Big data in agriculture – from FOODIE towards data bio

Karel Charvát*[1], Tomas Reznik[2] , Vojtech Lukas[3], Sarka Horakova[4], Karel Charvát Jr.[1], Michal Kepka[5], Marek Splichal[1], Simon Leitgeb[2], Jan Shanel[2], Karel Jedlicka[5], Jaroslav Smejkal[6]

1 Lesprojekt služby
2 Masaryk University
3 Mendel University
4 WirelessInfo
5 University of West Bohemia
6 Zetor company

Abstract

What's the role of Big Data in the farming ecosystem? Farmers need to measure and understand the impact of a huge amount and variety of data which drive overall quality and yield of their fields. Among those are local weather data, GPS data, ortophotos, satellite imagery, soil specifics, soil conductivity, seed, fertilizer and crop protectant specifications and many more. Being able to leverage this data for running long and short term simulations in response to "events" like changed weather, market need or other parameters is indispensable for farmers in terms of maximizing their profits. IoT (Internet of Technology) including field sensors and machinery monitoring. The experimentation in FarmTelemetry project demonstrates that one average Czech farm (i.e. around 1'000 hectares) could generate daily 20 MegaBytes of data. This could be only for Czech Republic something between 30 and 50 GB per one day. We may easily reach Terabytes of data a day from agricultural basic monitoring by sensors in Europe. Together with satellite data agriculture will need to manage extremely large amount of data. On one side there is growing whole ecosystem with a strong need to secure Big Data from different repositories and heterogeneous sources. In some cases, sharing of data could be common interest, but on other side, there are also different interests and data could help to one part of value chain to take bigger part of profit. From this reason Big data are sensitive topics and trusting of producers about data security is essential. The producers of seeds and chemicals want to maximize their business with farmers. Our team stated implementation of Big Data technologies in frame of European 7FP project FOODIE. This work currently the work continue as part of DataBio project.

Background

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. 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.

DataBio project is the Data-Driven Bioeconomy, focusing in production of best possible raw materials from agriculture, forestry and fishery/aquaculture for the bioeconomy industry to produce food, energy and biomaterials taking into account also various responsibility and sustainability issues. DataBio project re use and continue on development of FOODIE results DataBio proposes to deploy a state of the art, big data platform "on top of the existing partners" infrastructure and solutions - the Big DATABIO Platform. The work will be continuous cooperation of experts from end user and technology provider companies, from bioeconomy and technology research institutes, and of other partners. In the pilots also associated partners and other stakeholders will be actively involved.

Methods

The FOODIE project focused on building an open and interoperable agricultural specialized platform hub on the cloud for the management of spatial and non-spatial data relevant for farming production; for discovery of spatial and non-spatial agriculture related data from heterogeneous sources; integration of existing and valuable European open datasets related to agriculture; data publication and data linking of external agriculture data sources contributed by different public and private stakeholders al lowing to provide specific and high-value applications and services for the support in the planning and decision-making processes of different stakeholders groups related to the agricultural and environmental domains.

The **main idea** of FOODIE project is to:

> build an open and interoperable agricultural specialized platform hub on the cloud for the management of spatial and non-spatial data relevant for farming production
> discovery of spatial and non-spatial agriculture related data from heterogeneous sources
> integration of existing and valuable European open datasets related to agriculture
> data publication and data linking of external agriculture data sources contributed by different public and private stakeholders allowing to
> provide specific and high-value applications and services for the support in the planning and decision-making processes of different stakeholders groups related to the agricultural and environmental domains.

which is conceptualized in the following architecture diagram of FOODIE service platform (Figure 2). (Charvat et al. 2014, May), (Charvat et al. 2016, July).

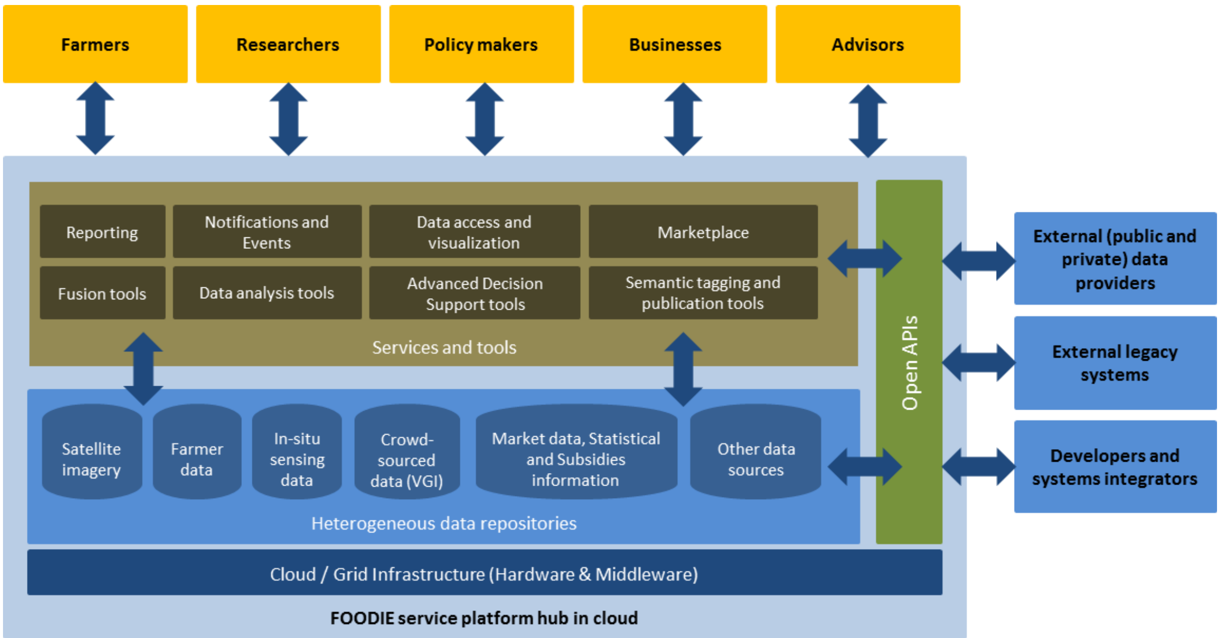


Figure 1. FOODIE service platform hub in cloud

DataBio use reference architecture based on BDVA reference architecture. The reference architecture of DataBio define set of basic modules, which could be implemented using different tools. This reference architecture is on Fig 2.

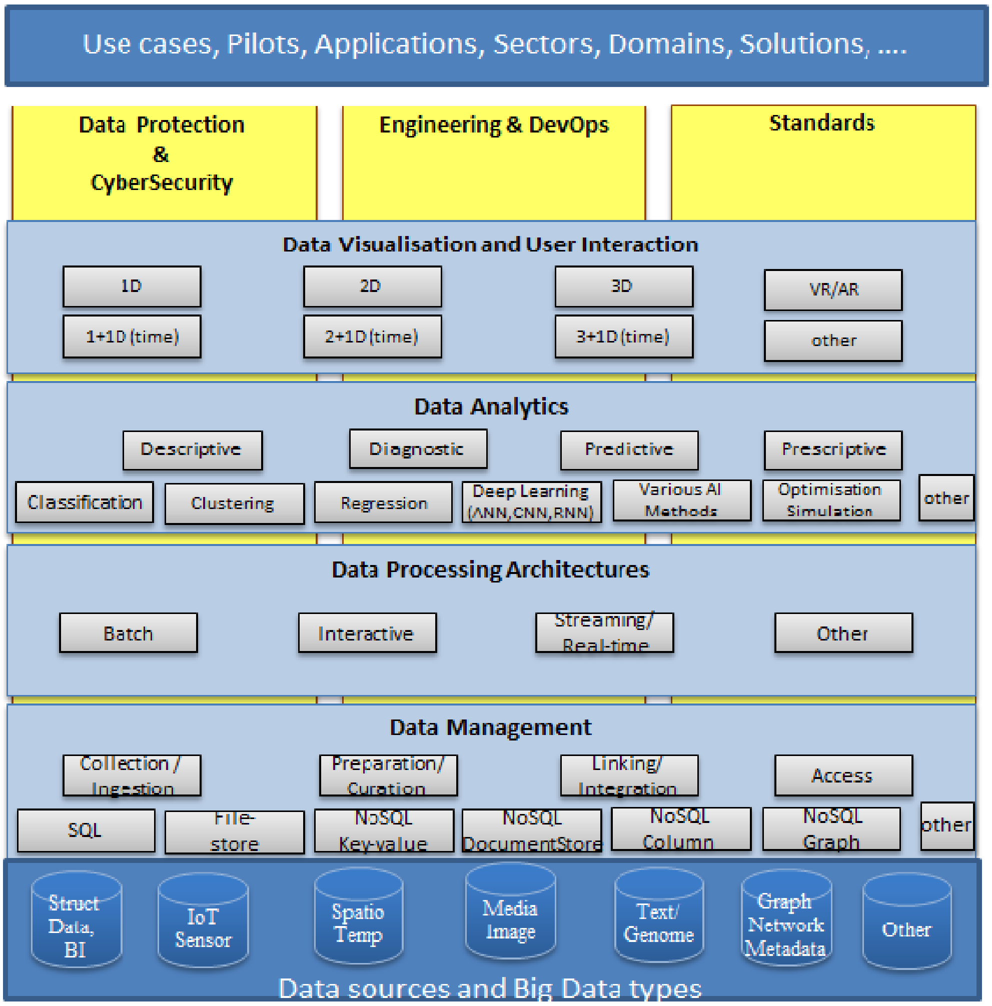


Figure 2. DataBio reference architecture

FOODIE Czech Pilots

The Farm Telemetry system was focused on monitoring of activities and utilization of individual tractors. Data collected by sensor units on tractor allow possibility of analyses and obtaining overall overview for individual farmer's fields (blocks). The field trials during season demonstrate that Farm Telemetry will deal with real Big data. This required optimization on the level of database structures and communication protocols. The main problem is that data from every farm has to be stored minimally for full season, to guarantee complex analysis of costs.

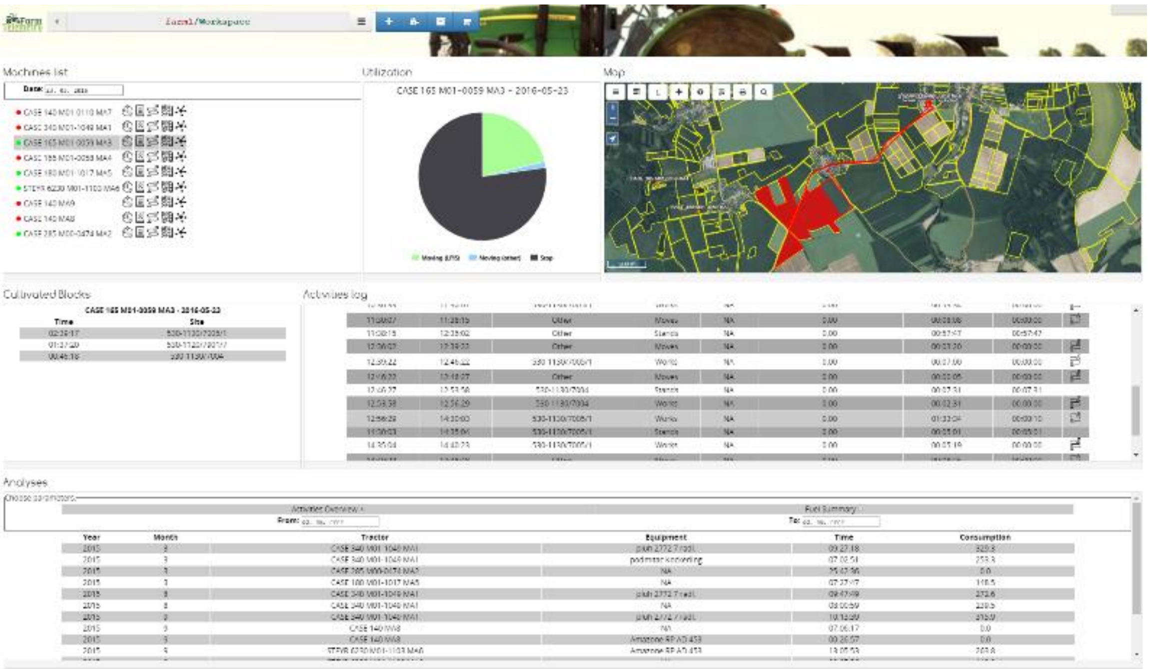


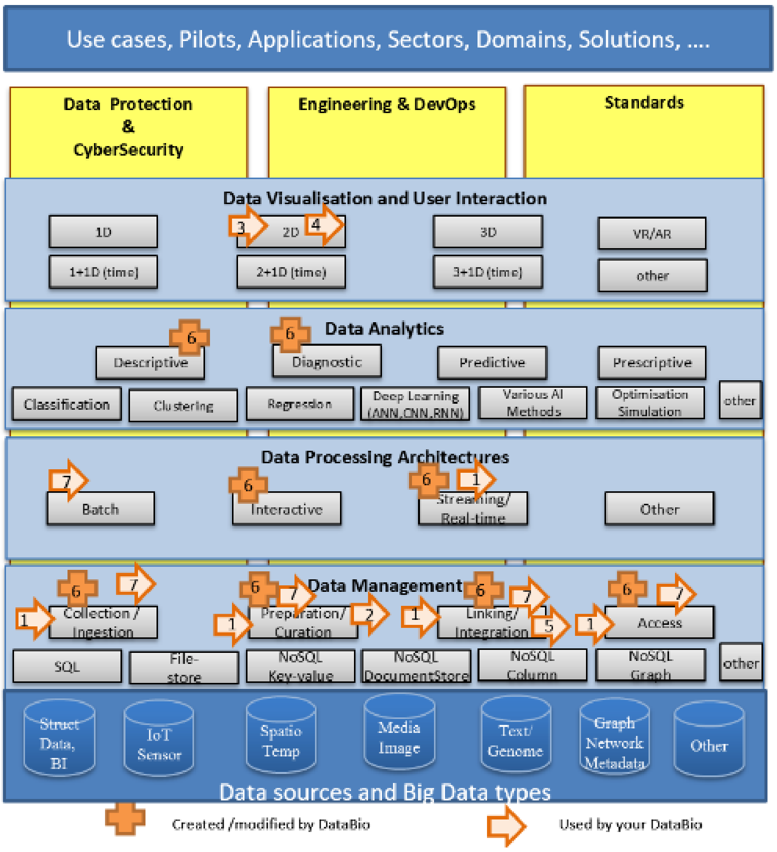
Figure 3. FOODIE farm telemetry

Yield potential zones are areas with the same yield level within the fields. Yield is the integrator of landscape and climatic variability and therefore provide useful information for identifying management zones (Kleinjan, Clay, Carlson, & Clay, 2007). This presents a basic delineation of management zones for site specific crop management, which 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. (Blackmore, Godwin, & Fountas, 2003), 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. The presence of complete series of yield maps for all fields is rare, thus remote sensed data are analysed to determine in field variability of crops thru vegetation indices.

Figure 4. Fertilizer planning application based on yield potential zones

Czech DataBio pilot machinery management

Machinery management is focused mainly on collecting telemetry data from machinery and analysing them in relation with other farm data. The main challenge is access to data and data integration, when farmer uses tractors and equipment from various manufacturers with different telematics solutions and different data ownership/sharing policy. In many cases farms or agriculture service organizations owns tractors of more than one brand/family. Although the communication protocols used in control units of farm machinery and data collection are subject of standardization, the telematics solutions including data ownership/usage policy are usually specific to each tractor brand/family and the level. Furthermore, attention shall be paid to ISO and CEN standards regulating data sharing in agriculture basing on the input coming from industry organizations like CEMA and AEF. Although this is not issue and can be even desirable for purposes of tractor producer's customer care responsible for solving technical problems on tractor, for farmers it can be hard or impossible to connect the data coming from tractor with other farm data relevant for agronomical / economical evaluation of machinery usage. Despite the fact tractor have telematics solution, farmer sometimes need to use third party device and software to obtain data for field specific analysis. Zetor Company is currently developing and testing modular telematics solution which is supposed to be part of all Zetor tractors. The solution will provides several levels of functionality ranging from basic telematics for customer care and basic location information for customer to field specific economic analysis and precision agriculture. The highest level of modular solution will offer connection to other data relevant for farm management like field boundaries obtained Land Parcel Information system (LPIS), elevation model and possibly yield potential maps derived from EO data. This connection will enable evaluation of economic efficiency and other analysis on the level of individual fields or even parts of the field. Expected utilisation of DataBio tools is on Fig 5.



1. SensLog
2. Micka
3. HsLayersNG
4. HSLayers-NG Cordova mobile application
5. Data Model for (Precision) Agriculture
6. Farm Telemetry
7. EO data processing pipeline

Figure 5. Machinery management architecture

Czech pilot cereals and biomass
The pilot aims to develop a platform for mapping of crop vigor status by using EO data (Landsat, Sentinel) as the support tool for variable rate application (VRA) of fertilizers and crop protection. This includes identification of crop status, mapping of spatial variability and delineation of management zones. The main focus of the pilot will be on the monitoring of cereal fields by high resolution satellite imaging data (Landsat 8, Sentinel 2) and delineation of management zones within the fields for variable rate application of fertilizers. The main innovation is to offer a solution in form of web GIS portal for farmers, where users could monitor their fields from EO data based on the specified time period, select cloudless scenes and use them for further analysis. This analysis includes unsupervised classification for defined number of classes as identification of main zones and generating prescription maps for variable rate application of fertilizers or crop protection products based on the mean doses defined by farmers in web GIS interface. The used tools from DataBio reference architecture are on Fig 6.

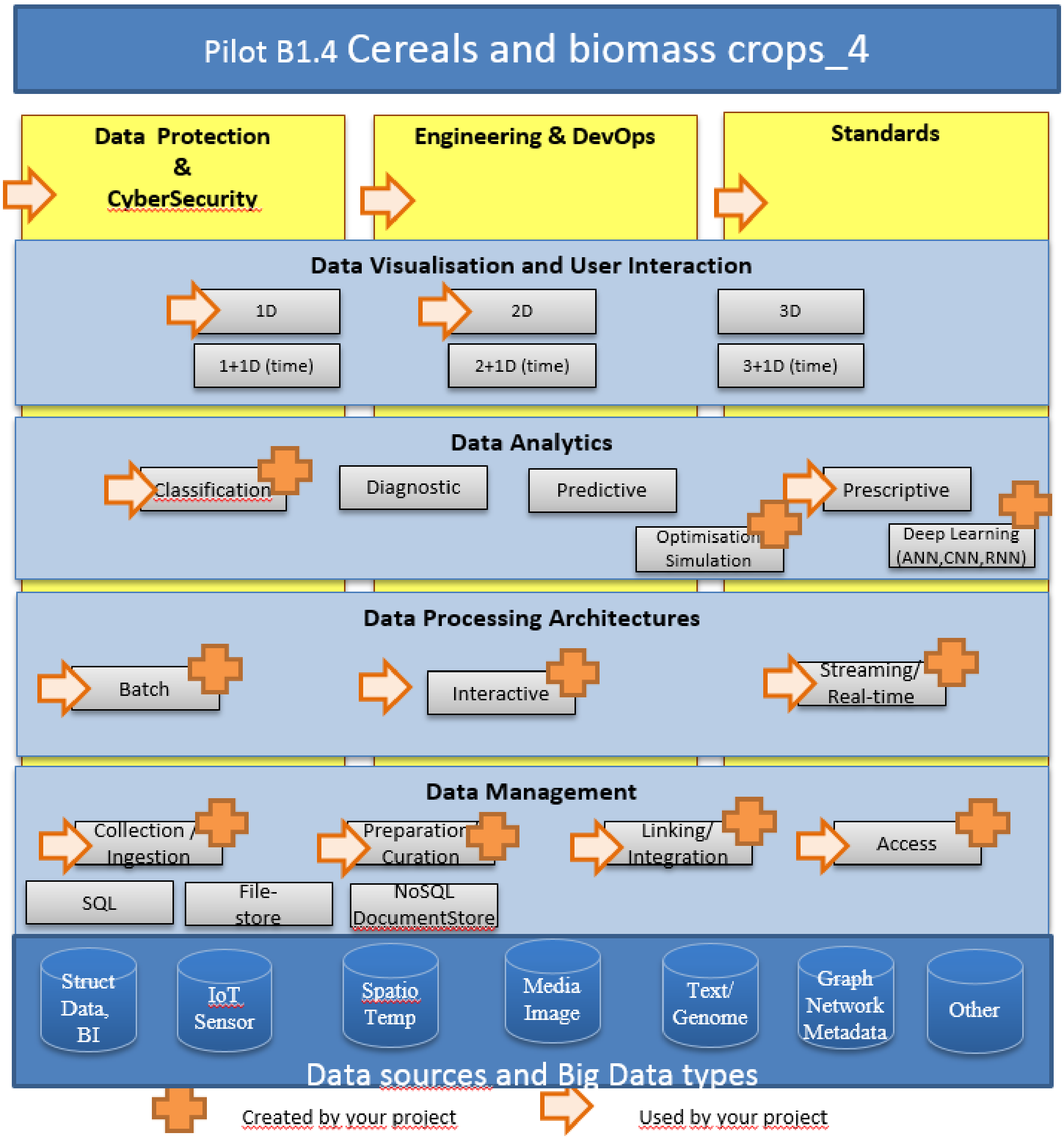


Figure 6. Cereal and biomass pilot architecture

Selected components

SensLog - SensLog is web-based sensor data management tool. Observations produced by external devices are processed by SensLog. SensLog is focused both on static monitoring devices and mobile devices with live tracking ability.

MIcKA is complex platform comprises a Web Catalogue Service as well as a metadata editor and manager provides the integrated solution required to publish and access digital catalogues of metadata for geospatial data, services, and related resource information. As such it enables to search for as well as maintain descriptions (metadata) on datasets, dataset series, Web services, etc. according to ISO/OGC/W3C standardization documents including Semantic Web principles.

Hlayers NG is a web mapping library written in JavaScript. It extends OpenLayers 4 functionality and takes basic ideas from the previous HSlayers library, but uses modern JS frameworks instead of ExtSJS 3 at the frontend and provides better adaptability.

HSLayers-NG Cordova mobile application - support of mobile platforms is important feature of HSL-NG development. Desktop HSL have responsive design but we are also working on special mobile application using Apache Cordova framework. Current version brings big part of HSL functionality (e.g. compositions, layer manager, search).

WebGLayer is a JavaScript library focused on fast interactive visualization of big multidimensional spatial data through linked views. The library is based on WebGL and uses GPU for fast rendering and filtering. Using commodity hardware the library can visualize hundreds of thousands of features with several attributes through heatmap, point symbol map. Users thus benefit from immediate and dynamic data visualizations, gain better understanding of data by applying filters, and develop the opportunity to discover relationships and patterns in the data.

Conclusion

Experience and tools from FOODIE are now used in DataBio in building generic Big Data platform for agriculture.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 732064 called "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)

References

Blackmore S, Godwin RJ, Fountas S 2003. The analysis of spatial and temporal trends in yield map data over six years. Biosystems Engineering.

Charvat K, Esbri MA, Mayer W, Campos A, Palma R, Krivanek Z 2014. FOODIE - Open data for agriculture. Ist Africa Conference Proceedings. 1–9 pp.

Charvat K, Horakova S, Wolfert S, Holster H, Schmid O, Pesonen L, Mildorf T 2012. Common basis for policy making for introduction of innovative approaches on data exchange agri-food industry. Final Strategic Research Agenda (SRA).

Kleinjan J, Clay DE, Carlson CG, Clay SA 2007. Productivity zones from multiple years of yield monitor data. *In*: Pierce FJ, Clay DC. GIS applications in agriculture.

Oerke EC, Gerhards R, Menz G, Herbert GW 2010. Precision crop protection - the challenge and use of heterogeneity. DOI: 10.1007/978-90-481-9277-9.

Řezník T, Lukas V, Charvát K, Charvát KJ, Horáková Š, Kepka M 2016. FOODIE data Model for precision agriculture. 13th International Conference on Precision Agriculture. 31 July – 4 August, St Louis, Missouri, USA.

Charvát K, Řezník V. Lukas V, Charvát K. Jr, Horáková S, Kepka M, Šplíchal M 2016. Quo Vadis precision farming. 13th International Conference on Precision Agriculture, July 31 – August 4, 2016, St. Louis, Missouri, USA.