CALLISTO: COPERNICUS ARTIFICIAL INTELLIGENCE SERVICES AND DATA FUSION WITH OTHER DISTRIBUTED DATA SOURCES AND PROCESSING AT THE EDGE TO SUPPORT DIAS AND HPC INFRASTRUCTURES

Stelios Andreadis[1], Ilias Gialampoukidis[1], Vasileios Sitokonstantinou[2], Beatrice Coloru[3], Han Vervaeren[4], Eva López[5], Vasileios Kalogirou[6], Panagiota Syropoulou[7], Elias B. Kosmatopoulos[1] Stefanos Vrochidis[1], Eliana Li Santi[8], Guido Vingione[8], and Ioannis Kompatsiaris[1]

[1]Information Technologies Institute, Centre for Research and Technology Hellas,
[2]National Observatory of Athens, [3]SMAT S.p.A., [4]De Watergroep, [5]Deutsche Welle,
[6]European Union Satellite Centre, [7]DRAXIS Environmental S.A., [8]Serco S.p.A.

ABSTRACT

CALLISTO is a project funded by the European Union's Horizon 2020 Research and Innovation Programme under the topic "Big data technologies and Artificial Intelligence for Copernicus" in 2020 and has started in January 2021, with a duration of three years. In this work we introduce the CALLISTO project and describe its concept and approach, the use cases of application and the expected impact.

Index Terms— Artificial Intelligence, Earth Observation, Copernicus, Data Fusion, Big Data

1. INTRODUCTION

Artificial Intelligence (AI) is already part of our daily lives and is forcefully entering the space sector to offer value-added Earth Observation (EO) products and services. Europe's Copernicus programme provides data on a free, full and open basis, while the recently launched Data and Information Access Service (DIAS) providers [3] index, store and exchange massive amounts of data and cloud infrastructure computational resources. However, there is a gap between these DIAS providers and the application end users, which can be bridged through dedicated AI solutions that will add value to the large volumes of offered satellite data.

Aiming to bridge the above gap, the CALLISTO project [1] provides a highly interoperable Big Data platform between DIAS infrastructures and Copernicus users, where the outcomes of machine learning solutions on satellite data, optimized on High Performance Computing (HPC), are semantically indexed and linked to crowdsourced, geo-referenced and distributed data sources, and served to humans in Mixed Reality environments, allowing virtual presence and situational awareness in any desired area of interest, augmented by Big Data analytics from state-of-the-art and scalable Deep Learning solutions. CALLISTO will be demonstrated in real use case scenarios in relation to policy-making, water management, journalism and security. The concept of CALLISTO along with its application can be seen in Fig. 1.

CALLISTO will be built upon the experience gained from the completed H2020 EOPEN project[8], while relevant projects are DeepCube [2], GEM [4] and RapidAI4EO [6].

This work has been supported by the project CALLISTO, funded by the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 101004152.

2. APPROACH

CALLISTO makes use of existing European data infrastructures, such as Copernicus' ONDA DIAS platform, to develop solutions that can be plugged into DIAS to enhance their capabilities and offer various types of other georeferenced and distributed data sources, Big Data analytics using Deep Learning technologies, semantic technologies, robotic technologies with Unmanned Aerial Vehicles (UAVs) and Mixed Reality (Virtual and Augmented) applications, addressing industrial and societal challenges for policy-makers, water utility operators, international broadcasters and entrusted entities for security applications.

The ONDA data model architecture enables the extraction of value from Copernicus data and information, made available through a dedicated ONDA API [5]. The objective of CALLISTO is to extend the ONDA data model by combining the satellite data with indexing functions for i) GNSS data provided by the Galileo-enabled mobile application, ii) UAV video content processed on-board, and iii) unstructured social media and Web data that are automatically geotagged with semantic and Deep learning technologies (Fig. 2). The new proposed indexing model also aims to address the effectiveness of data search for non-expert users, combining concept-based search, data fusion with 3D models, and semantic search.

The CALLISTO system consists of four layers, which are visualised in Fig. 1 from left to right. The **data collection layer** concerns the ingestion and indexing of additional data from distributed sources on a DIAS infrastructure, the indexing and generation of annotated data, and processing at the

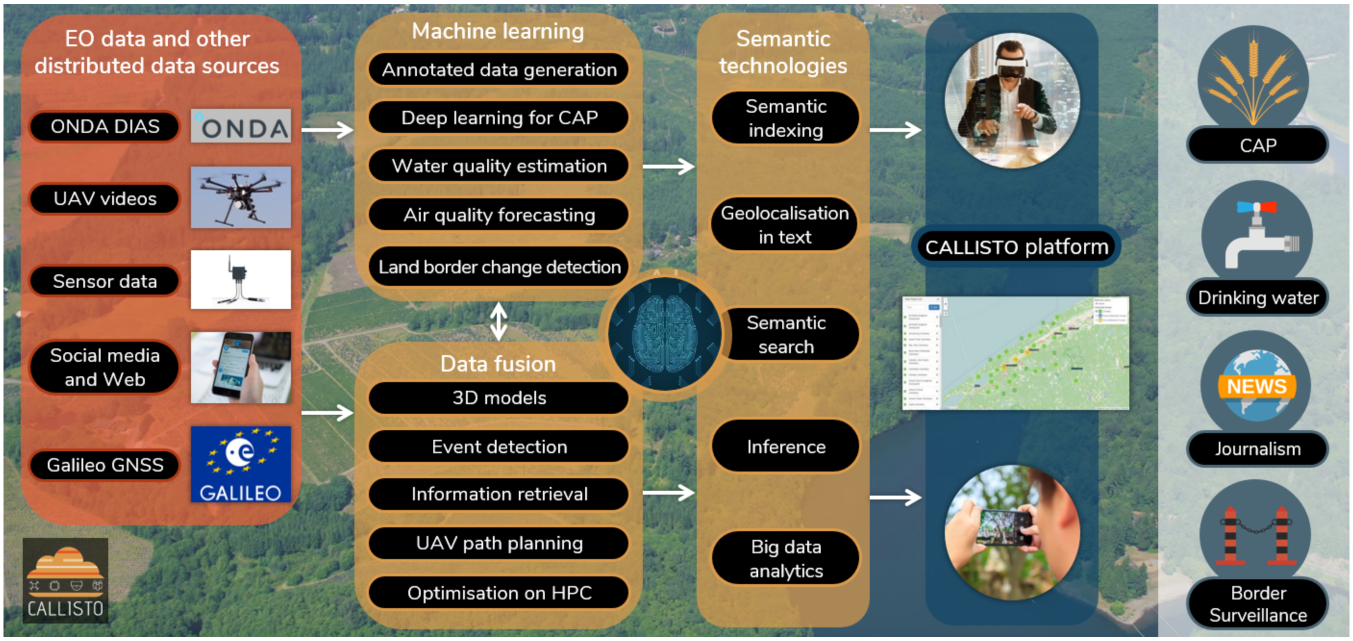

Fig. 1. The CALLISTO concept

edge (on-board of a UAV) and on the cloud (DIAS). The **core intelligence layer** includes the knowledge extraction from Deep Learning models, the fusion of EO and non-EO data, the realisation of UAV missions on demand, and the generation of 3D models from satellite imagery and visual content from UAVs. This layer also involves the automatic generation of annotated data repositories for feeding machine learning algorithms in training and validation phases, and optimisation on HPC infrastructure. The **semantic layer** is responsible for the creation of the CALLISTO ontologies, the automatic extraction of geoinformation when not available, and the retrieval of information via complex queries. Finally, the **user interaction layer** refers to the implementation of the platform, the user interfaces, the mobile application and the Mixed Reality environment to support Augmented and Virtual reality in the considered use case scenarios.

3. TECHNOLOGIES

In CALLISTO in-situ sensors provide hyperspectral measurements, a dedicated mobile application collects Galileo/GNSS data, Web scrapers and social media crawlers collect crowd-sourced information, and UAVs capture video recordings. On top of these data sources, complementary to the Copernicus data already offered and indexed on ONDA DIAS, AI technologies are provided. First, clustering techniques and Generative Adversarial Networks are proposed to generate annotated data to enrich the available data collections and Deep Learning is applied to extract meaningful knowledge, such as concepts, changes, and events based on data fusion. Detected events indicate the geographical coordinates where a UAV needs to fly, in an alert-driven way with minimum human intervention. Information retrieval from geo-referenced video content extracts additional in-situ knowledge, which can be used to validate and further enrich the outcomes of the Deep Learning analysis of Copernicus imagery. Machine Learning techniques are also applied for water quality estimation and air quality forecasting, while algorithms are optimised to be executed on HPC infrastructures, when necessary to boost scalability. Complex queries and association to Linked Open Data are supported through semantic representation (CALLISTO ontologies) and reasoning. Finally, 3D models are reconstructed from satellite data and allow the automatic creation of Virtual, Augmented and Mixed reality environments for virtual presence in any area of interest.

4. APPLICATION

Four use cases are foreseen in the CALLISTO project. The following subsections present the challenges that CALLISTO will have to address for each one of the use cases.

4.1. Edge Computing and Virtual Presence for the monitoring of CAP

The well-discussed reform and simplification of EU's Common Agricultural Policy (CAP) direct payment management and control measures is an ongoing process which aims at substituting the On The Spot Checks (OTSC) by an automated monitoring system. The current OTSCs are based on a random, limited sample (circa 5%) of farmers' declarations.

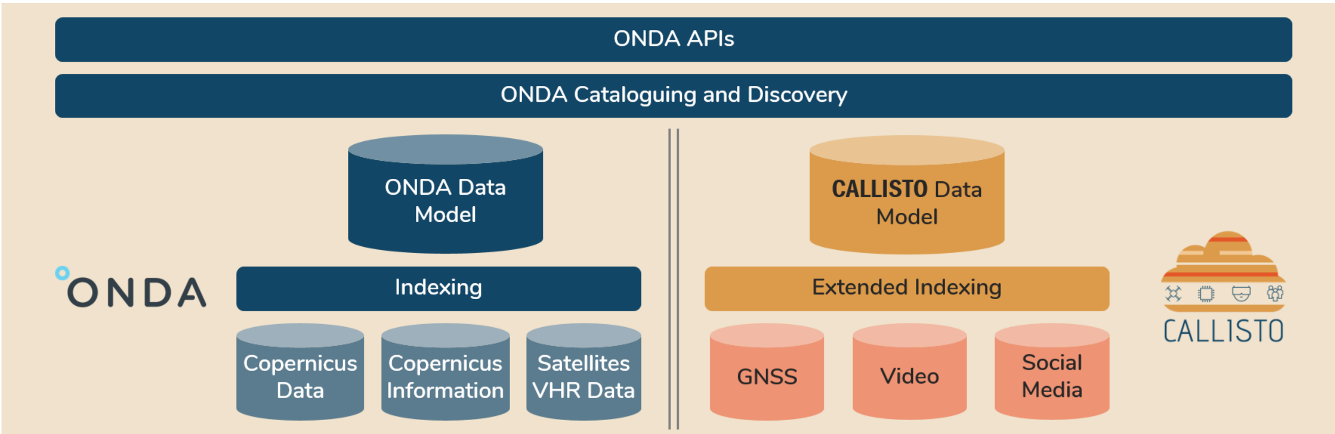

Fig. 2. The CALLISTO extension of the DIAS concept

These checks are based on field inspections or photointerpretation using high resolution images. It is apparent that the implementation of an automated monitoring system enhances not only the regularity and scalability of checks, but also the decision making of appropriate follow-up actions. Taking into consideration these, an implementing regulation (Regulation (EU) No. 2018/746) was adopted in May 2018, giving Member States the potential to simplify their controls by using a monitoring system instead of OTSC. However, no official monitoring system has been established since then. The idea behind such system is the exploitation of a time series of satellite imagery so for phenological cycle to be fully captured. Data provided by declaration are used for monitoring several activities, e.g. crop type identification for crop diversification and grassland mowing events.

4.2. Virtual presence in water resources for water quality assessment using EO and in-situ data

This use case comprises two scenarios. The first scenario concerns the De Blankaart and Kluizen Water Production Centres in Belgium, which contain several million cubic meters of raw water at the intake side of the station in large open basins or ponds, acting as drinking water buffer. Due to the increasing occurrence of extreme summers as well as intensive industrial or agricultural activities in the surroundings, the risk for eutrophication events with algae blooms in the storage basins is higher and the intake water quality needs close monitoring and management. To maintain and improve the water quality and to monitor changes in chemical, biological and physical parameters, accurate high spatial and temporal resolution water quality measurements are required. Thus, remote sensing techniques involving in-situ sensors and satellites can aid.

The second scenario concerns a Water treatment plant in the metropolitan City of Turin, Italy, managed by SMAT. SMAT produces every year more than 250 million cubic meters of quality drinking water, 16.3% of which is taken from the Po river. However, the river is exposed to a number of contamination threats from industrial plants and stores containing hazardous materials and it runs in proximity to highways and high-speed roads used by trucks transporting a number of dangerous substances.

4.3. Satellite Journalism

It is at the core of journalistic work to observe and report on political and social events and developments. One of the most important questions today is how mankind will deal with the global challenge of climate change. In December 2019, the President of the European Commission, Ursula von der Leyen, has announced the European Green Deal: a roadmap towards a climate-neutral continent and sustainable economy until 2050. To this end, several sectors need to fundamentally transform, many different actors to be aligned, advancements and drawbacks to be monitored and policy changes to be initiated. It is crucial for the success of Green Deal that all related political initiatives and developments are closely monitored and analysed by independent journalists too. Climate politics will affect us all and their acceptance will depend on how transparently and plausibly they will be implemented. However, investigative journalism in the field of climate politics requires very particular knowledge, skills and, most importantly, access to reliable large-scale data and information. Only if this information is freely available and accessible for control, analysis and coverage by independent journalists, the expected massive changes in our day-to-day lives will have a chance to be met with understanding and support.

4.4. Land border change detection

Surveillance of EU's external borders is vital for the internal security and protection of its citizens. For this reason, Border Surveillance constitutes one of the three key areas in the Copernicus Security Services. The service provides situational awareness capabilities on the EU's external land and sea borders, supporting the EU's external border surveillance information exchange framework (EUROSUR), managed by the European Border and Coastguards Agency (FRONTEX). Changes in land borders impose corresponding changes in their permeability. The constructions of preventive operational measures (e.g. fences), new buildings near a land border, development of communication network (e.g. trails, roads) inside a forested area, or changes in the infrastructure of Border Control Points (BCP), signal consequent changes in the permeability of the borders [7]. In a comprehensive situational awareness context, these "events" must be detected and analysed. The only feasible approach to monitor and detect such changes at scale, to inform decision makers and EU institutions, is through Earth Observation. The availability of Copernicus data streams, via DIAS infrastructures, makes this development possible. However, monitoring EU external borders with Big EO data, at scale, is a challenging scenario.

5. IMPACT

By fulfilling its objectives, CALLISTO is expected to have an eminent societal, technical and scientific impact, which is briefly described in the following subsections.

5.1. Societal Impact

CALLISTO provides annotated datasets, information and infrastructure as a collection of AI services to support the scientific community with new models and datasets, citizens with timely and reliable information, SMEs that develop new products and services, and industries in making use of state-of-the-art services in their operations.

In policy-making, CALLISTO involves ancillary crowdsourced information and inference at the edge using UAVs, which could lead to an all-inclusive solution for monitoring the compliance to CAP, giving an accurate and cost-efficient decision-making alternative to paying agencies for "smart sampled" parcels. In water sector, the CALLISTO system can improve the ability of water industries and water authorities to better observe and control their regions, without the need for physical presence and inspection. In journalism, CALLISTO supports reporters in researching and investigating areas that are usually hard to access as well as environmental issues through non-traditional data, such as EO and sensor data. Lastly, in security, the CALLISTO system informs timely on land changes at the European borders, contributing to increased situational awareness at strategic, operational and tactical levels.

5.2. Technical and Scientific Impact

Incorporating an effective and efficient orchestration of DIAS resources, edge processing on-board of a UAV with GPU, and HPC infrastructure, CALLISTO presents advanced data processing capacity with innovative types of computing. With the multimodal indexing, analysis and presentation of data deriving from satellites, UAVs, sensors and online users, CALLISTO constitutes an integrated solution for handling heterogeneous data. Furthermore, the collection of existing data sets and the generation of new annotated data, relevant to the CALLISTO problems and use cases, offer novel and open data repositories for training Deep Learning models and Machine learning in general. Finally, the proposed water quality estimation approach, using EO data on DIAS platform and enhanced by videos from UAV flights on-demand, provides a continuous and timely detection of all potential hazards in water safety and security.

6. CONCLUSION

In this paper we present in brief the CALLISTO concept and approach, the technologies it comprises, its application in four different use cases and its expected impact. The purpose of this work is to introduce the CALLISTO project to the scientific community and establish the basis for implementation, application and validation, evaluation, and demonstration in the following three years.

REFERENCES

[1] CALLISTO project. https://cordis.europa.eu/project/id/101004152.

[2] DeepCube project. https://cordis.europa.eu/project/id/101004188.

[3] Data and Information Access Services. https://www.copernicus.eu/en/access-data/dias.

[4] GEM project. https://cordis.europa.eu/project/id/101004112.

[5] ONDA API. https://www.onda-dias.eu/cms/services/catalogues/advanced-api/.

[6] RapidAI4EO project. https://cordis.europa.eu/project/id/101004356.

[7] V. Kalogirou. Imagery Intelligence-derived analytics for security applications: Challenges and opportunities. In *ESA Living Planet Symposium 2019*, Milan, Italy, 2019.

[8] G. Vingione, G. Scarpino, L. Marzell, T. Pettengell, I. Gialampoukidis, S. Andreadis, S. Vrochidis, I. Kompatsiaris, B. Valentin, L. Gale, W.-K. Lee, W. Lee, M. Gienger, D. Hoppe, V. Sitokonstantinou, I. Papoutsis, and C. Kontoes et al. EOPEN: Open interoperable platform for unified access and analysis of earth observation data. In *Proc. of the 2019 conference on Big Data from Space (BiDS'19)*, page 1–4, Munich, Germany, 2019.