EO4EU - AI-AUGMENTED ECOSYSTEM FOR EARTH OBSERVATION DATA ACCESSIBILITY WITH EXTENDED REALITY USER INTERFACES FOR SERVICE AND DATA EXPLOITATION

Babis Andreou, Kakia Panagidi,
Stathes Hadjiefthymiades

Dept. of Informatics
University of Athens
bcand@di.uoa.gr, kakiap@di.uoa.gr,
shadj@di.uoa.gr

Vasileios Baousis

European Centre
for Medium-Range
Weather Forecasts,
Vasileios.Baousis@ecmwf.int

ABSTRACT

A vast amount of Earth Observation (EO) data is produced daily and made available through online services and repositories. Contemporary and historical data can be retrieved and used to power existing applications, foster innovation and finally improve the EU citizens' lives. However, an undersized audience follows this activity, leaving huge volumes of valuable information unexploited. EO4EU provides innovative tools, methodologies and approaches that assist a wide spectrum of users, from domain experts and professionals to simple citizens to benefit from EO data. In this work, EO4EU project is described which delivers dynamic data mapping and labelling based on AI adding FAIRness to the system and data. EO4EU introduces an ecosystem for holistic management of EO data, bridging the gap between domain experts and end users, bringing in the foreground technological advances to address the market straightness towards a wider usage of EO data. EO4EU boosts the Earth Observation data market, providing a digestible data information modeling for a wide range of EO data, through dynamic data annotation and a state-of-the-art serverless processing by leveraging important European Cloud and HPC infrastructures.

***Index Terms*—** Digital Services and Platforms, Artificial intelligence, High-performance computing (HPC), Cloud Infrastructures, Earth Observation / Services and applications

1. INTRODUCTION

The rapid growth of Earth Observation (EO) data, especially in the last decade, combined with the ever-increasing development of analytical theories and tools, has generated a wide range of practical applications covering land, maritime, climate change and atmospheric monitoring. The comprehensive and systematic structuring and processing of this data and their transition to valuable knowledge has helped us model and predict natural processes, as well as understand the complex dynamics of our planet's environment. In the meantime, analytical solutions based mainly on Earth imagery have also affected a wide range of traditional industries, from defence and government to insurance, commercial and health applications, while it has signified the emergence of a dynamic sector of EO SMEs and startups that brings to the foreground a large diversity of processes, services and geospatial products by utilizing satellite data.

Despite the significant volume and plethora of Earth Observation (EO) data offered from current EU services and repositories, their access has not been yet extended beyond experts and scientists to the wider industry to deliver tangible applications that improve our health and lives and protect the planet. Unfortunately, a small part of the market has that kind of expertise and as follows high-value EO information remains unexploited, it is often fragmented, complex, diverse, and difficult to find, retrieve, download and process, while users must have some kind of domain expertise to find, access, understand how to pre-process data, find storage solutions and transform data into useful formats for analytics and Geographic Information Systems (GIS) [5]. Realistic simulations are considered computationally expensive, as the specifications for appropriate infrastructure to process this data are considerably high, ranging from powerful systems, to cloud platforms with additional hardware like GPUs and High Performance Computing Facilities. Moreover, often theoretical models do not translate well into practical applications despite the rapid growth of EO data and the growth of analytical theories and tools.

In this paper, the architecture of EO4EU project is described in order to provide innovative tools, services, methodologies and approaches utilising modern technologies that will assist a wide spectrum of users, from domain experts to simple citizens unaware of the plethora of data and capabilities offered by EU services, to access and process data and utilise the existing and future offered services. EO4EU aims to support the wider exploitation of EO data by delivering: (i)

Machine Learning (ML) methodologies for Semantic Annotation of existing and growing data sources, (ii) semantically enhanced knowledge graphs that will enable structuring of content around diverse topic areas and building step by step journeys from different sources into a unified approach, (iii) data fusion techniques to extend the scalability of existing distributed systems, (iv) Augmented and Virtual Reality for interactive user experience, and (v) advanced data analytics visualizations for improved learning and evidence-based interpretations of environmental observations. To further enhance the proposed approach, the project will utilize existing background technologies and will capitalize on available data sources and data exploitation initiatives, including Copernicus, GEOSS, INSPIRE, DestineE, Galileo, EGNOS, ECMWF, ESA/EPA, Datarade, CDP and Trucost Environmental.

The structure of this work is the following: in Section 2 the main design principles and EO4EU architecture is described followed by Section 3, in which some experiments took place in the first year of implementation and the use of Function-as-a-Service (FaaS). Section 4 is dedicated to conclusions and future actions.

2. EO4EU ARCHITECTURE

The EO4EU platform provides a holistic data processing ecosystem which further improves the access to the EU EO data offered by a variety of platforms and data repositories. Data sources include Copernicus services and associate platforms like DIASes, Gallileo/EGNOS service but also upcoming platforms and initiatives like DestinE. The extensibility and adaptability of the EO4EU platform to retrieve, process, fuse and deliver new datasets, without prior knowledge about their structure and format, is supported by machine learning algorithms and advanced semantic annotation mechanisms.

EO4EU is integrating a range of use cases, covering a wide spectrum of users and stakeholders, thus demonstrating the versatility of the proposed solution. Platforms which are deployed as use cases, are using data provided from EC data repositories and be applied to personalized health services, ship route optimization, food security taking into account climate change, forest management, soil erosion, environmental pest influenced by climatological conditions and civil protection decision making. EO4EU core mechanisms are deployed using microservices orchestration with Kubernetes clusters and higher abstraction services like Function as a Service (FaaS), which allow flexibility, extensibility, and scalability of the platform. Necessary storage for the EO4EU data store will be dynamically provided to the Kubernetes clusters with dynamic allocation of Persistent Volumes. Moreover, a set of visualization services and interfaces were designed, including a multi-layered user interface (GUI) for visual analytics, a Command Line Interface (CLI) and a respective Application Programming Interface (API), and an extended reality (XR) interface to further boost the adoption of the platform. The design of the EO4EU platform, data flows and APIs enable a wide spectrum of business models (processing distribution, integration of data, cloud services) to be easily materialized, thus offering a unique instrument to the modern IT industry. In the following sub-sections, the main pillars of EO4EU ecosystem are presented.

2.1. Data Sources Analysis

Massive batches of EO datasets of historical and daily earth observation measurements along with open-access data from national cohorts and pre-identified databases are linked with EO4EU. Also, real-time data collection streamlines focus on establishing a live connection with devices such as a smart wearable, mobile application, Internet of Things (IoT) sensors and devices. Open-source data collections from numerous sources, private and/or public entities, EU-funded initiatives and projects will also be utilized towards feeding EO4EU with further data sources and unlabeled. EO4EU is further data-enhanced through open access cohorts of the European Commission through the cloud-based platforms established to provide centralised access to Copernicus data and CAMs services, GEOSS, INSPIRE, DestinE, Galileo/ EGNOS programmes as well as direct access to open datasets and services provided by ECMWF and its entrusted services.

2.2. Knowledge Graph-based Decision Making

Knowledge Graph based Decision Making advances decision-making support for the needs of large-scale problems by adopting a knowledge-based decision-making view [1]. According to this view, decisions are considered as pieces of descriptive or procedural knowledge referring to an action commitment. In EO4EU, the synergy of decision-making and knowledge management is further strengthened by the incorporation of features[6], controlling the consequences of information overload, while hiding the complexity of optimization models and their formalisms behind generic and easy-to-use user interfaces.

EO4EU introduces Graph-Based Text Representation. Each document of a corpus is represented as a single graph. Specifically, every unique word of a document is depicted as a graph node and the co-occurrence between two words is denoted with an edge connecting the corresponding nodes. EO4EU' knowledge graph-based approach enables the extraction of informative features for each entity.

2.3. Generic ML pipeline for semantic annotation

EO4EU enables the semantic annotation of EO data, in a way that the requirement of use-case specific labelled data is minimized. Annotated/labeled images are needed to train supervised machine learning algorithms to recognize objects and

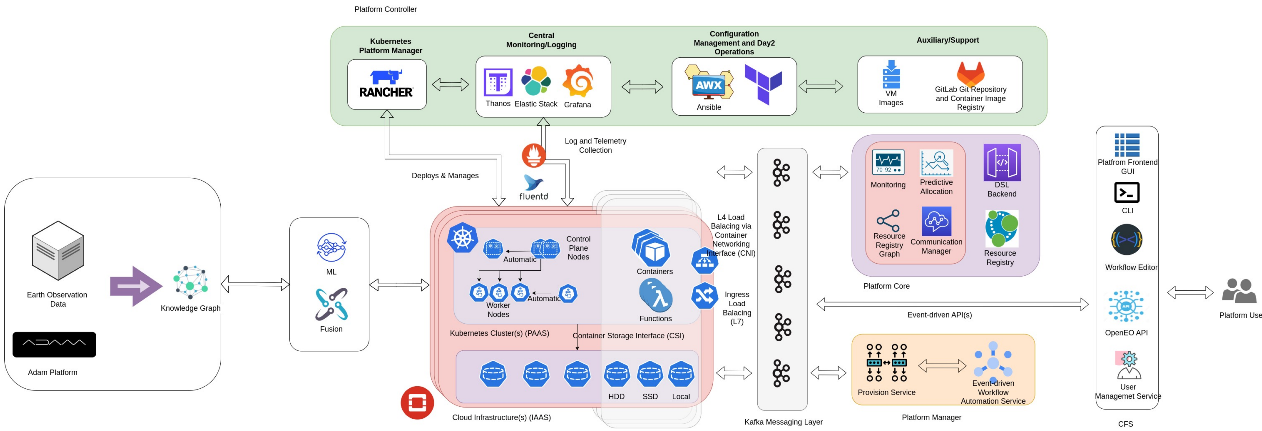

Fig. 1. EO4EU Architecture.

patterns in images. The Generic ML pipeline of EO4EU enables the learning of a robust and transferable representation of the input data in a latent space, in an unsupervised way [3]. ML provides the input representation over which the downstream tasks operates for the learning of task specific models.

As the volume of the data that may be transferred is a major barrier when dealing with EO data, an efficient compression mechanism that is tailored to the particularities of the data can address this issue more efficiently compared to standard of-the-shelf compression methods. Compression models are trained using Variational Autoencoders, i.e Variational Autoencoders with Quantized Latent Spaces (VQ-VAE) [4], as well as other methods that learn (lossy) latent variable representations. Of particular interest are discrete latent spaces which naturally led themselves to the compression task. The respective decompressor will be available to the entities that request the data, which will have the liberty to choose either to rescale the data back in the original size using the decompressor or to use them in their reduced-size form.

2.4. Systems and Services Orchestration

EO4EU' infrastructure provides computational resources and services on various levels for supporting a large variety of the workloads running in the platform. Highly performance x86 servers with modern GPU supporting accelerated AI and High Performance Computing (HPC) resources. Virtualization layer based on Linux/KVM for supporting virtualized workloads. Virtualization aims at minimizing the downtime and overall increase of the compute provisioning speed. The Infrastructure as a Service (IaaS) layer will be realized using OpenStack components for supporting user-driven IT provisioning and automations of the higher level components of Platform as a Service (PaaS). Kubernetes is be the main vehicle for running the operational workloads of the EO4EU platform, and also for supporting subject matter usage involving AI or other types of containerized applications. This way every service/workflow will be deployed isolated from the others, leaving a minimum footprint to the system concerning the workload and the current status of the instance. The Function as a Service (FaaS) orchestrator is able to reserve the needed OpenStack resource according to the workflow requirements. As a second step, a Kubernetes instance is deployed to the stack resources, leading to the deployment of the FaaS framework. The outcome is a FaaS system that is dynamic and versatile, processing a vast load of information with minimum delays.

2.5. Graphical User Interface

An intuitive web-based Graphical User Interface will provide the needed interaction capacity from the most advanced user to a simple viewer, minimizing cognitive load, using consistency as a key, preferring clarity over complexity, while keeping the user in control. A key User Interface (UI) feature is the adaptability provided through a fully configurable dashboard that will be able to personalize the viewable components depending on the user's needs. ASP.NET Core will be used for the development of the dashboard, ensuring the application robustness, high performance, cross-platform support, improved quality and easy maintenance. The dashboard will also rely on open standards to gather data from the infrastructure thus enabling the interfacing between the backend system and the dashboard for easy and secure communication and data transfer.

3. EXPERIMENTS

The main goal of EO4EU is to facilitate the access and the use of EO data to end users facing multiple restrictions, e.g. resources, access to multiple data sources etc. A user can access a GUI to create through the KG a query to the data sources and then to describe a pipeline of actions in order to manipulate a vast amount of information to her own needs.

Table 1. Example of a table format.

MB	1 POD(s)	Opt. PoD(s)	Faas(ns)	cl. Faas(ns)
0.5	92	8.26	45.32	20.1
1	250	18.2	62.89	31.94
10	1050	186	265.65	137.17
20	4413	486	543.16	273.79

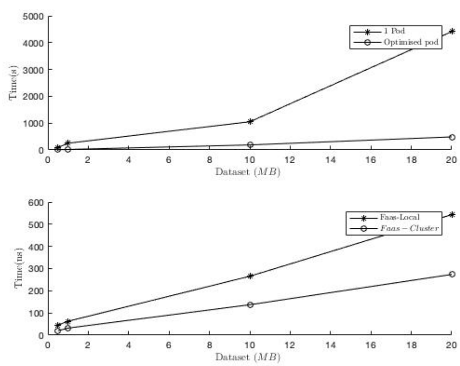

Fig. 2. Experiments of fusion algorithm conducted on 1 PC and a FaaS Cluster.

During this implementation process a benchmarking of the data was conducted to a simple but resource demanding scenario, i.e. the acquiring of a specific dataset including Sentinel S3 tiff images and then make a resampling of data and store the data again as new dataset. This simple function that is a prerequisite of all the machine learning algorithms to manipulate objects of the same size is quite demanding due to the read-write process. It is shown in the Figure below the needs in time for different datasets for 1 PC which has 12 cores and 16 GB RAM and Python 3.10. In the Table 1 it is presented also an optimised approach in which vectorization through numpy was applied in order to excellence the performance of a single resource. An OpenFaaS installation on top of a Kubernetes cluster constitutes the testing environment of the experiment. A proxy component is initiating the FaaS process providing the needed segmentation of the data. This way a batch of EO data is downloaded in the proxy component, divided in small segments (e.g. 1 segment = 1 picture) and after that a requests is lunched towards the OpenFaaS service. Finally, each OpenFaaS instantiated process returns to the proxy component a response with the processed data. The results of conducted experiments can be overviewed in the Table 1 and in the Figure 2. The OpenFaaS functionality [2] is measured in nanoseconds while for the same processing a simple unit needs more than 90 seconds. The experiments were performed in 100 iterations and the average values was computed for different sizes of files, i.e. 0.5Mb, 1Mb, 10Mb and 20Mb. The OpenFaaS functionality outperformed in all cases.

4. CONCLUSIONS

In this paper, the EO4EU platform was described. Its main goal is to provide innovative tools, services, methodologies and approaches utilising modern technologies that assist a wide spectrum of users to access and use data and capabilities offered by EU services. EO4EU will promote pre-operational European services and utilises existing platforms and services in a consolidated manner through the extensive use of disruptive technologies like ML methodologies for Semantic Annotation of existing and growing data sources, semantically enhanced knowledge graphs, data fusion techniques to extend the scalability of existing distributed systems and present all these with XR tools to the end user, for interactive user experience. EO4EU platform will provide preliminary user access to the end-users in Q4/2023 and be fully accessible within 2024. As future work, concerning the FaaS functionality, EO4EU will leverage the capability to work without downloading huge amounts of data at once and distribute that way the overall processing load in memory data caching.

REFERENCES

[1] Jianlin Feng. Knowledge graph embedding by translating on hyperplanes. 06 2014.

[2] Sachchidanand Singh and Mohit Sewak. Winning in the era of serverless computing and function as a service. 2018. doi: 10.1109/I2CT.2018.8529465.

[3] Gencer Sumbul, Marcela Charfuelan, Begum Demir, and Volker Markl. Bigearthnet: A large-scale benchmark archive for remote sensing image understanding. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, jul 2019. doi: 10.1109/igarss.2019.8900532.

[4] Aaron van den Oord, Oriol Vinyals, and koray kavukcuoglu. Neural discrete representation learning. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[5] Julia Wagemann, Stephan Siemen, Bernhard Seeger, and Jörg Bendix. A user perspective on future cloud-based services for big earth data. *International Journal of Digital Earth*, 14(12):1758–1774, 2021. doi: 10.1080/17538947.2021.1982031.

[6] Lingfei Wu, Yu Chen, Heng Ji, and Bang Liu. Deep learning on graphs for natural language processing. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, KDD '21, page 4084–4085, New York, NY, USA, 2021. Association for Computing Machinery. doi: 10.1145/3447548.3470820.