Linked Data for Common Agriculture Policy: Enabling Semantic Querying over Sentinel-2 and LiDAR Data

Dina Sukhobok[1], Héctor Sanchez[2], Jesús Estrada[2] and Dumitru Roman[1]

[1] SINTEF, Forskningsveien 1a, 0373 Oslo, Norway
{dina.sukhobok,dumitru.roman}@sintef.no
[2] TRAGSA, Calle Conde de Peñalver, 84 - 2 planta, 28006 Madrid, Spain
{jmev,hsanche1}@tragsa.es

Abstract. The amount of open and free satellite earth observation data combined with available data from other sectors (e.g. biodiversity, landscape elements, cadaster data) has the potential to enhance decision-making processes in various domains. An example of such a domain is agriculture, where the ability to objectively and automatically identify different types of agricultural features (e.g., irrigation patterns and landscape elements) can lead to more effective agriculture management. In this paper we show the possibility to publish and integrate multi-sectoral data from several sources into an existing data-intensive service targeting better and fairer Common Agriculture Policy (CAP) funds assignments to farmers and land owners. We show an end-to-end approach for integrating multi-sectoral data and publishing the result as Linked Data with the help of the DataGraft platform. To demonstrate the use of the resulted dataset, we developed a visualization system prototype showing various information about agricultural parcel features.

Keywords: Linked Data, Common Agricultural Policy (CAP), data integration

1 Introduction

One important decision-making problem in Europe is the Common Agricultural Policy (CAP) funds assignment to farmers and land owners. For the last 50 years, CAP has been the European Union's most important common policy, which has historically required a significant part of the budget. The main objective of the policy is to support the provisioning of a stable supply of safe food produced in a sustainable way at affordable prices for consumers. At the same time CAP aims to ensure a decent standard of living for farmers and agricultural workers, sets requirements for environmental protection and food safety[3].

The Tragsa Group, which is the part of the group of companies administered by the Spanish state-owned holding company Sociedad Estatal de Participaciones Industriales (SEPI), has developed an extension of the existing service of

[3] https://ec.europa.eu/agriculture/cap-overview_en

Spanish CAP funds assignments – the Common Agriculture Policy Assignment Service (CAPAS)[1,2]. Currently, the assignment of funds is based on several parameters provided by human operators, whose work is mainly concerned with ortophotos, satellite and aerial images, and interested parties such as the land owners. The CAPAS service aims to extend and improve the existing service with the purpose of achieving a better and more objective assignment of funds as well as to increase the transparency of the data and enabling its open use through the reliance on Linked Data technologies. The main advantage of using Linked Data technologies is the possibility to make data explicitly and publicly available in a machine-readable and human-understandable format. The improvement is based on the integration of new datasets, currently underused and the replacement of human-generated data (subjective) for more objective data collected and integrated from different cross-sectorial sources in an automated way. The main users of the developed service are the Spanish Environment and Agriculture Ministry, autonomous communities, Treasury Ministry, and Cadastre agencies. Secondary users are banks, other public administrations and farmers.

In this paper, we demonstrate the applicability of Linked Data technologies to integrate relevant multi-sectoral data from several sources into an existing CAP service. The resulting dataset can enable fairer fund specification, better environmental management and cadastral assessment. In addition, we present a CAPAS GIS viewer as a general showcase for various information about agricultural parcel features.

2 Linked Open Data from Sentinel-2 and LiDAR data

Data Sources. Various datasets were used in CAPAS service to produce an enhanced assessment of suitability for agricultural grant requests. The aim of CAPAS is to improve the efficiency of the existing Spanish CAP database, which is the foundation of the funds assignment service. This database is also known as LPIS or Land Parcel Identification System. The additional datasets come from two main external sources. One source is Sentinel-2[4] satellite data,[5] obtaining the following layers:

- Simple products generated by a single image of one specific date, including true color images, false color images and NDVI images, useful to identify the type of crop in a parcel.
- Complex products produced annually through several images taken at different dates.

Another source includes data from LiDAR[6] flights within the Spanish National Plan of Aerial Ortho-photography, obtaining the following layers:

[4] `https://sentinel.esa.int/web/sentinel/missions/sentinel-2`

[5] Sentinel-2 data are openly accessible at Sentinels Scientific Data Hub (`https://scihub.copernicus.eu/`) as raster images (JPEG2000).

[6] `https://en.wikipedia.org/wiki/Lidar`

– Landscape elements layer, defined as areas of natural vegetation within the agricultural parcels, including isolated trees, copses and hedgerows.
– Ecological value layer, defined as the total value of a specific parcel, calculating its protected surface under Natura 2000 protected sites specifications and the surface of the inventoried landscape elements.

For LiDAR data, we make use of the LiDAR files provided by the Spanish National Geographic Institute (IGN). After collecting data from external sources, we applied the developed data processing algorithms (e.g., algorithm for identification of crops, algorithm for detection of landscape elements[7]) in order to simplify the huge amount of data. The processed data have been made available as ESRI Shape Files[8].

Ontology description. Within scope of the CAPAS service we developed the proDataMarket[9][5] ontology for representing data relevant for service. The ontology has been divided into several sub-ontologies, where each sub-ontology contains concepts and properties specific for the particular domain. This modular approach also helped to handle the complexity of the model and made it easier to maintain. The sub-ontologies used for publishing the CAPAS service related data include Land Parcel Identification System (LPIS), Protected Sites, Sentinel data, Landscape Elements, and Assessment.

Linked data generation and publication. The publication of CAPAS service data as Linked Data was performed with the help of DataGraft [3,4][10]– a cloud-based platform for data cleaning and Linked Data generation. DataGraft provides facilities such as raw data cleaning and preparation (most often from tabular formats), mapping to standard Linked Data ontologies, and generating a semantic RDF graph. Data cleaning and preparation activities for the CAPAS service data included: generating and assigning unique identifiers to identify unequivocally entities; data type casting; geospatial data conversion from the Universal Transverse Mercator (UTM) system to the World Geodetic System 1984 (WGS84).

After the input data files were cleaned, they were mapped to the concepts defined by the proDataMarket ontology and made available through a SPARQL endpoint[11]. At this stage, data from two pilot areas was used in the service. The DataGraft capabilities of reusing and extending existing data transformations allowed inclusion of new data to the endpoint in a convenient way.

3 Demonstration Outline

During the demonstration, we will introduce both the enabler for publishing CAPAS service data as a Linked Data – the DataGraft platform, and a visualization

[7] `https://blog.prodatamarket.eu/wp-content/uploads/2017/04/paper_submitted2.pdf`
[8] `https://www.esri.com/library/whitepapers/pdfs/shapefile.pdf`
[9] `http://vocabs.datagraft.net/proDataMarket/`
[10] `https://datagraft.io/`
[11] `https://rdf.datagraft.net/4037585987/db/repositories/capas-2`

system prototype – the CAPAS GIS viewer[12]. The scenario demonstrated will cover uploading raw CAPAS data in the DataGraft platform, using a prepared transformation for data cleaning and Linked Data generation and the demonstration of the visualization system. The CAPAS GIS viewer allows users to configure visualizations and browse data for two pilot areas from existing Spanish CAP database as well as processed data obtained from Sentinel-2 and LiDAR (see Figure 1).

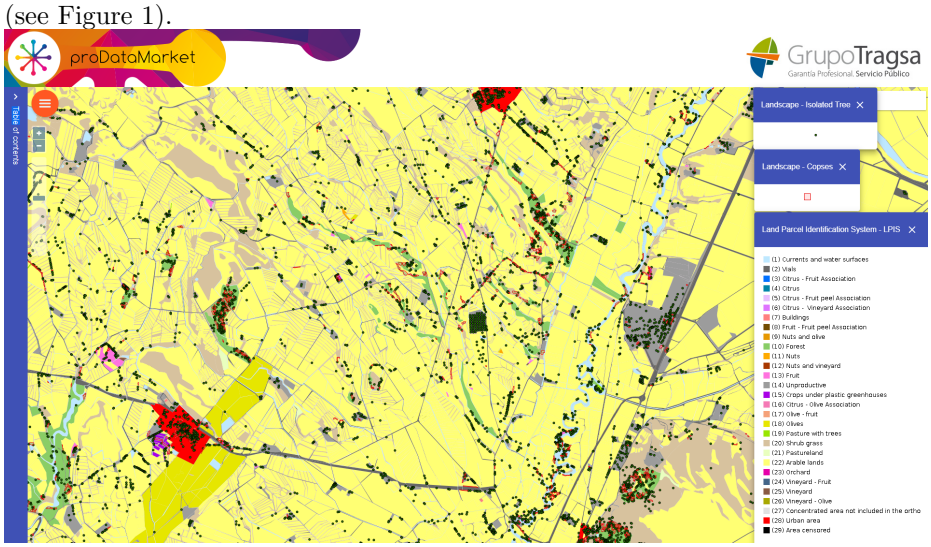

Fig. 1. The CAPAS GIS viewer

Acknowledgements This work is partly funded by the EC H2020 project proDataMarket (Grant number: 644497).

References

1. Navarro, Mariano, et al. "CAPAS: A Service for Improving the Assignments of Common Agriculture Policy Funds to Farmers and Land Owners." Challenge+ DC@ RuleML. 2015.
2. Estrada, Jesús, et al. Enabling the Use of Sentinel-2 and LiDAR Data for Common Agriculture Policy Funds Assignment. ISPRS Int. J. Geo-Inf. 2017, 6, 255.
3. Roman, Dumitru, et al. "Datagraft: Simplifying open data publishing." ESWC (Satellite Events) 2016: 101-106.
4. Roman, Dumitru, et al. "DataGraft: One-stop-shop for open data management." To appear in the Semantic Web Journal (SWJ) Interoperability, Usability, Applicability (published and printed by IOS Press, ISSN: 1570-0844), 2017, DOI: 10.3233/SW-170263.
5. Shi, Ling, et al. The proDataMarket Ontology for Publishing and Integrating Cross-domain Real Property Data. To appear in the journal "Territorio Italia. Land Administration, Cadastre and Real Estate". n.2/2017.

[12] `http://capas.prodatamarket.eu`