ADVANCED VISUALISATION OF BIG DATA FOR AGRICULTURE AS PART OF DATABIO DEVELOPMENT

Karel Charvat[1], Tomas Reznik[2], Vojtech Lukas[3], Karel Charvat Junior[1], Karel Jedlicka[4], Raul Palma[5], Raitis Berzins[6]

[1]Lesprojekt sluzby, Martinov, Zaryby, The Czech Republic
[2]Masaryk University, Faculty of Science, Department of Geography, Brno, The Czech Republic
[3]Mendel University, Faculty of Agronomy, Department of Agrosystems and Bioclimatology, Brno, Czech Republic
[4]University of West Bohemia, Faculty of Applied Sciences, Department of Geomatics, Pilsen, The Czech Republic
[5]Poznan Supercomputing Centre, Poznan, Poland
[6]Baltic Open Solutions Center, Riga, Latvia

ABSTRACT

There is an increasing tension in agriculture between the requirements to assure full safety on the one hand and keep costs under control on the other hand, both with respect to (inter)national strategies. Farmers need to measure and understand the impact of huge amount and variety of data which drive overall quality and yield in their fields. Among others, those are local weather data, Global Navigation System of Systems data, orthophotos and satellite imagery, data on soil specifics etc. A strong need to secure Big Data arises due to various repositories and heterogeneous sources. Data storage and visualisation requirements are in some cases competing as they are a common interest as well as a threat that helps one part of a value chain to gain a higher profit. As demonstrated in this paper, handling (Big) data is therefore a sensitive topic, where trust of producers on data security is essential.

Index Terms — precision agriculture, big data, yield productivity zones, visualisation

1. INTRODUCTION

The agriculture sector is of strategic importance for (European) society and economy. Due to its complexity, agri-food operators have to manage many different and heterogeneous sources of information. (Precision) Agriculture requires collection, storage, sharing, and analysis of large quantities of spatially and non-spatially referenced data. These data flows currently present a hurdle to uptake of precision agriculture as the multitude of data models, formats, interfaces and reference systems in use result in incompatibilities. In order to plan and make economically and environmentally sound decisions a combination and management of information is needed [1].
DataBio (Data-Driven Bioeconomy) project [2] aims at demonstrating the benefits of Big Data technologies in raw material production in agriculture, forestry and fishery/aquaculture for the bioeconomy industry to produce food, energy and biomaterials responsibly and sustainably. DataBio project re-uses and further develops the FOODIE project results [3]. DataBio deploys a state of the art big data platform "on top of the existing partners" infrastructure and solutions - the Big DataBio Platform. The Big DataBio Platform also comprises experts from bioeconomy and technology research institutes, end users, technology providers, and other partners. Furthermore, associated partners and other stakeholders are also actively involved in the pilots to verify the Big DataBio platform capabilities. The Big data technology (BDT) is a new technological paradigm that is driving the entire economy, including low-tech industries such as agriculture where it is implemented under the banner of precision farming (PF). Following the BDT and PF schemes, a Big data analytics system then provides farm managers with highly localized descriptive, prescriptive, and predictive plans. Descriptive plans offer better and more advanced way of looking at an operation, while prescriptive plans provide timely recommendations for operation improvement, i.e. seed, fertilizer and other agricultural inputs application rates, soil analysis, and localized weather and disease/pest reports, based on real-time and historical data. Finally, predictive plans use current and historical data sets to forecast future localized events and returns. The following sections focus on three new

domains of Big Data Visualisation and Analysis for Agriculture:

- yield predictions (see section 2);
- analysis and visualisation of Agriculture Linked Open Data (RDF, see section 3);
- 3D visualisation and analysis of agricultural data (see section 4).

2. YIELD PRODUCTIVITY ZONES

Estimating potential crop yield is a crucial activity performed in the assessment of seasonal production. In contrast, universal indirect methods are also being elaborated for the assessment of actual crop growth and yield based on remote sensing (see e.g. [4], [5] [6] or [7]). Their main motivation is to establish a model based on vegetative indices that is capable of identifying highly productive and less productive zones within a plot, which are linked to the actual or potential crop yield.

The concept of yield productivity zones was introduced in the FOODIE project and further developed in the DataBio project. It aims at the discovery, verification, and user-friendly visualisation of long-term high and low yield productivity zones, as they are areas where crop yield has been significantly above or below the average yield for the whole plot for several years, and because they require the site specific tailoring of fertilizer application rates to ensure the most efficient use of nutrients for the determined level of yield production. In addition, the identification of such zones can help to minimize environmental pollution by residues of agrochemical substances (fertilizers, pesticides). Addressing within-field spatial variability by delineating yield productivity zones is crucial for site specific crop management in precision agriculture [8] and follows the concept of sustainable food production ([9] as well as [10]). The term "long-term" in this context means for as long as agronomical practices do not change significantly, i.e. they remain comparable. The motivations to discover high/low crop yield productivity zones are therefore economic and environmental – to increase profitability by saving on fertilizer, fuel, and manpower on the one hand, and to avoid the overuse of agrochemicals in order to reduce soil and/or water pollution on the otherhand.

The calculation of yield productivity is based on the relationship between vegetation indices from satellite data and crop yield recorded during harvest. The ESPA repository of LANDSAT satellite images and Copernicus Open Access Hub were used as the main data sources offering surface reflectance products, main vegetation indices (NDVI, EVI) and clouds identification by the CFmask algorithm. A selection of scenes from recent eight years was made for specific farm area to collect cloud-free data related to the second half of vegetation period. Yield productivity zones were calculated as the relation of a pixel in individual scenes to the mean value of the plots and, in the next step, as the aggregation of all scenes into one layer. Fig. 1 depicts yield productivity zones prediction together with yield maps of spring barley for year 2017 for the Rostěnice Farm in the Czech Republic.

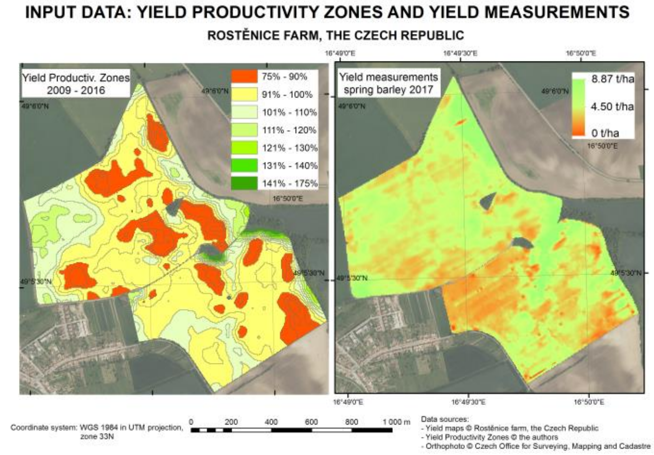

Fig. 1. Yield productivity zone prediction derived from satellite images (left) and the map of yield measurements as computed by a harvester (right).

3. OPEN DATA MODEL

Linked data are increasingly becoming one of the most popular methods for publishing data on the Web. There are various reasons for that: on the one hand, linked data define simple principles for publishing and interlinking structured data that is accessible by both humans and machines. Improving the data accessibility lowers the barriers on finding and reusing such data, while providing machine-readable data facilitating the integration of such data into different applications. On the other hand, linked data allow to discover more useful data through the connections with other datasets, and to exploit it in a more useful way through semantic queries. As a result, there is a growing number of datasets becoming available in linked data format, as can be seen in the Linking Open Data cloud diagram [11], which includes in its latest version (26 January 2017) more than 1100 datasets. Although large cross-domains datasets exist (like dbpedia or freebase) and also some domains are already well covered, like Geography, Government, and Bioinformatics, this is still not the case for agricultural related data. Relevant thesauri may be discovered, like AGROVOC from FAO (Food and Agriculture Organisation) or the National Agricultural Library's Agricultural Thesaurus (NALT). Nevertheless, there is still a lack of datasets related to the agricultural facilities and farm management activities. This is in part due to the lack of standardized models for the representation of such data.

The FOODIE project addressed such issue with the development of the FOODIE data model [12]. To ensure the

maximum degree of data interoperability, the model follows and extends the INSPIRE (INfrastructure for SPatial InfoRmation in Europe) generic data models, especially the data model for Agricultural and Aquaculture Facilities (AF). For instance, a feature was added into the generic INSPIRE data models on a more detailed level than Site. The key motivation was to represent a continuous area of agricultural land with one type of crop species, cultivated by one user in one farming mode (conventional vs. transitional vs. organic farming). Such concept is called Plot and represents the main element in the model since it is the level to which the majority of agro data is related. The Management Zone feature is even one more level lower than the Plot feature. The Management Zone feature enables a more precise description of the land characteristics in fine-grained areas.
Additionally, the FOODIE data model [12] includes concepts for crop and soil data, treatments, interventions, agriculture machinery, etc. Moreover, the model re-uses data types defined in ISO standards (among others, ISO 19101, ISO/TS 19103, ISO 8601 and ISO 19115) as well as standardization efforts published under the INSPIRE Directive (like structure of unique identifiers). The FOODIE data model was specified in UML (as the INSPIRE models) as well as transformed into an OWL (Ontology Web Language) ontology in order to enable the publication of linked data compliant with FOODIE data model [13].

4. FARM MACHINERY AND 3D VISUALIZATION

A lot of visualisation techniques may be found in the domain of Precision Agriculture [14]. This paper presents two specific techniques, a farm machinery visualisation (re-using the data modelling principles described in section 3) and a 3D interactive visualization of yield productivity zones (as described in section 2 with data model specifics; depicted in Figure 3). A newly developed tool called HSLayers NG (see also https://github.com/hslayers/hslayers-ng) has been common for both visualisation techniques. Such tool supports 3D visualization together with methods to query and visualize RDF (Rich Description Framework) data.

Monitoring of machinery fleet movement and especially its spatiotemporal changes brings new insights about the consequences of human decisions from many areas. Economic reasons are related to economic evidence for a farmer, including fuel consumption, efficiency of trajectory etc. to revenue authority or subsidies management. On the other hand, ecologic motivations aim to decrease of environmental burden caused e.g. by high CO_2 emissions due to a lack of movement optimisation, water pollution by nitrogen due to excessive fertilisation. The main focus of this visualization technique is therefore given to the interactivity through utilizing the concept of Multiple Coordinated Views and dynamic queries to emphasize the impact of changes of various phenomena on the traffic volume (see also Fig. 2). More details on cartographic visualization may be found in [16].

Fig. 2. Interactive cartographic visualization of agricultural machinery monitoring. It allows users to filter according to days, hours of a day, speed of a tractor, fuel consumption etc.

The second visualisation technique allows users to explore yield productivity zones with respect to topography (represented as DTM – Digital Terrain Model, DSM – Digital Surface Model, slope, slope orientation, and topography wetness index) according to [17]. A so-called perspective view was implemented in the application. The perspective view contains a three dimensional model of the farm plots and shows the area of interest in a perspective projection taking into account the observer position and his or her line of sight (see Fig. 3 or directly the application under the URL http://ng.hslayers.org/examples/rostenice/). Moreover, a user can add a Web map service of his choice (when respecting WGS84 coordinate system) to make custom 3D mash-ups. A farmer may check specific locations like areas with steep slopes, and see how the machinery deals with them (e.g. checking the machinery tracklogs from the first visualisation technique).

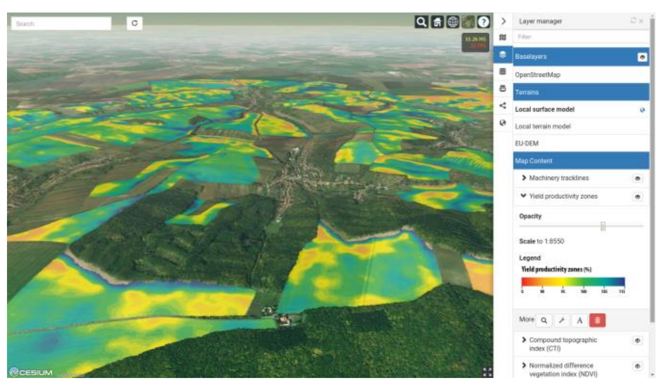

Fig. 3. Perspective view. Perspective visualization of the yield productivity zones of Rostěnice Farm (Czech Republic) portrayed on top of the local surface model.

5. CONCLUSIONS

The presented data models, their semantic equivalents, concept of yield productivity zones and visualization techniques represent the cornerstones for the Big data technology applications in the domain of Precision Farming as understood by the consortium of the DataBio project. The ongoing work focuses on several associated aspects, starting from validations of yield productivity zones prediction, through optimising the performance of developed 3D visualisation application and support of its further visualisation capabilities to the development of semantic end user applications.

Acknowledgments: This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 732064 entitled "Data-Driven Bioeconomy" (DataBio), the European Union's Seventh Framework Programme for research, technological development, and demonstration under grant agreement No. 621074 titled "Farm-Oriented Open Data in Europe" (FOODIE).

6. REFERENCES

[1] T. Reznik, V. Lukas, K. Charvat, K. Charvat jr., Z. Krivanek, M. Kepka, L. Herman, H. Reznikova, "Disaster Risk Reduction in Agriculture through Geospatial (Big) Data Processing," *ISPRS International Journal of Geo-Information* 2017, 8 (6), Basel, Switzerland, 11 p., DOI: 10.3390/ijgi6080238.

[2] DataBio [Data-Driven Bioeconomy official project website], created 2017, accessed 30 April 2018, available at: https://www.databio.eu/.

[3] FOODIE [Farm Oriented Open Data in Europe official project website], created 2014, accessed 30 April 2018, available at: https://www.foodie-project.eu/.

[4] M.E. Bauer, "The role of remote sensing in determining the distribution and yield of crops," *Advances in agronomy*, Brady, N.C., Ed. Academic Press: 1975; 27, pp 271-304.

[5] P.C. Doraiswamy, J.L. Hatfield, T.J. Jackson, B. Akhmedov, J. Prueger, A. Stern, "Crop condition and yield simulations using landsat and modis," *Remote Sensing of Environment* 2004, 92, pp. 548-559.

[6] D.B. Lobell, "The use of satellite data for crop yield gap analysis," *Field Crops Research* 2013, 143, pp. 56-64.

[7] R. Stampach, P. Kubicek, L. Herman, "Dynamic visualization of sensor measurements: Context based approach," *Quaestiones Geographicae* 2015, 34 (3), pp. 117-128. DOI: 10.1515/quageo-2015-0020.

[8] F.J. Pierce, P. Nowak, "Aspects of precision agriculture," *Advances in Agronomy* 1999, 67, pp. 1-85.

[9] R. Bongiovanni, J. Lowenberg-Deboer, "Precision agriculture and sustainability," *Precision Agriculture* 2004, 5, pp. 359-387.

[10] R. Gebbers, V.I. Adamchuk, "Precision agriculture and food security," *Science* 2010, 327, pp. 828-831.

[11] The Linked Open Data Cloud, created 2007, accessed 2 May 2018, available at: http://lod-cloud.net/.

[12] T. Reznik, K. Charvat jr., K. Charvat, S. Horakova, V. Lukas, M. Kepka, "Open Data Model for (Precision) Agriculture Applications and Agricultural Pollution Monitoring," In Proceedings of Enviroinfo and ICT for Sustainablity 2015, ACSR-Advances in Computer Science Research, 22, Atlantis Press, Paris, France, pp. 97-107

[13] R. Palma, T. Reznik, M.A. Esbri, K. Charvat, C. Mazurek, "An INSPIRE-Based Vocabulary for the Publication of Agricultural Linked Data," *Ontology Engineering* 2016, 9557, pp. 124-133, DOI: 10.1007/978-3-319-33245-1_13.

[14] Jurik, V.; Herman, L.; Sasinka, C.; Stachon, Z.; Chmelik, J. When the Display Matters: A Multifaceted Perspective on 3D Geovisualizations. In Open Geosciences 2017, 9 (1), pp. 89-100. DOI 10.1515/geo-2017-0007.

[15] T. Reznik, K. Charvat, Z. Krivanek, S. Horakova, K. Charvat jr., M. Kepka, J. Jezek, "The Cartographical Visualization of Crowdsourcing Data in Precision Agriculture," In *Proceedings of the International Cartographic Conference* 2017, Washington, United States of America, 7 p.

[16] J. Jezek, K. Jedlicka, T. Mildorf, J. Kellar, D. Beran, „Design and Evaluation of WebGL-Based Heat Map Visualization for Big Point Data," *Rise of Big Spatial Data - Lecture Notes in Geoinformation and Cartography* 2017, Springer, Cham, Switzerland, pp. 13 – 26, DOI: 10.1007/978-3-319-45123-7_2.

[17] H. Skokanova, "Can we combine structural functionality and landscape services assessments in order to estimate the impact of landscape structure on landscape services?," *Moravian geographical reports* 2013, 21, pp. 2-15, DOI 10.2478/mgr-2013-0016.