Visualisation of Big Data in Agriculture and Rural Development

Karel JEDLIČKA[1], Karel CHARVÁT[2]
[1]*University of West Bohemia, Univerzitní 8, Plzeň, 306 14, Czechia*
Tel: +420377639210, Email: smrcek@kgm.zcu.cz
[2]*Lesprojekt s.r.o, Martinov 197, 277 13 Záryby, Czechia*
Tel: +420604617327, Email: charvat@lesprojekt.cz

Abstract: 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). Big data analytics system will then provide pilot managers with highly localized descriptive (better and more advanced way of looking at an operation), prescriptive (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) and predictive plans (use current and historical data sets to forecast future localized events and returns). Presentation will be focused on two completely new domains of Big Data Visualisation and Analysis for Agriculture

Keywords: Agriculture, Rural Development, Land Use, Big Data, Visualisation

1. Introduction

The agriculture sector is of strategic importance for European society and economy. Due to its complexity, agrifood operators have to manage many different and heterogeneous sources of information. 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.

Big data is moving into agriculture in a big way. Number of new technologies is influencing farming nowadays: Sensors on fields and crops are starting to provide data points on soil conditions, as well as detailed info on wind, fertilizer requirements, water availability and pest infestations, GPS units on tractors, can help determine optimal usage of agriculture machinery, Unmanned aerial vehicles, or drones, can patrol fields and alert farmers to crop ripeness or potential problems, RFID-based traceability systems can provide a constant data stream on farm products as they move through the supply chain, from the farm to the compost or recycle bin. Individual plants can be monitored for nutrients and growth rates. 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). But there is necessary to mention, that farmers primary focused not on (big) data, but on a knowledge generated from this data. Therefore it is important to better understand basic terms such as data, information and knowledge for every discussion about knowledge or information management.

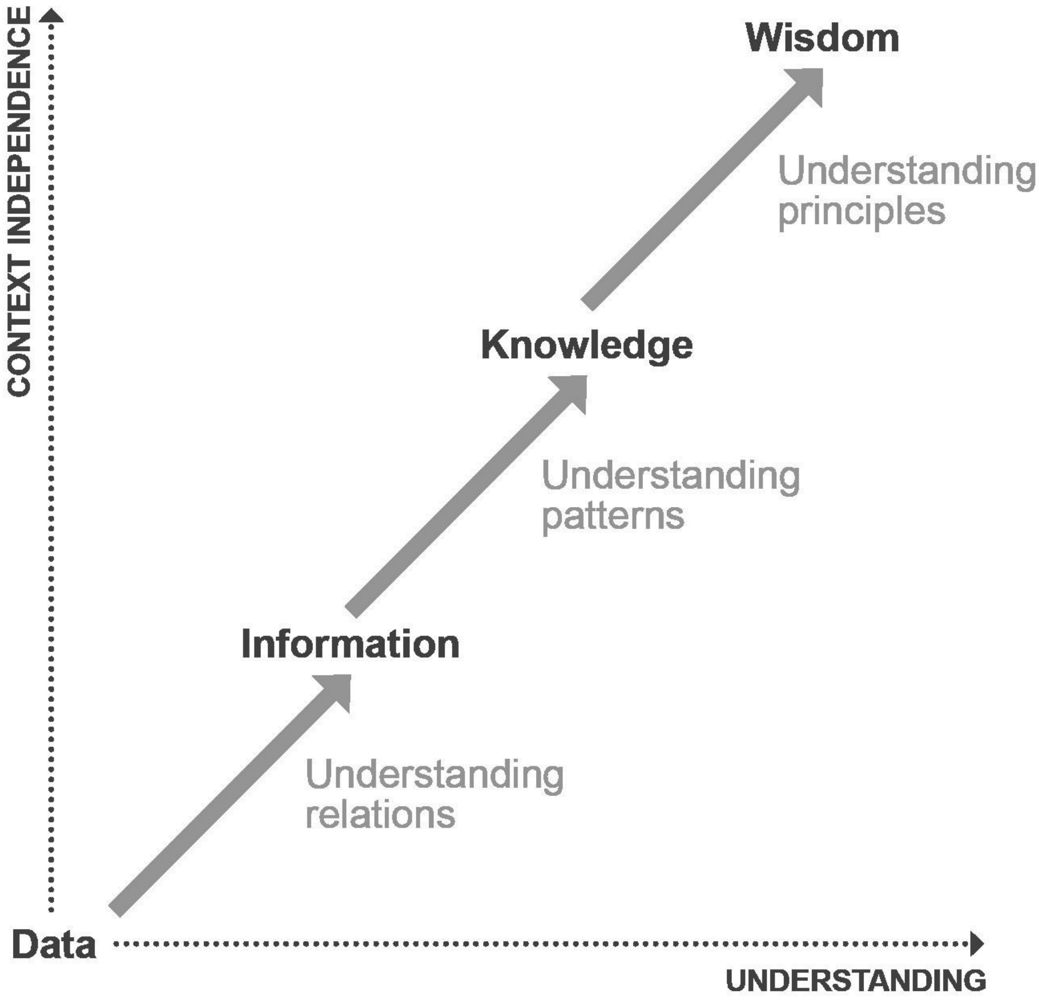

Figure 1. The Data-Information-Knowledge-Wisdom hierarchy of Ackoff [1],[2][1]

For better explanation we will use Data-Information-Knowledge-Wisdom hierarchy Ackoff (see figure 1) including Data: as symbols; Information: data that are processed to be useful; provides answers to "who", "what", "where", and "when" questions; Knowledge: application of data and information; answers "how" questions and Wisdom: evaluated understanding. This contribution describes on best practise examples, how DataBio Reference Architecture and analytical tools support transformation of agriculture related data into farmer's wisdom.

2. Objectives

One of the methods, how to transform Big Data into Wisdom is their advanced visualisation. The objective of this paper is focused on Big Data Visualisation for Agriculture and rural development, particularly on 3D visualization of:

- Agriculture and rural development data
- Predicted crop yield
- Linked Open Data related to agriculture

Displaying the data in 3D environment helps a user to explore an area of interest in more natural way than a traditional map.

[1] Picture adopted from https://www.linkedin.com/pulse/we-have-knowledge-use-well-howard-thomas/

3. Methodology

The focus of this contribution is on visualisation any customer data related to agriculture and rural development. To sum up, following types of data can be distinguished:

- non-structured data
- data structured in:
    - general graph (linked data, RDF)
    - tree (hierarchical data, such as XML or JSON based datasets)
    - tables and relationships (entity-relationship-attribute model, database)

In fact, all the data structures can be represented as a graph, but the different types of structure influences a typical way of storage of the data (see examples in parentheses).

Moreover there is often useful to deal with a spatial aspect of the data, as the majority of data can be usually related to some particular place or area. Spatial aspect of the data is usually represented by:

- mosaics:
    - regular rasters with squared cells (imagery ~ thus all Earth Observation Data, continuous phenomena ~ such as temperature distribution among an area of interest, small and middle scales of digital terrain models, etc.)
    - irregular triangulated meshes (used for detailed digital terrain model representation)
- vectors
    - point, line, polygon (representing discrete objects ~ both man made features such as buildings, roads, and natural objects such as forests, rivers, water areas, etc. such a representation can be used e.g. for objects extracted from imagery as well as objects gathered from other sources - maps, directly measured data, etc.)

Also it should be mentioned that both raw and processed data can be visualized, from the agriculture point of view:

- Raw data for picking the right dataset for further data processing.
- Processed (transformed / harmonized / analyzed / …) data for exploration the results and decision support.

The methodology followed in the developments consisted in selecting a proper, open and well extendable technologies (see chapter 4) and developing three best practise examples to show, how such a technology set can be used for visualization of various data types (see chapter 5).

4. Technology

From the above mentioned diversity of the data storage structures it comes clear that robust and easily customizable applications should be used for data visualization

The raw data visualization is usually done by a generic client application, handled by a professional, who can deal with different data structures and formats. These applications are not described here as there is many of them in the market or as an open source solutions (QGIS[2] for example, to mention at least one application of many).

[2] https://www.qgis.org/

The common technological framework selected for further best practises development consists of HSLayers NG[3] ~ a JavaScript based library for geodata visualization, with an integrated Cesium plugin[4] for 3D visualisation. This framework was then extended for handling various types of data, namely:

- Web Map Service ~ for raster and imagery data
- GeoJSON ~ for vector data
- Resource Description Framework (RDF) ~ for linked data
- OpenStreetMap live data pump ~ for vector data from OSM

The framework development was then divided into following steps:

- Integration of CESIUM with HSlayers NG (see folders examples/3d-olu and components/cesium at https://github.com/hslayers/hslayers-ng repository for details).
- Installing own CESIUM server (http://cesium.cenia.cz)
- Publishing local DEMS on own CESIUM server

Created framework allows a developer to create applications his/her own tailored applications. Last but not least the framework supports responsive design.

5. Developed best practise applications examples

The applications mentioned further are best practice examples of processed data visualization tailored for the purposes of the DataBio project Data Experimentation and Proof of Concept phases. The applications were created by using the above mentioned framework. The work about this topic started in previous project FOODIE[5] and now continues as a part of Czech agriculture pilots of DataBio project[6]. To speed up a development, three new large scale testbeds were developed as part of INSIRE Hack[7].

5.1 Open Land Use Perspective Visualization

The Open Land Use perspective visualization is available at following address:
http://ng.hslayers.org/examples/3d-olu

This application visualises the Open Land Use Map (OLU)[8] on top of the EU-DEM terrain model in a perspective view; note that a geographical extension of OLU for Africa is planned in DataBio project[9]. Displaying the data in 3D environment helps a user to explore an area of interest in a more natural way than a traditional map. Moreover, the user can explore other datasets, such as Open Transport Map[10] or Smart Points Of Interests[11], [3] or even add a Web Map Service of his/her choice (as long as the WMS is in WGS84 coordinate system) to make custom 3D mash-ups.

It is worth to mention that the application uses a strategy to visualize a continent-wide dataset (OLU) in 3D environment. The strategy consists of two features (adopted from [4]):

[3] https://github.com/hslayers/hslayers-ng
[4] https://cesiumjs.org/
[5] http://www.foodie-project.eu/
[6] http://databio.eu/
[7] http://www.plan4all.eu/inspire-hack-2017/
[8] http://sdi4apps.eu/open_land_use
[9] http://www.plan4all.eu/2017/07/open-land-use-map-4-africa/
[10] http://opentransportmap.info
[11] http://sdi4apps.eu/spoi

- Generalized data is portrayed when an observer is far from the Earth surface visualized on the virtual globe visualized by Cesium plugin. Detailed data are segmented by municipalities.
- When the observer moves closer following steps are processed:
  - A visible spatial extent on Earth's surface is calculated from an actual view.
  - Bounding boxes[12] for municipalities are then compared to the visible spatial extent.
  - Finally, displayed are the only municipalities, whose bounding boxes intersected with the Earth's surface visible spatial extent.
  - These steps are done repeatedly each time when observer position nor line of sight is changed, as it influences the actual view.

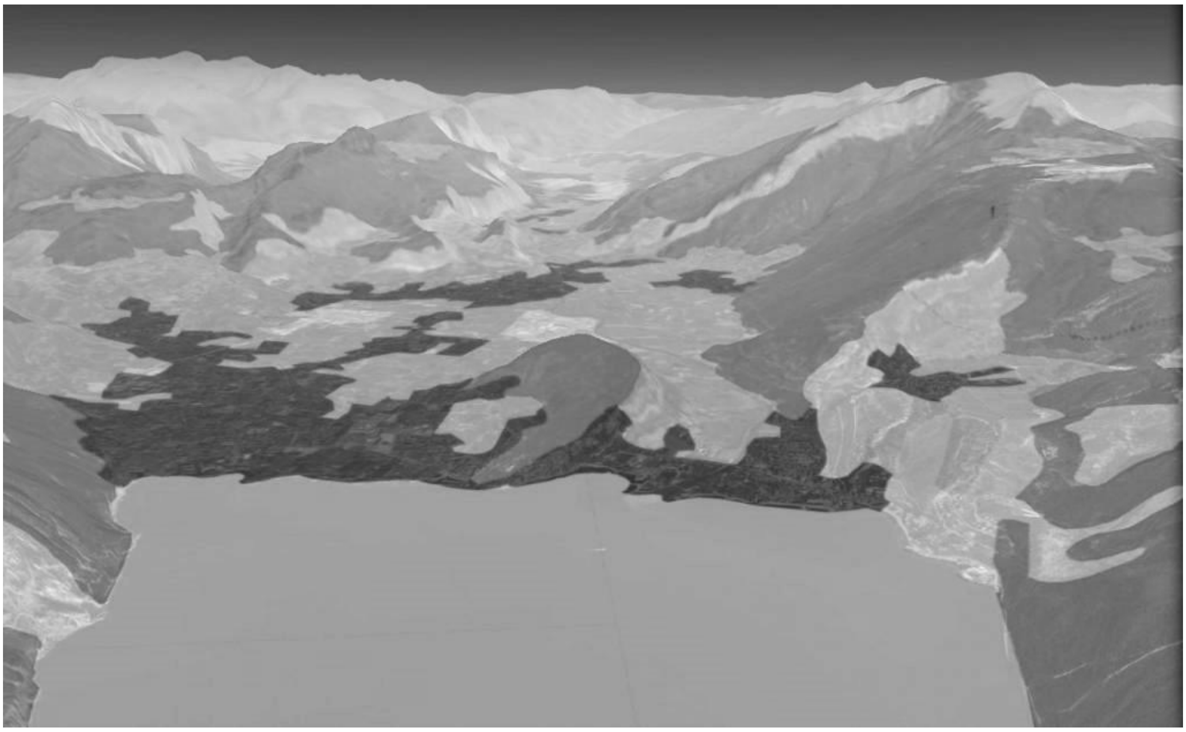

Figure 2. 3D Open Land Use visualisation

5.2 Perspective visualization of estimated yield - Rostěnice farm

As the yield is an integrator of landscape and climatic variability, it provides useful information for identifying management zones [5]. These zones are understand as areas with the same or similar yield level within the fields. Such identified zones present a fundamental delineation for site specific crop management.

The zones delineation is usually based on yield maps over the past few years. Similar to the evaluation of yield variation from multiple yield data described by Blackmore et al. [6] the aim is to identify high yielding (above the mean) and low yielding areas related as the percentage to the mean value of the field. Also the inter-year spatial variance of yield data is important for agronomists to distinguish between areas with stable or unstable yields. But the presence of complete series of yield maps for all fields is rare,

---

[12] Bounding box is an expression of the maximum extents of a 2-dimensional object (e.g. point, line, polygon) or set of objects within its (or their) 2-D (x, y) coordinate system, in other words min(x), max(x), min(y), max(y).

thus satellite imagery can be analysed to determine the desired in field variability of crops thru vegetation indices where yield maps are missing [7].

The estimation of yield potential zones from multi-temporal satellite data is established as the general model in FOODIE platform [8]. As the main data source, ESPA repository of LANDSAT satellite images is used, which offers surface reflectance products, main vegetation indices (NDVI, EVI) and clouds identification by CFmask algorithm. A selection of scenes from recent 8 years was made for the Rostěnice farm area to collect cloud-free data related to second half of vegetation period. An estimated yield was calculated for separate scenes as the relation of each pixel to mean value of whole field. In last step all scenes were combined and median value of yield potential was calculated. After fully operation of Sentinel 2A/B satellites, calculation of yield potential can be enhanced by its vegetation products.

The Rostěnice application (available at http://ng.hslayers.org/examples/rostenice) visualises the crop yield dataset calculated from (satellite imagery as mentioned above) on top of the EU-DEM terrain model in a perspective view.

Such a visualization can help a farmer to better understand the farm fields. The farmer can explore the relation of the estimated crop yield to the topography, slope, orientation and topography wetness index in his/her field. The farmer can check the parts of fields with steep slopes and how the machinery deals with them (by checking the machinery tracklogs, where available), see more e.g. in [9].

The application uses a digital terrain model of fifth generation and the digital surface model of first generation produced by ČÚZK[13].

Figure 3 Rostěnice Farm Visualisation

5.3 Linked data integration to 3D virtual environment - Vineyards example

This application (available at http://ng.hslayers.org/examples/produce-3d) visualises management zones of different vine types, describing the type of vine by color and label, and the productivity of each zone using vertical exaggeration of a polygon representing particular vineyards. After clicking on a vineyard, attribute data from the vineyard layer are

---

[13] http://www.cuzk.cz/en

displayed together with the linked data from dbpedia[14] and other semantic sources. The aim of this example is to show integration of linked data and spatial data together.

Figure 4. Linked data integration to 3D virtual environment

6. Discussion & Conclusion

6.1 Technology

The described technological framework is functional and have a potential to visualize various types of data by using appropriate approach. There still runs ongoing development of the framework consisting in extending the dimension of visualizable data from three to four dimension by using a CZML[15] data format for time aware data. There is also plan to implement a graphical user interface to allow user to add his/her own data without a need of coding.

6.2 Developed applications

3D visualisation is bringing new potential into analysis of Big data in the field of agriculture. Current experiments with the Open Land Use perspective visualization cover all Europe with geographic (OLU, EU-DEM, OTM) and linked data (SPOI) and have 27 million of objects. Moreover, a further geographical extension of Open Land Use Map to Africa is planned in the DataBio project.

The exploration of yield estimated from satellite data with respect to topography (represented as DTM, DSM, slope, slope orientation, and topography wetness index) is takes an advantage of a perspective visualization of 3D data comparing to two dimensional map (as can be seen in the Rostěnice application). Further planned development heads two major directions: displaying and analysis of machinery paths on field; and a real challenge ~ enabling the estimation of yield from satellite imagery as a web service.

---

[14] http://wiki.dbpedia.org/

[15] https://github.com/AnalyticalGraphicsInc/czml-writer/wiki/CZML-Guide

Acknowledgements

The research reported in this paper has been supported by the Farm Oriented Open Data In Europe - FOODIE project (http://foodie-project.eu/, CIP-ICT-PSP-2013-7, Pilot B , no.621074), Data-Driven Bioeconomy – DataBio ICT-15-2016-2017; Big Data PPP: Large Scale Pilot actions in sectors best benefitting from data-driven innovation.

References

[1] Bellinger, G., Castro, D., Mills, A., *Data, Information, Knowledge and Wisdom*, on-line: http://www.systems-thinking.org/dikw/dikw.htm (cit. 9. 4. 2018)

[2] Charvát, K., Horáková, Š., Mildorf, T., Křivánek, Z. Strategic Research Agenda for introduction of innovative approaches on data exchange in agri-food industry, *IST-Africa 2013 Conference Proceedings*, Paul Cunningham and Miriam Cunningham (Eds), IIMC International Information Management Corporation, 2013, ISBN: 978-1-905824-38-0.

[3] Vohnout, P., Cerba, O., Kafka, S., Fryml, J., Krivanek, Z., & Holy, S. (2014, May). SmartTouristData approach for connecting local and global tourist information systems. In *IST-Africa Conference Proceedings, 2014* (pp. 1-6). IEEE.

[4] Charvát, K., Mildorf, T., Bye, B., L., Berre, A.,J., Jedlička, K. Open Data, VGI and Citizen Observatories INSPIRE Hackathon. *In International Journal of Spatial Data Infrastructures Research.* In Press 2018

[5] Kleinjan, J., Clay, D. E., Carlson, C. G., & Clay, S. A. (2007). Productivity zones from multiple years of yield monitor data. *In F. J. Pierce, & D. C. Clay, GIS applications in agriculture*. CRC Press, Boca Raton.

[6] Blackmore, S., Godwin, R. J., & Fountas, S. (2003). *The Analysis of Spatial and Temporal Trends in Yield Map Data over Six Years. Biosystems Engineering.*

[7] Charvát K., Řezník T., Lukas V., Charvát K., Jr., Horáková, Š., Kepka, M., Šplíchal, M. QUO VADIS PRECISION FARMING, *In 13th International Conference on Precision Agriculture Proceedings*, July 31 – August 4, 2016, St. Louis, Missouri, USA

[8] Řezník, T., Charvát, K., LUKAS, V., Charvát Jr, K., Horáková, Š., & Kepka, M. (2015). Open Data Model for (Precision) Agriculture Applications and Agricultural Pollution Monitoring. *Proceedings of the Enviroinfo and ICT for Sustainability*.

[9] Řezník, T., Lukas, V., Charvát, K., Charvát Jr, K., Horáková, Š., Křivánek, Z., & Herman, L. (2016). Monitoring of in-field variability for site specific crop management through open geospatial information. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 41*.