ORIGINAL SCIENTIFIC PAPER

Promoting digital innovation: Identifying the key elements in agricultural open data ecosystem in Croatia

Dragica Šalamon[1], Martina Tomičić Furjan[2], Alen Džidić[1], Filip Varga[1,3], Larisa Hrustek[2], Bastiaan van Loenen[4]

[1]*Faculty of Agriculture, University of Zagreb, Svetošimunska 25, Zagreb, Croatia, (dsalamon@agr.hr)*
[2]*Faculty of Organization and Informatics, University of Zagreb, Pavlinska 2, Varaždin, Croatia*
[3]*Centre of Excellence for Biodiversity and Molecular Plant Breeding, Svetošimunska 25, Zagreb, Croatia*
[4]*Faculty of Architecture and the Built Environment, Knowledge Centre Open Data, Delft University of Technology, Julianalaan 134, Delft, Netherlands*

Abstract

Open data is critical for digital innovation in the sector of agriculture. We aimed to identify the four key elements of the open data ecosystem for the agricultural domain in Croatia. 357 data files available via the Croatian Open data portal, the internet pages of the Ministry of agriculture, the Croatian agency for agriculture and food, as well as the Paying Agency for Agriculture, Fisheries and Rural Development were analysed. The main G2B publishers of the agriculture domain have not yet reached their potential in opening data. There are still many raw and aggregated datasets not openly available. The published data lack a suitable format for data linking and information on licencing.

Key words: open data, agriculture, ecosystem elements, publisher, dataset

Introduction

As the basis of digital transformation, access to more comprehensive data with broad usage capabilities enables innovation for the progress of the European economy and the betterment of citizens (Davies, 2012). European and Croatian strategic logistics and legislation for the opening and reuse of data as an institutional framework for the implementation of open data policy has been developing intensively since 2013 (Musa et al., 2018). In 2018, a code of conduct was developed for the exchange of agricultural data under the contract, and in 2019, EU Member States signed a declaration of cooperation "A smart and sustainable digital future for European agriculture and rural areas", which recognises the potential of digital technologies for the agricultural sector and rural areas and supports the establishment of data spaces.

Given the projected global increase from 33 zettabytes of open data in 2018 to 175 zettabytes of open data in 2025 (Gens et al., 2017) the current European Data Strategy predicts a near doubling in the number of data experts in the EU since 2018 (to 10.9 million), as well as an increase in the population with basic digital skills from 57% to 65%. The value of the data economy in the EU is estimated at EUR 829 billion Euro in 2025, representing almost a triple from 2018 (EC, 2020). The expected reversal of the processing and analysis of data from centralized computing infrastructure to network-edged computing by 2025 (Panetta, 2016) will provide additional opportunities for data manufacturers to manage the data generated and a new perspective for the development of SMEs' and MMEs' products and services. For this reason, a significant impact on the way production, consumption and life is expected from rationalized energy consumption, better traceability of products, materials

and food, to a healthier life. The strategy highlights the importance of SMEs, and increasing the functionality of G2B ("government to business") and B2G ("business to government") and B2B ("business to business") data exchange (EC, 2020). Innovation based on open access to data and open use and reuse of data between sectors will also contribute to the European Green Plan.

The priority sector for addressing issues such as connectivity, data processing and storage, computing power and cybersecurity is precisely the European agricultural information space within which a European neutral platform for the exchange and assembly of agricultural data is planned into repositories, including private and public data. The role of sector data spaces is to foster the development of an open data ecosystem (businesses, civil society and individuals) where new products and services are created on the basis of more accessible data. For the precise and tailored application of production approaches at farm level, it is necessary to strengthen digital, in particular data, literacy in the sector (EC, 2020; Šalamon et al., 2021). The flow of production data, in particular in use with other data from agricultural but other sectoral data sources is essential for agricultural production. For the agriculture sector, they include socio-economic, legislative and administrative, environmental and natural resources data (Šalamon, 2019).

As technologies and innovations arise on the basis of open data, not the other way around (Musa et al., 2018), the limiting factor in the early development of the open data ecosystem in the agriculture sector is the amount of open data and repositories. With a focus on the G2B principle, we decided to identify key elements of the open data ecosystem (Zuiderwijk et al., 2014) of the agricultural sector in Croatia and to assess their status in order to provide input to the next steps towards an open data ecosystem in the agriculture sector in Croatia.

Material and methods

The very definition of the open data ecosystem as a functional whole and cyclically sustainable system of people, practices, values and technologies, which make up the infrastructure of open data within a narrower context is still in development (Zuiderwijk et al., 2014; van Loenen et al., 2018). Therefore we decided to focus on the four key elements of the open data ecosystem according to Zuiderwijk et al. (2014): (i) open online data repositories; (ii) search, assessment, review of data and licences; (iii) cleaning, analysis, enrichment and visualisation of data; (iv) interpretation, discussion and feedback to the data provider and other stakeholders.

From the G2B perspective, the data sets of the agricultural sector were analysed, in a narrow sense and without the analysis of policy acts, heralds and proceedings, available through the following repositories: (i) the Open Data Portal of the Republic of Croatia (https://data.gov.hr/) in the thematic field of Agriculture, forestry and fisheries; (ii) the website of the Ministry of Agriculture (https://poljoprivreda.gov.hr/); (iii) the website of the Croatian Agriculture and Food Agency (https://www.hapih.hr/), and (iv) the website of the Paying Agency for Agriculture, Fisheries and Rural Development (https://www.apprrr.hr/), in relation to the number of documents available and the publishers of datasets.

The search, assessment and review element of data and licenses includes: (i) an analysis of access options; (ii) information on the data format (raw or aggregated; and the format and size of the files). The third element is the purification of raw data, analysis, enrichment, visualization and connectivity. For the fourth key element of the ecosystem, the ability to provide feedback to a publisher that publishes a particular set of data has been analysed.

Results and discussion

A total of 357 files (5.95 GB) available from four publishers were analysed, and over ten databases and related applications from these publishers were listed. On its website, the largest number of files are provided by the Ministry of Agriculture of the Republic of Croatia (202 analysed), while the Payment Agency offers the largest amount of data open trough downloadable files on the account of spatial data (5.1 GB). In addition to the Ministry, the Croatian Agency for Agriculture and Food, and the Agency for Payments in Agriculture, Fisheries and Rural Development make central public publishers of open data in the agricultural ecosystem in Croatia. According to the number of open files, the most common format is .pdf (not considered to be an open format), followed by .xslx and .aspx (Graph 1). Almost no dataset was published in an open format (.csv). Additional data sets are opened by the National Bureau of Statistics, as the only publisher of agricultural data that also references its data through the Open Data Portal of the Republic of Croatia.

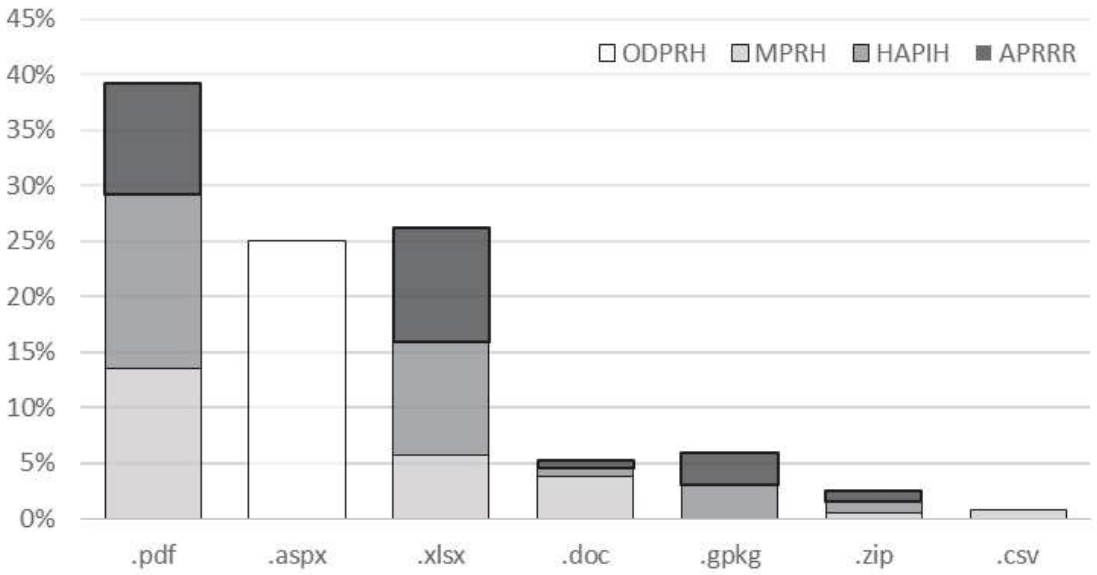
In addition to the publication of open documents, the Ministry of Agriculture of the Republic of Croatia (see https://poljoprivreda.gov.hr/) also creates databases for different target groups (for consumers, or narrower professional communities). These are Food bank, Hagris, SVIS, Lysacan, VetIS, Owner, Honey of the Croatian apiaries, Eggs of the Croatian farms, KOLK-VETI and By-products. They are available through the eponymous or other (e.g. e-donations, eDPZ) web or installation applications of the Ministry. Access to databases is possible with registration, but there is no information on how users are registered. The databases Honey of the Croatian apiaries, and Eggs of the Croatian farms offer limited search options for users that did not register. Feedback is enabled through the user system by contacting the person in charge of each application for both registered users and the ones not registered.

Access to open data sets of the Ministry was analysed based on 202 data sets available through documents in .xlsx, .csv, .doc, .html, .png, .zip, and .pdf formats (0.33 GB) (Graph 1). Predominantly .pdf files, which are not considered to be an open format, that contain aggregated data with certain map visualisations are shared (Graph 1). Animal information is available in the form of data aggregated by cities, without the specification of the animal species. The farm registry is listed, but there is no access to that data. The assessment of breeding values in Croatia is only available for cattle in .html form, and milk control and management of cattle are available through the Owner web application. Breeding associations or similar address books have different levels of availability. Some are open in .html form and others accessible only through applications with registered users. For some address books, visualization is also enabled on the interactive map. Gene bank collections offer aggregated information on the number of samples of reproductive tissues or genomic collections, without the possibility of searching. Individual data sets are available in multiple formats, e.g. .csv, .xlsx and .pdf. For some, the download format information is wrong. Feedback is possible for individual thematic data groups, while information on how to use the data or licenses are generally not available.

The Croatian Agriculture and Food Agency offers the possibility to access information on demand using defined application forms for: access to data; re-use of the data, and supplement or clean the data. Some information from this publisher is freely available on the website https://www.hapih.hr/. 36 documents (0.13 GB) were analysed with data related to seed and planting, plant protection, soil, food safety, livestock production, quality control of livestock products, viticulture, fruit growing and vegetable growing that this publisher opens through its website. Predominantly open .pdf files contain mostly aggregated data with

individual map views (Graph 1). In addition to the files freely available on the publisher's website, an additional 14 sets of data were analysed in .xlsx documents (0.17 GB) consisting of raw data related to livestock production provided on demand for scientific and educational purposes. Files that are freely available are not linked, there is no information about licenses, are predominantly of simple structures, while the raw data provided can be purified for different purposes. Visualizations are limited to aggregated data in annual reports.

For the publisher Paying Agency for Agriculture, Fisheries and Rural Development (APRRR) (https://www.apprrr.hr/) 75 files out of a total of 5.32 GB were analysed. Most data can be accessed without registration, but some data is available through the Arkod portal, Agronet and Portal ponuda, which require registration. The Arkod portal allows visualization on an interactive map, while Agronet and Portal ponuda allow for data browsing. Terms of use of Arkod spatial information are available on request. The ability to feedback on data is not directly obvious.



Graph 1. Shares of used formats of the number of analyzed files published by central public publishers of open agriculture data in Croatia.
ODPRH – Open dana portal (agriculture); MPRH – Ministry of agriculture; HAPIH – Croatian Agency for Agriculture and Food (Internet documents and documents on demand); APRRR – Paying agency.

The prevailing file types available from this publisher are .pdf and .xlsx, fewer sets are .gpkg files, while exceptional data is present in .doc files, and several archived sets of information are available in compressed .zip files (Table 1). The file format does not provide related data, the license information is generally not present, the data is predominantly aggregated and of simple structure.

Through the Open Data Portal of the Republic of Croatia in the thematic area Agriculture, forestry and fisheries, 30 sets of data were analysed, of which 29 were published by the State Bureau of Statistics and one by APRRR. The National Bureau of Statistics issues aggregated sets of .aspx format data with metadata, and all (sub)selected data can be downloaded in .htm, .json, and various .csv, .xls format options. Each set of data has the specified source, responsible service, contacts of responsible persons and copyright. The APRRR data set lists aggregated farm register statistics and the register published in 2015, but the link is not valid.

Although complementary owners of public data in the agricultural sector from a G2B perspective, such as peripheral public services (cities, counties, municipalities), centres of excellence (Varga et al, in press) and the research and scientific community in this survey are not covered, four data sets sources, i.e. the four data publishers researched, are currently responsible for the fundament of the open agricultural sector ecosystem in Croatia. Step-ups in opening potential are databases, applications and interactive visualizations of the Ministry of Agriculture and APRRR. The additional potential of data in food production in Croatia is carried by non-governmental organisations (LAGs or breeders' associations, etc.) from a B2G perspective. The large representation of .pdf file formats with aggregated data does not currently allow for database and G2G ("government to government") interoperability and efficiency in data usage.

For the final assessment of this ecosystem, in addition to the estimated key elements, an additional three elements should be analysed: user trajectories clarifying the routes of use of open data sets (Walker et al., 2020); the data quality management system; and different types of metadata that can be linked to ecosystem elements (Zuiderwijk et al., 2014). The first step in further evaluation of the system is a stakeholder analyses (Hrustek et al., in press) and the social elements of the open data ecosystem, as well as infrastructure and technological elements and tools for working with open data in this sector. Additional value provides a link to the data space of the agricultural sector, namely the availability of socio-economic, legislative and administrative, environmental and natural resources data with application in the agricultural sector (Šalamon, 2019).

Conclusions

Although there are steps in the sector to open up the potential of open data in the form of databases, applications and interactive visualisations, central agricultural sector publishers can improve the amount and quality of published open data significantly. The lack of raw data hinders faster access to the consolidated versions of the data, and the opening of predominantly aggregated data prevents the faster development of new knowledge and conclusions from the large amount of information aggregated by central public authorities in the sector. Predominantly used formats are not open and unsuitable for data connectivity, and the lack of licensing information restricts the proper use of open data. Education, data literacy and systematic investment of human capacity and knowledge within the agricultural sector are crucial in enabling the publication of open data and the reuse of such data from central sources, but also for incentifying peripheral publishers.

Note

This work was created as a product of the Twinning Open Data Operational (TODO) project funded by the EU Horizon 2020 Research and Innovation Programme (GA: 857592 – TODO).

References

Davies T. (2012). Open Data: Infrastructures and ecosystems. ACM Web Science Conference 2012: 1-6.

EC (2020). A European strategy for data. European Commision, Brussels.

Gens F., Del Prete C., Lundstrom S., Fearnley B., Nakamura T., Turner V., Carter P., Carvalho L., Fitzgerald S., Gillen A., Matsumoto S., Mohan D., Newmark E., Ng S., Parker R., Schubmehl D., Vesset D., Tom Mainelli T., Rick Villars R., Sutherland H. (2017). Worldwide IT Industry 2018 Predictions. IDC FutureScape, Doc # US43171317.

Hrustek L., Tomičić Furjan M., Džidić A., Šalamon D., Varga F., van Loenen B. (2021). Key stakeholder groups for digital innovation of agriculture in Croatia. Proceedings of 55th Croatian & 15th International Symposium on Agriculture (in press).

Musa A., Bevandić D., Herak D., Jadrijević L., Kovačić M., Lukša Z., Vrček N. (2018). Priručnik o ponovnoj uporabi informacija za tijela javne vlasti: otvoreni podaci za sve. Musa A. (ed.), 59. Zagreb, Hrvatska: Povjerenik za informiranje, Hrvatska.

Panetta C. (2016). Gartner's Top 10 Strategic Technology Trends for 2017. Smarter with Gartner. Gartner.

Šalamon D. (2019). Overview of open data in Croatia available for use in the sector of agriculture. Book of Abstracts of the International Conference Innovations: Guarantee for Future of Agribusiness in Croatia. Svržnjak K. (ur.). Zagreb: Hrvatsko agroekonomsko društvo, 31-32.

Šalamon D., Džidić A., Varga F., Čavrak I., Bosnić I. (2021). Data literacy in sector of Croatia-Learning about and with open data. Book of Abstracts of the International Conference Innovations: Inovation in education. Hrvatsko agroekonomsko društvo (in press).

Varga F., Grdiša M., Nikolić T., Guberović E., Bosnić I., Welle Donker F., Šalamon D. (2021). Dalmatian pyrethrum spatial records quality from Croatian Open Databases. Proceedings of 55th Croatian & 15th International Symposium on Agriculture (in press).

van Loenen, B., Vancauwenberghe, G., Crompvoets, J., Dalla Corte, L. (2018). Open Data Exposed. In: Van Loenen B., Vancauwenberghe G., Crompvoets J. (eds.). The Hague: TMC Asser Press, 1-10 (Information Technology and Law Series; vol. 30).

Walker J., Thuermer G., Simperl E., Carr L. (2020). Smart rural: The open data gap. In *The Conference: Data for Policy 2020*, Data for Policy community platform on Zenodo.

Zuiderwijk A, Janssen M, Davis C (2014). Innovation with open data: essential elements of open data ecosystems, Information Polity 19 (1-2): 17-33.