure enables the organization of such data in a hypercube.

For the data related to agriculture a NoSQL database is developed with MongoDB, named AgriDB. Meteorological data, such as humidity, wind speed, temperature, rainfall, and so on, are acquired from 43 different sensors in Cyprus and they are stored on a daily basis. In addition, normalized difference vegetation index (NDVI) values from Sentinel-2 are obtained together with the aforementioned data to calculate crop evapotranspiration.

Shortwave and longwave downward irradiances from CSN's Limassol instrumentation, including global and diffuse horizontal irradiances, direct normal irradiance, longwave irradiance, and erythemal UV irradiance. Collected data are sent to a storage server in NRT as CSV files and then saved to the SolarDB, similar to AgriDB. Sky-camera images are stored online on the manufacturer's cloud infrastructure and retrieved through an File Transfer Protocol (FTP) connection once per day to the data center. The instruments installed at stations affiliated with the Meteorological Service of Cyprus are cataloged in the Meteorological Service's database.

The CEODE is planned to be the backbone of a Regional Exploitation Platform, a JupyterLab as an analytical frame-

Authorized licensed use limited to: Cyprus University of Technology. Downloaded on May 20,2025 at 08:04:34 UTC from IEEE Xplore. Restrictions apply.

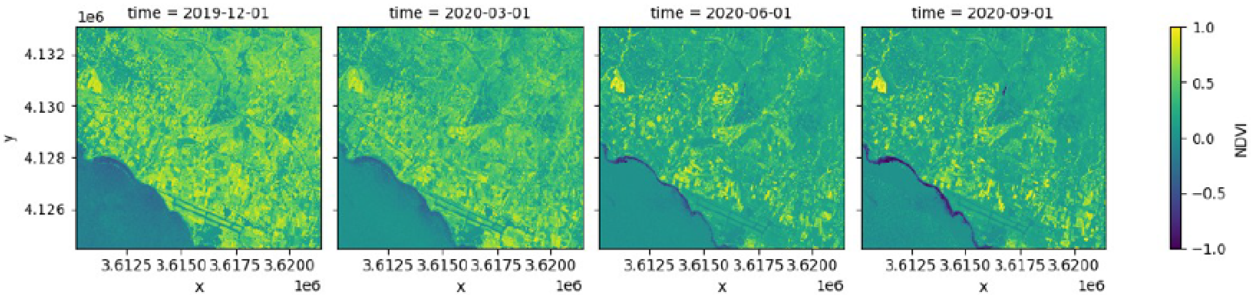

Fig. 3. Vegetation mapping at a main agricultural area of Cyprus. (Acheleia, Paphos) using CEODC.

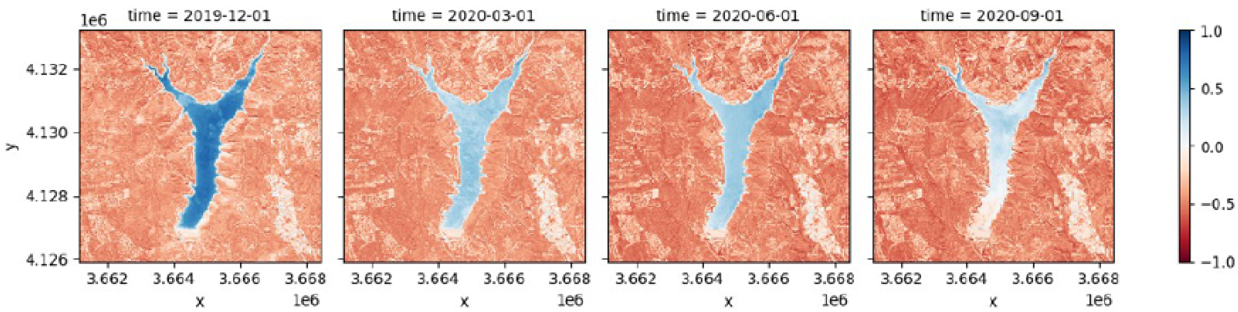

Fig. 4. Mapping the water extent of Kouris' Reservoir (Limassol, Cyprus) using CEODC.

work, and the host for different downstream applications or services. For example, the visualization tool "weblarda" – available through LARDA – can be utilized where users can visualize any selected data.

2.4. Efficiency of Cyprus EO Data Cube

As already mentioned above, two different connections for data acquisition are established. Earth Observation Data Access Gateway offers the capability of downloading Sentinel-1 and Sentinel-2 data with a bandwidth of approximately 11 MB/s. On the other hand, the connection to NASA's EO data regulates bandwidth based on traffic. The processing chain of Sentinel-1 and Sentinel-2 needs approximately 234.1 and 121.3 seconds per image respectively. In contrast, for each MODIS' HDF's pre-processing, about 27.3 seconds are needed.

3. PRELIMINARY RESULTS

To demonstrate the capabilities of CEODC, two case studies are presented. At first, NDVI is employed to explore the variations of vegetation over an important agricultural area of Cyprus, at Acheleia, Paphos (Fig. 3). The second case study, utilizes the Modified Normalized Difference Water Index (MNDWI) to observe differences on the reservoir's water over the largest reservoir of Cyprus, Kouris located in Limassol (Fig. 4).

4. CONCLUSIONS

In this work, a holistic framework is described which will assist in the research exploration and exploitation of EO contributing in environment's sustainability and society's resilience. The framework consists of EODAS, GBS and LLs to acquire data which is handled and stored in the CEODE. The main purpose of this framework is to make Cyprus the reference point in the EMMENA region for space-based environmental monitoring and space technologies. The CEODE is expandable and can handle data from other disciplines in the future, e.g., water resources and maritime LLs. Through, a descriptive review [15] of the literature there is a need of a Central Data Hub in the EMMENA region and this is the gap that our proposed architecture will address. At the moment, we are covering a big part of environment and climate monitoring in EMMENA. With the capabilities of Big Earth Data Analytics and Artificial Intelligence we solve issues for environmental monitoring and society's resilience.

Authorized licensed use limited to: Cyprus University of Technology. Downloaded on May 20,2025 at 08:04:34 UTC from IEEE Xplore. Restrictions apply.

5. ACKNOWLEDGMENTS

The authors acknowledge the ‘EXCELSIOR’: ERATOSTHENES: Excellence Research Centre for Earth Surveillance and Space-Based Monitoring of the Environment H2020 Widespread Teaming project (www.excelsior2020.eu). The ‘EXCELSIOR’ project has received funding from the European Union’s Horizon 2020 research and innovation programme under Grant Agreement No 857510, from the Government of the Republic of Cyprus through the Directorate General for the European Programmes, Coordination and Development and the Cyprus University of Technology.

6. REFERENCES

[1] Huadong Guo, Lizhe Wang, and Dong Liang, “Big Earth Data from space: a new engine for Earth science,” *Science Bulletin*, vol. 61, no. 7, pp. 505–513, Apr. 2016.

[2] Huadong Guo, Stefano Nativi, Dong Liang, Max Craglia, Lizhe Wang, Sven Schade, Christina Corban, Guojin He, Martino Pesaresi, Jianhui Li, Zeeshan Shirazi, Jie Liu, and Alessandro Annoni, “Big Earth Data science: an information framework for a sustainable planet,” *International Journal of Digital Earth*, vol. 13, no. 7, pp. 743–767, July 2020.

[3] Xinwei Wang, Guohua Wu, Lining Xing, and Witold Pedrycz, “Agile earth observation satellite scheduling over 20 years: Formulations, methods, and future directions,” *IEEE Systems Journal*, vol. 15, pp. 3881–3892, 2020.

[4] P Soille, A Burger, D De Marchi, P Kempeneers, D Rodriguez, V Syrris, and V Vasilev, “A versatile data-intensive computing platform for information retrieval from big geospatial data,” *Future Generation Computer Systems*, 2018.

[5] Corina Vaduva, Michele Iapaolo, and Mihai Datcu, *A Scientific Perspective on Big Data in Earth Observation*, pp. 155–188, Springer International Publishing, Cham, 2020.

[6] Ionuț Iosifescu Enescu, Lucia de Espona, Dominik Haas-Artho, Rebecca Kurup Buchholz, David Hanimann, Marius Rüetschi, Dirk Nikolaus Karger, Gian-Kasper Plattner, Martin Hägeli, Christian Ginzler, Niklaus E. Zimmermann, and Loïc Pellissier, “Cloud optimized raster encoding (core): A web-native streamable format for large environmental time series,” *Geomatics*, vol. 1, no. 3, pp. 369–382, 2021.

[7] Badr-Eddine Boudriki Semlali, Chaker El Amrani, Guadalupe Ortiz, Juan Boubeta-Puig, and Alfonso Garcia-de Prado, “SAT-CEP-monitor: An air quality monitoring software architecture combining complex event processing with satellite remote sensing,” *Computers & Electrical Engineering*, vol. 93, pp. 107257, July 2021.

[8] Gregory Giuliani, Bruno Chatenoux, Thomas Piller, Frédéric Moser, and Pierre Lacroix, “Data Cube on Demand (DCoD): Generating an earth observation Data Cube anywhere in the world,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 87, pp. 102035, May 2020.

[9] Martin Sudmanns, Dirk Tiede, Stefan Lang, Helena Bergstedt, Georg Trost, Hannah Augustin, Andrea Baraldi, and Thomas Blaschke, “Big Earth data: disruptive changes in Earth observation data management and analysis?,” *International Journal of Digital Earth*, vol. 13, no. 7, pp. 832–850, July 2020.

[10] Karine R. Ferreira, Gilberto R. Queiroz, Lubia Vinhas, Rennan F. B. Marujo, Rolf E. O. Simoes, Michelle C. A. Picoli, Gilberto Camara, Ricardo Cartaxo, Vitor C. F. Gomes, Lorena A. Santos, Alber H. Sanchez, Jeferson S. Arcanjo, José Guilherme Fronza, Carlos Alberto Noronha, Raphael W. Costa, Matheus C. Zaglia, Fabiana Zioti, Thales S. Korting, Anderson R. Soares, Michel E. D. Chaves, and Leila M. G. Fonseca, “Earth observation data cubes for brazil: Requirements, methodology and products,” *Remote Sensing*, vol. 12, no. 24, 2020.

[11] “The Australian Geoscience Data Cube — Foundations and lessons learned,” vol. 202.

[12] Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore, “Google Earth Engine: Planetary-scale geospatial analysis for everyone,” *Remote Sensing of Environment*, vol. 202, pp. 18–27, Dec. 2017.

[13] Bing Zhang, Zhengchao Chen, Dailiang Peng, Jon Atli Benediktsson, Bo Liu, Lei Zou, Jun Li, and Antonio Plaza, “Remotely sensed big data: evolution in model development for information extraction [point of view],” *Proceedings of the IEEE*, vol. 107, no. 12, pp. 2294–2301, Dec. 2019.

[14] Ziheng Sun, Laura Sandoval, Robert Crystal-Ornelas, S. Mostafa Mousavi, Jinbo Wang, Cindy Lin, Nicoleta Cristea, Daniel Tong, Wendy Hawley Carande, Xiaogang Ma, Yuhan Rao, James A. Bednar, Amanda Tan, Jianwu Wang, Sanjay Purushotham, Thomas E. Gill, Julien Chastang, Daniel Howard, Benjamin Holt, Chandana Gangodagamage, Peisheng Zhao, Pablo Rivas, Zachary Chester, Javier Orduz, and Aji John, “A review of earth artificial intelligence,” *Computers Geosciences*, vol. 159, pp. 105034, 2022.

Authorized licensed use limited to: Cyprus University of Technology. Downloaded on May 20,2025 at 08:04:34 UTC from IEEE Xplore. Restrictions apply.

[15] Marinos Eliades, Silas Michaelides, Evagoras Evagorou, Kyriaki Fotiou, Konstantinos Fragkos, Georgios Leventis, Christos Theocharidis, Constantinos F. Panagiotou, Michalis Mavrovouniotis, Stelios Neophytides, Christiana Papoutsa, Kyriacos Neocleous, Kyriacos Themistocleous, Andreas Anayiotos, George Komodromos, Gunter Schreier, Charalampos Kontoes, and Diofantos Hadjimitsis, "Earth observation in the emmena region: Scoping review of current applications and knowledge gaps," *Remote Sensing*, vol. 15, no. 17, 2023.

[16] J Bühl, M Radenz, W Schimmel, T Vogl, J Röttenbacher, and M Lochmann, "pylarda v3. 2," *Zenodo [code]*, vol. 10, 2021.

[17] Konstantinos Fragkos, Argyro Nisantzi, Ilias Fountoulakis, Silas Michaelides, Georgia Charalampous, Kyriakoula Papachristopoulou, Charis Kontoes, Diofantos Hadjimitsis, and Stelios Kazadzis, "Introducing the solar radiation and energy laboratory of the eratosthenesrsquo; centre of excellence: Overview of activities," *Environmental Sciences Proceedings*, vol. 26, no. 1, 2023.

[18] Rahma Djiroun, Kamel Boukhalfa, and Zaia Alimazighi, "Designing data cubes in olap systems: a decision makers' requirements-based approach," *Cluster Computing*, vol. 22, pp. 783–803, 2019.

Authorized licensed use limited to: Cyprus University of Technology. Downloaded on May 20,2025 at 08:04:34 UTC from IEEE Xplore. Restrictions apply.