20(2), pp.64-221, 2022 **ISSN 1334-4684**

INTERDISCIPLINARY DESCRIPTION OF COMPLEX SYSTEMS

Scientific Journal

Scientific Journal

INTERDISCIPLINARY DESCRIPTION OF COMPLEX SYSTEMS

INDECS, volume 20, issue 2, pages 64-221, year 2022
Published 29th April 2022 in Zagreb, Croatia
Released online 29th April 2022

Office

Croatian Interdisciplinary Society
c/o Faculty of Mechanical Engineering & Naval Architecture
I. Lučića 1, HR – 10 000 Zagreb, Croatia
E-mails: editor@indecs.eu (for journal), ured@idd.hr (for publisher)

Published bi-monthly by *Croatian Interdisciplinary Society* (http://idd.hr) as online (ISSN 1334-4676) and printed (ISSN 1334-4684) edition. Online edition, http://indecs.eu, contains freely available full texts of published articles. Printed by Redak d.o.o. (HR) in 50 pieces.

Journal INDECS is financially supported by Croatian Ministry of Science and Education.

Content of the journal INDECS is included in the DOAJ, EBSCO, EconLit, ERIH PLUS, Ulrich's and Web of Science Core Collection.

INDECS publishes original, peer-reviewed, scientific contributions prepared as reviews, regular articles and conference papers, brief and preliminary reports and comments to published articles. Manuscripts are automatically processed with the system Comet, see details here: http://journal.sdewes.org/indecs.

The accessibility of all URLs in the texts was checked one week before the publishing date.

TABLE OF CONTENTS

REGULAR ARTICLES

Interdisciplinary Description of Complex Systems 20(2), ii, 2022

EDITORIAL: TRENDS AND PROSPECTS OF OPENING DATA IN PROBLEM DRIVEN SOCIETIES

This special issue of the interdisciplinary **INDECS journal is** planned to offer the results of research on opening and sharing data in the context of the European Data Strategy striving towards a single European data market, the implementation of the Open Data Directive[1] by EU Member States and other EU measures promoting open data being expected soon.

In the last two decades open data has revolutionised the data landscape in the European Union. In the beginning of the 21st century government data was typically available at high cost accompanied by restrictive licences that only a few could afford. How different is the data landscape in 2022: in the EU27 thousands of government datasets are provided as open data, including many high value datasets and this may only be the beginning of the next wave of open data foreseen if the European Data Strategy is fully implemented. Europe is working towards this goal with the soon to be expected draft *Implementing act on high-value data sets* as part of the Open Data Directive, the proposed *Data Act*[2], and the proposed *Data Governance Act*[3].

Although the availability of open government data has significantly grown over the past twenty years, the provision is still mostly provider led, leaving even today many users in the dark when they are trying to find, access and reuse open government data for their purposes. As such the supply of open government data is not yet based on the major challenges we have to encounter in our information societies anno 2022. The upcoming *Interoperable Europe Act* will help the Member States further on their way to digital transformation facilitating the creation of eco-system of integrated digital public services. This Act will be adding to the building blocks upgrading governing digital services in the EU agreed upon earlier this year by the European Commission, European Parliament and European Council: the Digital Services Act[4] and the Digital Markets Act[5].

Academic research on and with open data so far has been mostly single disciplinary and single domain, providing new insights limited to these disciplines or domains, ignoring the fact that most challenges go beyond these and require interdisciplinary and multi-domain research approaches to obtain a more comprehensive understanding of the nature of the challenges and ways to overcome them.

The ongoing project TODO (Twinning Open Data Operational) that has received funding from the European Union's Horizon 2020 research and innovation programme (grant agreement No 857592) aims to leverage the interdisciplinary scientific excellence and innovation capacity of the University of Zagreb in the field of open data and multi-domain research approach on the open data life cycle to boost the supply and the use of open government data in Croatia and beyond. All of the planned activities of the project, especially the capacity building and the transfer of the knowledge and the experience are fostered by the expert research groups from the two universities with pronounced excellence in open data research

1 Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information. OJ L 172.

2 European Commission, 2022, Proposal for a of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act). COM(2022) 68 final.

3 European Commission, 2020, Proposal for a of the European Parliament and of the Council on European data governance (Data Governance Act). COM(2020) 767 final.

4 See for the text of the proposal: Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC. COM/2020/825 final.

5 See for the text of the proposal: Proposal for a Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act), 2020/0374 (COD).

*DOI: 10.7906/indecs.20.2.e.1

and ecosystem development: the Delft University of Technology and the University of Aegean.

The decision to organise a special issue of research focusing on the prospects of open data was a logical step in building the multi-disciplinary teams. The TODO consortium researchers were introduced to a number of domain focused open data research topics at the National Open Data Conference 2021 (NODC2021), September 20th to September 22nd, 2021 in Zagreb. Most of the articles of this special issue are addressing the Croatian open data ecosystem and are a development of the initial domain/discipline approach. Several articles present a breakthrough in connecting the disciplines to address the real world challenges by exploiting the open data and by improving the open data ecosystem.

This special issue entitled "Trends and Prospects of Opening Data in Problem Driven Societies" identifies several of societal challenges and explains the role of open data in making our world a little better. The trends we draw from the submission to this special issue are:

Trend 1: Extending the open data supply;
Trend 2: Applying open data to real world challenges;
Trend 3: Improving the open data ecosystem.

The first trend (Extending the open data supply) is exemplified by the following articles:

1. *Improving the availability of space research spatial data* where the authors Nevistić and Bačić provide an overview of planetary spatial data archives, data storage and retrieval methods, and their shortcomings in the context of easy search, download and interpretation of data, with the aim of establishing Spatial Data Infrastructure of Celestial Bodies that would make space data better accessible to the public and non-planetary scientists.
2. *Open National CORS data ecosystems: A cross-jurisdictional comparison* by Supinajaroen *et al.* explores the divergence in the openness of the National Continuously Operating Reference Stations (NCORS) in the Netherlands, Germany and Sweden. NCORS networks collect and process data from the Global Navigation Satellite Systems (GNSS) to provide precise positioning data to support spatially related activities.
3. In *Identifying and overcoming the barriers towards open data of public undertakings,* Boone and Van Loenen consider the third wave of open data where open government data is complemented by open data of the public undertakings. They assess expected legal, organisational and technical barriers for the case studies of the Port of Rotterdam and Schiphol Airport in Amsterdam.
4. *Firefly occurrences in Croatia – One step closer from citizen science to open data* by Virić *et al.* deals with data deficient nature protection issues and species, which can be addressed by applying the citizen science approach as well as with the value of data collected by non-experts.

The second trend discovered in the submissions for this special issue, applying open data to real world challenges is evident in the following articles:

1. *Urban dog spaces: the openness of dog-related data in the City of Zagreb, Croatia* by Varga *et al.* is an assessment of data provided via official websites and portals of the city required for construction and maintenance of urban infrastructure. Five-star system of ranking the data formats was used for the published data and the quality of data was cross-checked with the field survey and citizen science collected data.
2. *Importance of the open data approach for multimodal travel improvement* by Mandzuka *et al.* examines *Multimodal Journey Planners* (MJPs). MJPs provide travellers with better and more complete information when choosing a mode of transport so they

can select the most suitable option for their needs. The *open data* approach is crucial for defining a system that responds to the end-users' actual needs and aspirations. In this research, the importance of traffic data collection, acquisition and distribution according to the open data concept is described.

2. *Open election data: Evidence from Croatia in a comparative perspective* by Đurman *et al*. compares the seven major groups of electoral data available for the electoral process in EU27 and the United Kingdom, focusing on the temporal aspect of the timeliness of pre- during and post- election process data as well as providing additional details on the open electoral data available in Croatia.
3. *Open access on GNSS permanent networks data in case of disaster* by Latinčić *et al.* pointed out that although open access to Global Navigation Satellite Systems (GNSS) permanent networks is highly beneficial for natural disaster management, access is currently often restricted. A high percentage of GNSS permanent network providers that participated in the research presented agreed that these data should be freely available in instances of natural disasters.

The third identified trend: Improving the open data ecosystem, deals with the research not focused on the data sets itself, but rather in the components of the ecosystem required for its effective re-use and value generation:

1. In the article *Framework for federated learning open models in e-Government applications*, Guberović *et al*. develop a concept of the Federated Learning Open Model (FLOM) as an example for the third generation e-Government machine learning tool in the cohabitation of ethical computing and intelligent services. The authors apply the proposed FLOM framework to the horizontally partitioned data environment with the example of the agricultural commodity price prediction, as well as the vertically partitioned data environment on the example of the loan approval prediction.
2. In the *Serious games for building data capacity*, Di Staso *et al.* recognized the need for the fast awareness raising and the capacity building of the public institution employees and provide an overview and the assessment of twelve available teaching games covering that potential.
3. *Towards digital innovation: Stakeholder interactions in agricultural data ecosystem in Croatia* is an article in which Hrustek *et al*. analyse the requirements and the potential for data flow in Croatian agriculture data ecosystem, focusing on data supply from this data rich sector. In complex systems such as agriculture is, effective cooperation in promoting of the best management practices and sustainable value creating depends on understanding the myriad of stakeholders operating often in a decentralized data ecosystem. Identifying the stakeholders and their relationships is achieved by superimposing the stakeholder importance with respect to the estimated data supply based on the on-line queries and semi-structured interviews.

Basically, the most important challenge in the research community today is enabling the multi- and inter-disciplinary collaboration. The speed of the societal challenge of data empowered development and sustainability achievement in EU27 depends on the three trends identified in the submissions for the "Trends and Prospects of Opening Data in Problem Driven Societies". We can consider the dedicated issues on open data research as the important step in enabling multi-domain and interdisciplinary approach in applying open data to real world challenges through the maturation of open data ecosystems.

Zagreb, 28th April 2022

Guest editors:
Bastiaan van Loenen
Dragica Šalamon

Interdisciplinary Description of Complex Systems 20(2), 64-77, 2022

IMPROVING THE AVAILABILITY OF SPACE RESEARCH SPATIAL DATA*

Zvonimir Nevistić and Željko Bačić**

University of Zagreb, Faculty of Geodesy
Zagreb, Croatia

DOI: 10.7906/indecs.20.2.1
Regular article

Received: 14 January 2022.
Accepted: 3 March 2022.

ABSTRACT

The rapid development of space technology and the increased interest in space exploration have resulted in the intensive observation of celestial bodies, mostly in the solar system, over the past decade with the prospect of an upward trend in the future. Large amounts of collected data on space bodies impose the need to develop the Spatial Data Infrastructure of Celestial Bodies at the general level to enable standardized organization and storage of these data, and their efficient use and exchange. To approach the development of such an infrastructure, it is necessary to investigate what data, as well as how and to what extent, are collected through space observation. It is also necessary to investigate how this data can be obtained. This paper provides an overview of planetary spatial data archives, data storage and retrieval methods, and their shortcomings in the context of easy search, download and interpretation of data, all with the aim of establishing Spatial Data Infrastructure of Celestial Bodies that would make space data more accessible to the public and non-planetary scientists.

KEY WORDS

planetary data, spatial data infrastructure of celestial bodies, SDICB, space data archives

CLASSIFICATION

JEL: C89

*This is the extended version of the abstract published in: Vujić, M. and Šalamon, D., eds.: Book of abstracts of the National Open Data Conference. University of Zagreb, Faculty of Traffic and Transport Sciences, Zagreb, 2021.

**Corresponding author, η: znevistic@geof.hr, + 385 1 4639 538
University of Zagreb, Faculty of Geodesy, Kačićeva 26, 10 000 Zagreb, Croatia

INTRODUCTION

Today, with the advancement of technology and the widespread use of the Internet a lot of information is easily accessible to everyone with Internet access, and spatial information is one of the most important elements to support decision-making in many disciplines. Organizations around the world spend millions of dollars each year on the production and use of spatial data [1], and yet there is often a problem of lack of information within organizations to perform certain tasks. To address this issue, organizations often use data from other sources and share data with each other. The large amount of spatial data that appeared in last 15 years does not facilitate their use. On the one hand, it is challenging to find and access spatial databases that are distributed through various portals of government agencies and other web portals [2]. On the other hand, there is a great redundancy of data where money and human resources are spent to collect and maintain duplicate data [3]. This has triggered the development of the Spatial Data Infrastructure (SDI) concept that solves the problem of finding spatial data and reducing their redundancy [4] and enables better data management which can achieve economic and environmental benefits. The SDI concept connects existing spatial data into a single network, and to be successfully implemented, it is necessary to harmonize and standardize existing data sets [5]. The application of this concept provides a basis for searching spatial data, their assessment and application at all social levels and facilitates integration with other data sets.

Spatial data are increasingly being collected through space observations. Technological development has enabled numerous scientific studies of planets and other celestial bodies, and today spacecrafts visit more and more planets, satellites, comets, and asteroids. Data collected by space research are of particular interest because their interpretation provides a better understanding of the Earth and its dynamics and provides answers to important questions, such as the impact of global warming [6], provides a better understanding of the solar system and is used to mitigate hazards on Earth and contribute to the development of science in general [7]. Planetary science is one of the fastest growing-scientific disciplines [8] which integrates many other scientific disciplines to determine the origin, physical processes, and other characteristics of objects in space [9]. Today, space data increasingly converges with terrestrial geo-scientific visualizations and analyzes such as GIS and web maps. The National Aeronautics and Space Administration (NASA) currently has 2 petabytes of space spatial data, and large amounts of data are archived each year by new missions. Space data stored in various archives are not suitable for immediate use, so they are accompanied by metadata to facilitate their use for the non-planetary community [10]. These archives have changed significantly over the past few years. Their primary purpose was to provide data storage for planetary scientists and their research. With the increase in the number of multidisciplinary missions, the number of other scientists and public who want to access this data has increased. The collection and distribution of space data face numerous challenges. One of the challenges is the standardization of data storage methods which is crucial for accurate and precise analysis and scientific research. Today, this problem is of great interest, given that access to data is available to everyone, but there are still no adequate ways of storing data or their distribution and search of data through archives is often limited [11, 12]. To solve these problems, it is necessary to develop standards that would allow interoperability and data exchange between different communities and to create archives that will satisfy all users' needs and in which will be easy to find, share and interpret data. One solution is development of Spatial Data Infrastructure of Celestial Bodies (SDICB) as an extended concept of traditional (terrestrial) SDI, which should address the challenges of collecting, managing, finding, and using planetary data. Such a concept would support research missions in the space community and maximize the value of spatial data collected on planets [13] and other celestial bodies. The SDICB must serve the wider community whose members do not have to be spatial and space data experts and who do not understand all aspects of spatial data storage, retrieval, and use. Currently available solutions are often technology-focused and should instead focus on simplifying data access and improving data usability.

This paper provides an overview of planetary spatial data archives, data storage and retrieval methods, and their shortcomings in the context of easy search, download and interpretation of data, all with the aim of establishing SDICB that would make space data more accessible to the public and non-planetary scientists.

The first part of the paper explains planetary data and their importance and where this data can be found. The most frequently used archives and portals for access to space data are explained, and the ways of access to data as well as the main shortcomings of archives are investigated. Numerous interoperable initiatives have been launched to solve the existing problems of current ways of distributing space data. The mentioned initiatives serve as a basis for the development of the SDICB, the concept which is explained together with the future work of this research.

PLANETARY DATA

Since ancient times, mankind has always explored the stars and planets and used their observations for various applications by which they revolutionized perception and understanding of the world. With the advancement of technology, planetary science has greatly advanced, and planets have become objects of scientific research using in-situ and remote sensing methods. Our knowledge of the solar system is increasing with advances in technology and spacecraft are visiting more and more planets, satellites, comets, and asteroids. Planetary science is the science of studying planets, moons, and other space bodies in the solar system and space [8]. Many other scientific disciplines can be applied in planetary science such as geology, geography and GIS, geodesy and remote sensing, geophysics, mineralogy, volcanology, geomorphology, and others. All these disciplines are applied to determine an accurate assessment of the shape, origin, and other processes on space bodies. Scientists are working to improve our understanding of planets and other space bodies through the study of the atmosphere, surface, and their interior. The main goal of space missions is to try to understand the origin of planets and their physical processes and systems, as well as the characteristics of asteroids and other objects in space [9].

Nowadays, planetary data increasingly converges with terrestrial geo-scientific visualizations and analyzes tools. To present this data as faithfully as possible, a spatial component is needed, and today more and more space spatial data are being collected. Space spatial data are all data with a spatial component, obtained by remote sensing methods, navigation methods, georeferencing of images or collected in-situ by rovers with enough spatial information to be projected on the space body. The space spatial data stored in various archives are not easy for use by non-professional scientists, so the appropriate metadata is provided, along with different types of image data to facilitate their use for the non-planetary community [10].

The conversion of raw (unprocessed) space data into spatially enabled data involves great efforts in time and knowledge. For this to be achievable, missions must be developed and calibrated appropriately, and software and platforms for distribution of this data require continuous development and maintenance [14]. Coordinates on Earth are easily obtained from Global Navigation Satellite Systems (GNSS), system of satellites that provide autonomous geo-spatial positioning with global coverage and allow small electronic receivers to determine location on Earth, or similar positioning systems. Such sophisticated systems do not exist for planetary research. Therefore, to obtain spatial data, reference systems from astronomical observations and remote sensing of the planet`s surface (e.g., radar measurements, laser altimetry) must be realized. Cartographic maps from which the coordinates of characteristics on the planets can be easily extracted can be produced by various techniques. The most common technique is photogrammetry, and today maps are often calculated relying on the known position of grounded space missions and the known, previously derived, planetary rotation parameters.

The basic division of space spatial research data is into fundamental data, which include geodetic control networks, topographic data, and ortho-images, while the second group is framework data which include composite maps, planetary nomenclature, and geological maps. All these data have relevant information about the position in the reference coordinate system of the planet or body they are referring to. In the space archives we can find many other data, for example altimetric data for each space body, data on plasma and atmosphere of planets, orbital data, radar data, geometric shapes of planetary models and their parts, spectrometric data, etc. The most common types of data are geological data that can be found in all archives for each planet and other bodies. Geodetic and geophysical data include data on reference systems, planetary rotation and shape, topography, and gravitational models. It is also important that each data has corresponding metadata because without information on how they were taken, with which instrument and for what purpose, they would not be usable. The formats in which data is most often stored are VICAR (Video image access and retrieval) format, PDS (Planetary Data System), FITS (Flexible Image Transport System) and Raw formats such as JPEG.

For planetary data cartography, GIS, and remote sensing have an important role. Mapping of space bodies began with the invention of telescope, and with the advancement of technology, today the most important role in the mapping of space bodies has launch space missions that carry high-resolution cameras for data collection and based on this data cartographic maps are created. The biggest difference in the use of space to Earth data is the shape and size of the body being observed. International Astronomical Union (IAU) defines for all major bodies in our solar system geodetic parameters, which enables their mapping, and which are the basis for the data interoperability [12]. Maps of each planet are thematic maps, and the most common are topographic and geological maps. Nomenclature and geographical names also play an important role in data collection, and the IAU is the competent authority for planetary and satellite nomenclature [15] used to uniquely name features on the surface of space bodies so that the features can be easily found and described. Consistent and accurate nomenclature is crucial for efficient data exploitation. In addition to cartography, remote sensing is one of the most important sources of data in solar system research since it is not physically possible to visit most of the bodies. Large amounts of data are collected for each body, and GIS tools serve for conducting spatial analyzes over that collected data. Shape, topography, gravitational field, and rotation parameters are the basic characteristics of every space bodies. The part of geodesy that deals with the study of these characteristics and their influences on geophysical processes is planetary geodesy. The objectives of planetary geodesy are to establish a network of planetary geodetic control points, define reference systems, determine planetary rotation parameters, establish global and local planetary surface models, and create geographic image data, maps, and information systems [16]. Remote sensing is the most used method of data collection for planetary geodesy (radar observation, laser instruments or altimeters), but other methods, such as observing the orbital motions of satellites are also used [17]. All these methods and parameters play an important role in space science and without them it would not be possible to obtain spatially enabled data.

The innovative nature of space missions creates new technologies and study techniques that will later benefit the public on Earth and contribute to the education of future generations. Space research also contributes to the public in the form of a growing understanding of the solar system and the importance of our planet in it [8]. Space data archives have drastically changed over the years. From the storage of data for space scientist and their research, today with the increasing number of missions, archives have become of big interest for public and other (non-space) scientists.

Collection and distribution of space data face several challenges. One of them is the standardization of cartographic methods and data which is critical for accurate and precise analysis. Today, this problem is of great interest since access to data is easy and accessible to everyone. But there are still no adequate ways to store this data as well as distribute it to users. One of the main problems is extremely decentralization of space community where each organization has its own archives and data sources, but also the standards and formats they use. The solution to these problems is the development of standards that would enable the interoperability of these data between different communities and that would create the foundations for the establishment of a SDICB. Therefore, it is necessary to create spatial data that will satisfy users and that will be easy to find, share and interpret.

SPACE RESEARCH DATA ARCHIVES AND PORTALS

Space research data is archived and made available to users through data archives and portals. Access and download to all of the data is free of charge to everyone. New missions increase the number of data every year, which does not make it easier to search and access the data.

The largest archive of space research data is NASA`s Planetary Data System (PDS), https://pds.nasa.gov. PDS provides access to data and related documentation through 9 nodes. Six nodes are science discipline nodes, focusing on atmospheres, geosciences, cartography and imaging sciences, planetary plasma interactions, ring-moon systems, and small bodies. There are two support nodes: Engineering node, and the Navigation and Ancillary Information Node. The ninth node is the project management group. The primary purpose of this archive was long-term preservation of data for the space scientist`s research, so, data retrieval methods are focused on the data itself and not on the user needs. As the number of users of archived data increased and expanded to the non-planetary community over the past few years, additional data search capabilities based on various attributes are needed. These capabilities depend on the metadata quality of archived products. Although this archive is not adapted for the non-planetary community, it is the most used source of space research data today [11]. To find and download the data in the PDS, it is necessary to know the mission and the instrument of the mission that collected the data, which makes it difficult to search. Each dataset in the archive can be accessed through multiple nodes which emphasizes the redundancy of existing solutions of accessing the data. All products in the PDS are peer-reviewed, well-documented and available free of charge to scientists and the public. The data is stored in the PDS3 / 4 standard developed by NASA, and only some data can be viewed before download.

Following the PDS, the European Space Agency has created the Planetary Science Archive (PSA). PSA is an archive of data from ESA`s space missions but also distributes data from some of NASA`s missions. PSA supports and uses NASA`s PDS3 / 4 standard for archiving the data, and data is distributed through a web data browser which is based on filtering data options by type, mission, and instrument. In addition to NASA, United States Geological Survey (USGS) has the largest database of space data. USGS has developed several web searches tools. Integrated Software for Imaging and Spectrometers (ISIS) is one of USGS`s portals which consists of software for downloading, processing, and calibrating radiometric and geometric data of space research. It supports the missions of NASA, ESA, ISRO and JAXA. USGS independently creates documentation and metadata for distributed products. Map-a-Planet (Astropedia) is also under the jurisdiction of the USGS and provides to users projected images and ready-to-use products. It supports standardized WMS web mapping services for all bodies, and data search is enabled through the map view. All data can be visualized before download. On this portal it is possible to download GIS-ready products for use in standard GIS tools. According to many research and surveys, Astropedia is the most user-friendly portal and the easiest to use, especially for non-planetary users. Imaging Node Annex provides users with spatial data and products derived from PDS data (mosaics, maps, SHP files, databases). For each product the link to the original publications and information about the quality of the data is provided. Portal creates own metadata for products that are provided to users. Astrogeology USGS is a Web portal that offers users to download WMS services for more than 30 different space bodies with over 100 images. Access to WMS services of this portal is possible directly through GIS applications ArcMap, QGIS and GDAL.

These are just some of the most used portals and tools for downloading and retrieval of space research data. Numerous space missions are developing their own portals as well as national space agencies. There are also several open tools to facilitate the search of PDS and PSA, specialized portals and software for image data, and tools for data conversion or product generation from raw data formats.

But despite the fact that there are many places where data can be searched and downloaded, users, especially non-planetary users, have a general problem when searching for the data: “Where to download data and how to search for it?”. Borden and Bishop [11] in their research explored which archives and portals space scientists use the most for their research and what are their experiences using these interfaces. Once the data is archived, it becomes publicly available to users for future research. However, it is often the case that archived data is unused, so studies and initiatives are being launched to provide users with the best ways to facilitate access to data and use of data for all future potential users, under which are considered not only space but also scientists of other interests, and the public. In their research, Borden and Bishop [11] concluded that more than half of the respondents use NASA`s PDS portal to search for data or one of the search tools within the PDS science nodes, and the Geoscience node is the most used. 80 % of respondents download data, 10 % process it online using different tools before downloading it, while the rest of respondents only view the data online. More than a quarter of the respondents pointed out that the data after the download were not usable, or that the data are available in unknown formats that could not be used for further research. Some portals offer to user transformation of data before downloading to standard formats of desktop applications, and respondents believe that this example should be followed by all services. 70 % of the respondents pointed out the problem with the scale and resolution of the downloaded data, and 51 % of them think that the data were in incomprehensible coordinate systems [11].

SHORTCOMINGS OF THE ARCHIVES

The survey by Borden and Bishop [11] conducted among space scientists who frequently use archives points to many shortcomings in accessing, retrieving, and downloading space data. And what about other, non-planetary scientists and the public, what are their possibilities of finding planetary data? Space research is interdisciplinary research and not only space but also scientists of other professions are interested in it. Space exploration can give us a better understanding of our planet, so the data collected by space missions should be easily, efficiently, and understandably available to everyone. By analyzing the space data archives and portals with their services, 15 shortcomings have been identified from the perspective of non-planetary users and listed here:

- dispersion of data because of existence of several access points (different data archives, within the same archive, many websites, and portals of individual missions),
- data is inconsistent and incomprehensible to non-planetary scientists because metadata is not created to make it easier to understand,
- data is stored in formats that cannot be used in standard processing tools and converting them into standard formats creates a risk of losing or reducing the quality of data,
- related to data formats, there is a problem with their interoperability, because archives use formats that cannot be used in standard analysis and processing tools (e.g. GIS tools),
- without detailed knowledge of the mission and the instruments on the mission that the data are observed, it is not possible to find data of interest. The data can only be searched with prior knowledge of which search tool to use,
- archives and services are data-oriented and not user-oriented, the user interface is outdated, data search is not intuitive, and the emphasis is on long-term data storage rather than user needs. Improving archives to become more user-friendly would simplify data retrieval,
- most data cannot be visualized before downloading,
- searches are based on knowledge of data facts, such as knowing the mission and the instrument by which the data was collected,
- no ability to search the entire archive but parts according to the specified criteria. Data retrieval services must be intuitive,
- data filtering options give too much data with the same and similar descriptions and labels, which leads to the problem of choosing usable data for certain purposes,
- supporting documentation is very difficult to find for some data and is often incomprehensible to non-planetary users,
- instructions for using the archive are poorly distributed and abbreviated,
- too much time is required to search and download data of interest,
- there is no requirement that the data must be spatially located on the object which would allow more correct use of data and comparison with other data,
- archives use outdated standards, especially in terms of coordinate systems and frameworks.

Given the growth and large investments in space missions and technologies, these problems must be solved as soon as possible, and as one of the solutions to these problems is the establishment of SDICB.

SPACE INTEROPERABILITY INITIATIVES

The motivation to support common, interoperable data formats and standards in space research is not only to improve access to data and products but also to address the problem of distributing increasing amounts of data. Use of standards increases the reach of data, enables better visualization and analysis of data, and increases their efficiency [12]. Increase of space data has led to various initiatives in the space community that seek to increase the interoperability of data and provide easier access to data for all user groups.

Basic for spatial analysis are geodetic coordinates, and coordinate systems and frameworks which define the precise position of an object in relation to the agreed origin [18]. It is important that coordinate systems are understandable to all data users, otherwise the possibility of spatial errors increases significantly. Working Group on Cartographic Coordinates and Rotational Elements (WGCCRE) of International Astronomical Union (IAU) defines geodetic reference frameworks for all major bodies of the solar system. This includes the definition of the north, the prime meridian and the equator, and the definition of the geometric body which approximates the shape of the body and the definition of the vertical date. By adopting IAU recommendations, it is possible to exchange coordinates (spatial information) and understand spatial relationships using the same system. WGCCRE publishes every three years recommendations on coordinate systems and related parameters, which can be used to produce cartographic products of solar system bodies. The recommendations facilitate the use and comparison of multiple data sets and promote the standardized use of a set of parameters for mapping the solar system. Hare et al. [19] proposed a method to support the coordinate systems of solar system bodies within OGC standards. Within this standard, it is necessary to define a minimum set of information so that the user understands not only the data layer but also the current coordinate system and / or map projection. The EPSG (European Petroleum Survey Group) code is commonly used to define coordinate systems. If the coordinate system is not part of the EPSG database and if it is not presented with the minimum parameters for their definition, there is no simple mechanism for its explicit definition. To help address these issues within the space community, set of codes outside the EPSG has been proposed that the OGC would adopt as a standard. The proposed codes would include the definition of cartographic coordinates and rotational elements according to IAU recommendations and the numerical code of the space body defined according to the NAIF. Although these codes exist for several years, their use is still not recommended [12].

The USGS is a major provider of tools for processing and analyzing maps and other spatial data of NASA missions. The tools support Integrated Software for Imagers and Spectrometers (ISIS) and a specialized software package for image processing and processing of other space research data [20]. Although the software can load and export different formats, processing is only possible in the specialized ISIS3.cub format. Since 2007, the Geospatial Data Abstraction Library (GDAL) has added support for ISIS3 formats to improve interoperability with other applications. Current support is focused on the use of ISIS products and does not allow the development of new data. In the future, it is expected to add the possibility of developing ISIS3 data within the GDAL library [12].

Most of the data in the archives are stored in the original (raw) format of the mission instrument. For such data to be used for further analysis in GIS tools, they need to be georeferenced. Unfortunately, the PDS format is not widely recognized in GIS tools. Two formats recommended for use in space missions that are commonly used are GeoTIFF and GeoJPEG2000. GeoJPEG200 was approved by the PDS, and in 2008 it was used for the first time in Mars HiRISE mission. Another format used by the astronomical community, and whose use is also recommended for space exploration, is GeoFITS. The format is compatible with PDS standards and is supported by many open-source software tools and catalogs.

For web service interoperability, the consortium that defines standards is the Open Geospatial Consortium (OGC). Several space missions support WMS / WFS standards. This allows users to search and visualize data projected on a map in JPEG or PNG format. Of the currently published WMS / WFS services, only some include support according to IAU recommended for coordinate systems. Also, the WMTS service is used, which enables faster delivery of map layers but is not as flexible as the previous two and cannot generate images at any scale. WCS / WCPS network services have also found application in space research [21].

IAU recommendations do not cover other standards relevant to digital mappings, such as attribute feature representations, symbols, color scales, and metadata. Digital maps must use standards so consistent cartographic products can be developed. Attributes and symbols for digital maps of space bodies are defined in the Digital Cartographic Standard for Geological Map Symbolization [22] developed by the USGS. Recommended cartographic symbols are mostly the same as for the Earth maps. The use of these standards facilitates the understanding and readability of geological or thematic maps because users are familiar with feature attributes and symbol styles [23].

Most existing space data portals often include minimal metadata and therefore have limited search capabilities for external users. The methods defined by the OGC CSW standard can facilitate user access, so users do not need to build new search tools. One of the main advantages of using the OGC CSW standard is the ability to index one data portal into another. Products served by such mutually indexed portals must provide references to data creators and source data portals. Today, more and more initiatives are encouraging the use of OGC standards in space research. Well-structured and stored metadata are extremely important for achieving interoperability. PDS metadata, in which most of space data is archived, are not supported for use on widespread spatial data portals. Most geospatial portals require metadata as defined by the FGDC or ISO. Methods for converting PDS to FGDC / ISO metadata standards should be possible, especially because FGDC metadata standards require only a few minor add-ons to properly support space data [25]. This conversion is of great importance and should be supported so that existing FGDC / ISO tools can be used for space data.

Nomenclature is also important for any cartographic product, providing context for research and analysis. The main institution for nomenclature is the IAU, which through the Gazette of the Planetary Nomenclature collects requests and publishes official nomenclature that is used in all missions. The official IAU nomenclature is publicly available for download in KML and WFS format.

Many initiatives to improve the availability of space data have been launched through volunteer communities and various private organizations. One of these initiatives is MPASIT, which provides recommendations for the development of a comprehensive planetary spatial data infrastructure. Europlanet initiative aims to connect the European space community and project of geological mapping of Mars, Mercury and the Moon was also launched within it. The VESPA initiative deals with the availability and distribution of space data from various scientific domains. Several initiatives address the standardization of space data archiving such as PlanetServer, OpenPlanetary initiative provides an online framework to help collaborate between different institutions in planetary mapping, and CARTO initiative is focused on web solutions for spatial visualization and data analysis of space research data.

SPATIAL DATA INFRASTRUCTURE OF CELESTIAL BODIES

One of the solutions to the problem of rapid increase in the amount of space research data is the establishment of an efficient SDI. The concept of SDI is widely applicable to any spatial data and is not limited to Earth data and allows data interoperability through policies and standards by defining mechanisms for data storage and access. In the paper of Laura et al., this concept applied to space data is called Planetary Spatial Data Infrastructure (PSDI) [13, 25]. Although the name PSDI describes well its purpose, and since this concept would include not only data collected on planets but data from all space objects (satellites of planets, comets, etc.), we suggest that the more correct name for this would be SDICB.

The SDICB is an extension of the traditional terrestrial SDI that will allow standardized collection, management, and retrieval of spatial data from space exploration. Today, space data is stored in archives and portals and such archives are not user-centric, do not allow semantic data search, and are adjusted for space scientists and research. The existing archives, given their objectives and method of implementation, currently do not meet the main principles of the SDI. SDI must serve the wider community whose members do not need to be spatial data experts and who do not understand the intricacies of storing, retrieving, and using spatial data. Archives are technology-oriented and need to focus on simplifying data access and improving data usability. The main reasons why SDICB is needed in the space community is that its establishment would keep all data in one place, avoid duplication of data from different agencies and space research teams, harmonize the formats of data collected and achieve their interoperability, simplify access and downloading. Using these datasets would reduce the number of difficult-to-understand data access tools and allow other users outside the space community to access and use the data. The establishment of SDICB would increase user confidence in the interoperability and accuracy of the data, which would contribute to scientific research and decision-making.

Current methods of archiving data do not allow their use by non-professional users who are not involved in the data collection. There are several initiatives within NASA and other space agencies focused on increasing the availability of space research data to the public. Most space data users are unable to process and use raw data without investing much time and effort in understanding archived data. When accessing space data, there is a big problem of decentralization of the data, given that there are many places where data can be accessed and there is a big problem of data duplication. Insufficiently clear metadata, which does not adequately describe the data and does not provide good methods for searching and filtering data of interest, is also a big problem. SDICB would centralize data by establishing a single access point for spatial data of space exploration missions. Adopting recommendations from various initiatives to improve data availability and interoperability through a unique space research metadata catalog as shown in Figure 1 would provide standardized ways of accessing data for all users, make metadata understandable to all, and would allow flexible data access technologies.

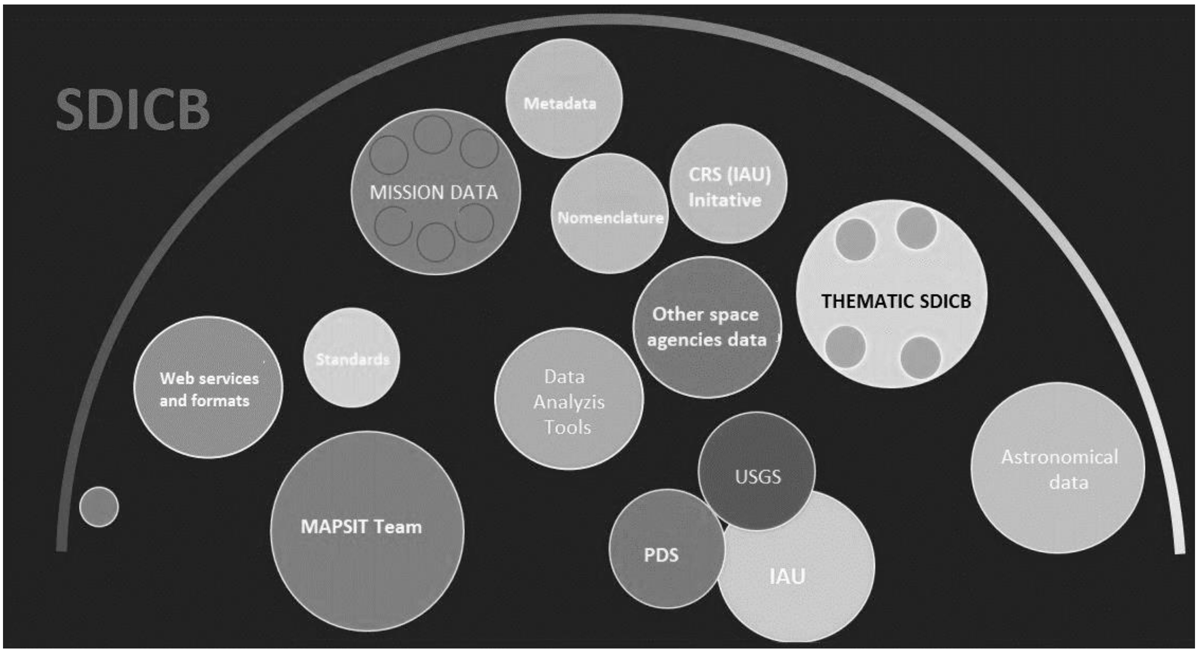


Figure 1. Data, recommendations, and initiatives included in the development of SDICB.

Given the great challenges and high costs associated with collecting space data, it is not surprising that the development of SDICB was not the primary problem of the space community. However, by treating and preparing the collected data as multi-purpose infrastructure products [26], this data will not remain unused, i.e., the currently inefficient management of this valuable data set will be avoided. As terrestrial SDI, SDICB would be a set of agreed standards, institutional cooperation agreements and policies to describe the framework within spatial data from space exploration will be collected and organized so that they are easily accessible and usable for a wide range of user.

As with terrestrial implementations, the SDICB concept would have 5 basic components: policy, access networks, standards, people, and data. But each of these components must be further expanded to meet space data requirements. The extension refers to answering various questions related to data and other components, and the answers to these questions provide the current state of space spatial data management and identify subjects that need to be engaged. The list of knowledge collected in this way is ideal for identifying not only the current state of the spatial data management system but also for identifying strategic gaps that should be addressed during the establishment and creation of SDICB [25, 27].

FUTURE WORK

After researching the ways of archiving space research data and their shortcomings, as well as research of various initiatives related to achieving interoperability in the space community, the SDICB concept is considered as the best solution to address these shortcomings. To define the framework for the establishment and implementation of such a concept, it is necessary to answer numerous questions and define all the elements for successful implementation. This includes establishing the definition of the SDICB with the main objectives, vision, and reasons for its establishment. The concept itself must include detailed studies on the needs and identification of users, stakeholders, and the legal basis (policies and standards) for establishment. It is necessary to solve the institutional and organizational structure (level of development and strategic plan) of the project establishment, to define the basic and thematic data, standards, and methods of collecting and storing relevant metadata. As part of the development of such a concept, a survey is currently being conducted to assess the current situation and user needs, and the results of which, along with additional research on the remaining elements of the concept, will serve in the future to establish guidelines for implementing the SDICB concept.

CONCLUSION

Today, there is a great increase in volume and art of space research data and the growing interest of scientists and the public in using this data. Data open and accessible to everyone does not make it easier to find and use it. The archives in which data are stored are intended for their long-term preservation and are focused mainly on data, not the users. To find and interpret data of interest requires too much effort and time, and searches are possible at the missions and instruments levels. Archives do not allow semantic data search, which indicates that there is a problem with insufficiently documented metadata. In addition, one of the main problems are data formats as well as their interoperability and use in standard tools. Over the past few years, several initiatives have emerged to standardize ways of collecting, storing, and distributing space research data to make them interoperable and accessible to a wide range of users. Each of these initiatives provides recommendations on how to store data and related metadata. One of the solutions to this burning problem in space community is the establishment of a SDICB that would effectively distribute data through an agreed set of standards based on recommendations, institutional cooperation agreements and policies. Successful implementation of such a concept would serve the wider community and provide an easy way to search, access and use data for all users, not just space scientists, as is the case with current ways of archiving data. This way of organizing data will allow maximum use of data and facilitate their use, especially for citizens. The scientific and research potential of archived data will increase, which will contribute to a better and easier understanding of, for example, the physics and dynamics of planets and other space bodies, and the understanding of physical and dynamic processes on Earth.

ACKNOWLEDGEMENT

This research is part of TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592.

REFERENCES

[1] Forniter Economics.: *Geospatial Data Market Study.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/937025/Frontier_Economics_-_Geospatial_Data_Market_Study.pdf, accessed 21st February, 2022,

[2] Li, W.; Wang, S. and Bhatia, V.: *PolarHub: A large-scale web crawling engine for OGC service discovery in cyberinfrastructure.* Computers, Environment and Urban Systems **59**, 195-207, 2016, http://dx.doi.org/10.1016/j.compenvurbsys.2016.07.004,

[3] Maguire, D. and Longley, P.: *The emergence of geoportals and their role in spatial data infrastructures.* Computers, Environment and Urban Systems **29**(1), 3-14, 2005, http://dx.doi.org/10.1016/j.compenvurbsys.2004.05.012,

[4] Hu, Y. and Li, W.: *Spatial Data Infrastructures.* In: Wilson, J.P., ed.: *The Geographic Information Science & Technology Body of Knowledge*, Association of American Geographers, Washington, 2017, http://dx.doi.org/10.22224/gistbok/2017.2.1,

[5] Ključanin, S.; Poslonček-Petrić, V. and Bačić, Ž.: *Basics of spatial data infrastructure*. In Croatian.
Dobra knjiga, Sarajevo, 2018,
[6] Breuer, D., et al: *Solar System (Astronomy, Astrophysics and Cosmology).*
Springer, Berlin & Heidelberg, 2009,
http://dx.doi.org/10.1007/978-3-540-88055-4,
[7] United States Congress: *Commerce, Justice, Science, and Related Agencies Appropriations for 2011: Hearings Before a Subcommittee of the Committee on Appropriations, House of Representatives.*
https://www.nasa.gov/pdf/428154main_Planetary_Science.pdf, accessed 9th January 2022,
[8] Hunter, M.: *Supporting the Life-Cycle of planetary geospatial data.*
Esri User Conference, San Diego, 2019,
https://www.esri.com/content/dam/esrisites/en-us/about/events/media/UC-2019/user-presentations/UC_18.pdf, accessed 11th January 2022,
[9] National Aeronautics and Space Administration (NASA): *Planetary Science.*
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, 2019,
https://science.jpl.nasa.gov/division/planetary-science/, accessed 11th January 2022,
[10] Mapping and Planetary Spatial Infrastructure Team: *Mapping and Planetary Spatial Data Infrastructure Roadmap 2019 – 2023.*
https://www.lpi.usra.edu/mapsit/roadmap/MAPSIT-Roadmap-2019-06-19.pdf, accessed 4th January 2022,
[11] Borden, R.M. and Bishop, B.W.: *Assessing Planetary Data Access and Use.*
4th Planetary Data Workshop, Arizona, 2019,
https://www.hou.usra.edu/meetings/planetdata2019/pdf/7004.pdf, accessed 23rd December 2021,
[12] Hare, T.M.; Rossi, A.P.; Frigeri, A. and Marmo, C.: *Interoperability in planetary research for geospatial data analysis.*
Planetary and Space Science **150**, 36-42, 2018,
http://dx.doi.org/10.1016/j.pss.2017.04.004,
[13] Laura, J.R., et al.: *Framework for the Development of Planetary Spatial Data Infrastructures: A Europa Case Study*.
Earth and Space Science **5**, 486-502., 2018,
http://dx.doi.org/10.1029/2018EA000411,
[14] Radebaugh, J., et al.*: Obtaining and Using Planetary Spatial Data into the Future: The Role of the Mapping and Planetary Spatial Infrastructure Team (MAPSIT).*
Planetary Science Vision 2050 Workshop, NASA, Washington, DC, 2017,
https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20170001387.pdf, accessed 21st December 2022,
[15] International Astronomical Union: *Gazetteer of Planetary Nomenclature.*
International Astronomical Union Working Group for Planetary System Nomenclature, Paris, 2021,
[16] German-Aerospace-Center: *Planetary Geodesy.*
https://www.dlr.de/pf/en/desktopdefault.aspx/tabid-170/226_read-397, accessed 4th January 2022,
[17] Bills, B.G.: *Planetary Geodesy*.
In: Finkl, C.W., ed.: *Encyclopedia of Planetary Science. Encyclopedia of Earth Science.* Springer, Dordrecht, 1997,
http://dx.doi.org/10.1007/1-4020-4520-4_313,
[18] Archinal, B.A., et al.: *Report of the IAU Working Group on Cartographic Coordinates and Rotational Elements: 2015,*
Celestial Mechanics and Dynamical Astronomy **130**(3), No. 22, 2018,
http://dx.doi.org/10.1007/s10569-017-9805-5,
[19] Hare, T.M., et al.: *Standards Proposal to Support Planetary Coordinate Reference Systems in Open Geospatial Web Services and Geospatial Applications.*
Lunar Planetary Science Conference XXXVII, Houston, 2006,

[20] Keszthelyi, L., et al.: *Utilizing the Integrated Software for Imagers and Spectrometers (ISIS) to Support Future Missions*.
Lunar Planetary Science Conference XLV, Houston, 2014,
https://www.hou.usra.edu/meetings/lpsc2014/pdf/1686.pdf, accessed 9th January 2022,
[21] Baumann, P.M; Dumitru, V. and Misev, A.D.: *Standards-Based Services for Big Spatio-Temporal Data*.
ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **XLI-B4**, 691-699, 2016,
http://dx.doi.org/10.5194/isprs-archives-XLI-B4-691-2016,
[22] Skinner, J.A., et al.: *Planetary Geologic Mapping Protocol–2018*.
The Planetary Geologic Map Coordination Group, Astrogeology Science Center, 2018,
https://astropedia.astrogeology.usgs.gov/alfresco/d/d/workspace/SpacesStore/01e32dcd-3072-4ac2-8e41-7cc5029bd2cf/PGM_Protocol_March2018.pdf, accessed 11th January 2022,
[23] Nass, A.; van Gasselt, S.; Jaumann, R. and Asche, S.: *Implementation of cartographic symbols for planetary mapping in geographic information systems*.
Planetary Space Science **59**(11-12), 1255-1264, 2010,
http://dx.doi.org/10.1016/j.pss.2010.08.022,
[24] Hare, T.M.: *Standards-Based Collation Tools for Geospatial Metadata in Support of the Planetary Domain*. Ph.D. Thesis.
Northern Arizona University, Flagstaff, 2011,
https://www.researchgate.net/publication/233796188_Standards-based_collation_tools_for_geospatial_metadata_in_support_of_the_planetary_domain, accessed 20th December 2021,
[25] Laura, J.R., et al.: *Towards a Planetary Spatial Data Infrastructure*.
ISPRS International Journal of Geoinformatics **6**(6), No. 181, 2017,
http://dx.doi.org/10.3390/ijgi6060181,
[26] Hendriks, P.H., Dessers, E. and van Hootegem, G.: *Reconsidering the definition of a spatial data infrastructure*.
International Journal of Geographical Information Science **26**, 1479-1494, 2012,
http://dx.doi.org/10.1080/13658816.2011.639301,
[27] Laura, J.R. and Beyer, R.A.: *Knowledge Inventory of Foundational Data Products in Planetary Science*.
The Planetary Science Journal **2**(1), 18-28, 2021,
http://dx.doi.org/10.3847/PSJ/abcb94.

OPEN NATIONAL CORS DATA ECOSYSTEMS: A CROSS-JURISDICTIONAL COMPARISON*

Warakan Supinajaroen, Bastiaan van Loenen and Willem Korthals Altes**

Delft University of Technology
Delft, Netherlands

DOI: 10.7906/10.7906/indecs.20.2.2
Regular article

Received: 14 January 2022.
Accepted: 20 April 2022.

ABSTRACT

Developments toward Open Government Data (OGD) also affect the data from National Continuously Operating Reference Station (NCORS), an infrastructure supporting standard and precise positioning in spatial activities. The application of OD policies on NCORS data (OD-NCORS) varies per country. This article explores the approaches and impact of OD-NCORS in three European countries: Germany, the Netherlands and Sweden. Understanding the differences in the implementations may benefit other countries in their strategies to implement OD for their NCORS. It may also provide insights for organisations considering OD for other data. The research found that the key factors affecting OD-NCORS implementation are the national governing nature and the existence of commercial CORS networks. There is no single approach for OD-NCORS implementation that fits every national context.

KEY WORDS

open data, data ecosystem, CORS, GNSS, Germany, Netherlands, Sweden

CLASSIFICATION

JEL: C89

*This is the extended version of the abstract published in: Vujić, M. and Šalamon, D., eds.: Book of abstracts of the National Open Data Conference. University of Zagreb, Faculty of Traffic and Transport Sciences, Zagreb, 2021.

**Corresponding author, η: w.supinajaroen@tudelft.nl; +31618189376; -

INTRODUCTION

Many public datasets in many countries are made available as Open Government Data (OGD). Also, the data from National Continuously Operating Reference Station networks (NCORS) are provided as OGD in many countries. NCORS networks collect and process data from the Global Navigation Satellite Systems (GNSS) to provide precise positioning data to support spatially related activities. NCORS governing entities have applied different approaches to ensure the stability and functionality of their NCORS under the OD landscape. The countries' experiences and lessons should benefit many other countries facing similar OD-NCORS implementation challenges.

This article explores the approaches, impacts and results of OD implementation on NCORS ecosystems in three members of the European Union (EU): the Netherlands, Sweden, and Germany. After the introduction, an overview of NCORS from a data perspective is elaborated together with an NCORS data ecosystem conception. After that, the concept is applied to explore the cases. The findings are discussed in the next part. Lastly, based on the exploration, this article provides considerations for stakeholders in implementing OD on NCORS in other countries and the implementation of OD on other datasets.

WHAT IS NCORS

CORS is a collective term for ground stations that observe signals from GNSS such as Global Positioning System (GPS), Global Navigation Satellite System (GLONASS), BeiDou and Galileo. In general, CORS can refer to a station or a network of ground stations. The latter one is called a CORS network. CORS can be categorised into three hierarchical levels: global, national and local networks [1]. At national levels, many national governments have established CORS as an infrastructure [2-4] to support several positioning activities. The CORS networks operated by the national governments can be considered National CORS networks or NCORS. In many countries, there are commercial CORS (CCORS) networks.

The functions of NCORS are first to support positioning standardisation through access to the national spatial reference frame, the fundamental function of NCORS data is to allow access to the geodetic reference frame [5-7]. Secondly, NCORS support positioning accuracy through correction services. Thirdly, NCORS serves also national time synchronisation [8].

NCORS has become a requirement for many development initiatives. For smart city initiatives, the precise positioning data from CORS can support the location services such as traffic management, autonomous vehicles and emergency services [9, 10]. NCORS supports digital cadastral [2, 3] as part of E-government initiatives [4]. Precision Agriculture (PA) applies spatial technologies [11] as "right treatment in the right place at the right time" [12]. The PA functions such as precision soil preparation, seeding, crop management, precision harvesting, and data analysis and evaluation benefit from NCORS data [11, 13, 14].

A DATA PERSPECTIVE OF NCORS

The function of NCORS is to provide data that improve the accuracy of and standardise other positioning data. NCORS data serve as a national geospatial data framework – the data that other spatial data are built upon [15]. NCORS data serve the correction services for GNSS positioning. With these two functions, users can collect, process, and distribute the high precision spatial data that are interoperable with other spatial data in the national reference frame. The flows of NCORS data in serving the two functions are provided both in real-time data through Real-Time Kinematic (RTK), Differential GNSS (DGNSS) and Precise Point Positioning (PPP) services. For archive data, NCORS data is provided, such as RINEX.

AN NCORS DATA ECOSYSTEM

An ecosystem refers to an integrated system composed of a biotic community, its environment, and its dynamic interactions [16]; these components could not be separated or viewed in isolation [17]. The elements in an ecosystem coexist and influence each other. External forces also affect an ecosystem, such as climate changes and natural disasters that influence the components in the ecosystem. The ecosystem concept was applied to explain, understand, model and replicate the factors or elements and their relations in many academic and practical fields.

Data science scholars have defined several data ecosystem conceptions [16-20]. Data ecosystems are the networks of socio-technical components to collaborate on creating, managing, and sustainable data-sharing initiatives [19]. A data ecosystem is a frame of the relations of the actors, which can be individuals or entities and technical components to reach the common data goals [20]. In OD, the completion of an OD ecosystem must be (1) user-driven – to satisfy different types of users, (2) circular – to allow the stakeholders in satisfying and contributing to the data value chain, (3) inclusive – to stimulate the participation from non-government actors, and (4) skill-based – to provide OD and relevant knowledge to people [21].

From these conceptions, a data ecosystem refers to the data existence and the interacted elements with a function(s) to facilitate the flow of the data. In this article, an NCORS data ecosystem is defined as *"A system where technical and institutional elements at the national level interact to facilitate the flow of NCORS data from providers to users"*.

The Elements in an NCORS Data Ecosystem

The primary functions of NCORS data support the standardisation and interoperability of spatial data at the national level. This way, it should be considered part of the National Spatial Data Infrastructure (SDI). The elementary aspect of an NCORS data ecosystem can adopt from an SDI. Van Loenen [22] explains that the core elements of SDI are 1) Data, 2) People: the actors in different sectors that require, build, use and enforce the actions for the existence of spatial data, 3) Policies: plans or courses of action to achieve the goal of the spatial data, 4) Institutional framework: the responsibility arrangement of different actors involving the spatial data, 5) Technology: the scientific method, instrument, data and material, directly and indirectly, used to enable the spatial data chain, 6) Standards: the common specification, quality, or requirements which allow the flow of spatial data between the processes and organisations, and 7) Financial resources: the interconnected resources to drive all elements.

In the NCORS data ecosystem, two main elements can be distinguished 1) the central and 2) the surrounding elements, Figure 1. An NCORS data chain and NCORS stations are the central elements. The surrounding elements are people, financial resources, institutional frameworks, technology, standards, and policies. These elements interact to facilitate the functions of the NCORS network and the flow of NCORS data from providers to users.

The NCORS data chain contains different data formats: from GNSS signals to precise positioning data. The NCORS data chain carries the data value which can be perceived in different positioning services, for example, Real-time Kinematic (RTK), Differential-GNSS (DGNSS), Precise Point Positioning (PPP), PPP-RTK, Post-Processing Kinematic (PPK), and data archive.

Mechanisms in an NCORS Data Ecosystem

The NCORS data ecosystem's elements have a function(s) in collectively facilitating the NCORS data chain. These collective functions in an NCORS data ecosystem can be considered in three mechanisms: (1) governance, (2) provision and (3) utilisation.

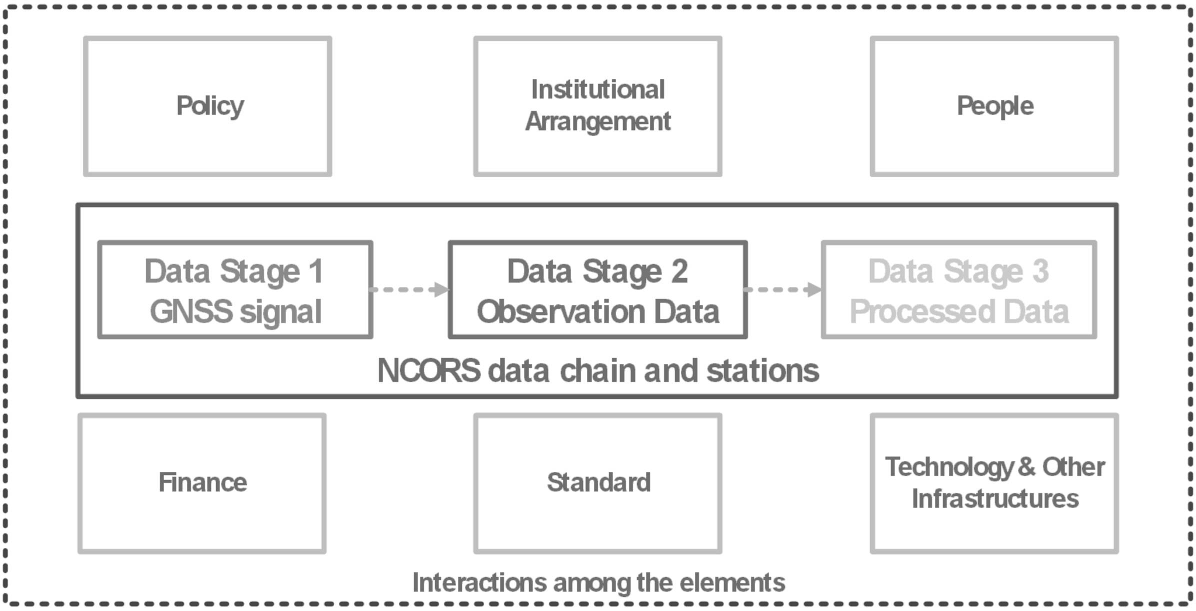


Figure 1. An NCORS data ecosystem, adapted from [23].

The governance of NCORS data is the mechanism to govern the technical and non-technical elements with different functions to achieve the individual and the collective goal(s) [24, 25]. The governance can be perceived through five aspects of SDI governance: vision, leadership and control, self-organisation ability, communication channel, and financial resource sustainability [18, 26]. The governance involves the provision and the utilisation. The governance also involves external factors such as political support and facilitating conditions that significantly affect the NCORS data ecosystem.

Provision is the collective actions of the stakeholders in collecting, processing, and distributing NCORS data. The quality of the provision can be measured by the attributes of data availability [18], such as known – user awareness about the availability and user knowledge about the access to such availability; attainable – users can access and use the data; and usable – the characteristics of the NCORS data that meet user requirements. In many cases, CCORS expands the availability of overall national CORS services [27, 28]. It is, therefore, part of the NCORS data ecosystem.

Utilisation is the mechanism that determines the use of NCORS data which is a critical driving force to initiate and maintain an NCORS data chain. The utilisation provides feedback to data governance and provision. It also influences governance through political and public support. The utilisation of NCORS data can be considered in many indicators, for example, the number of users or accounts, the number of downloads, the volume of downloaded data, time of use, and registered devices [27-29]. There are benefits and drawbacks of each indicator in explaining utilisation. For instance, one user might use many devices or the other way around. This research considers NCORS data utilisation as basic utilisation – the use within original users or purposes, and optimal utilisation – the utilisation that expands beyond original users or purposes.

OPEN DATA AND NCORS

OD means the data that is made available without any technical, legal, financial, and intellectual barriers to access, use, reuse, and share the data [30, 31]. The implementation of OD may affect the mechanisms in an NCORS data ecosystem (Figure 2). OD frames the data policy of NCORS governance. In terms of the provision of NCORS, OD affects the degree of NCORS data availability directly through both “known” and “attainable” attributes. Lastly,

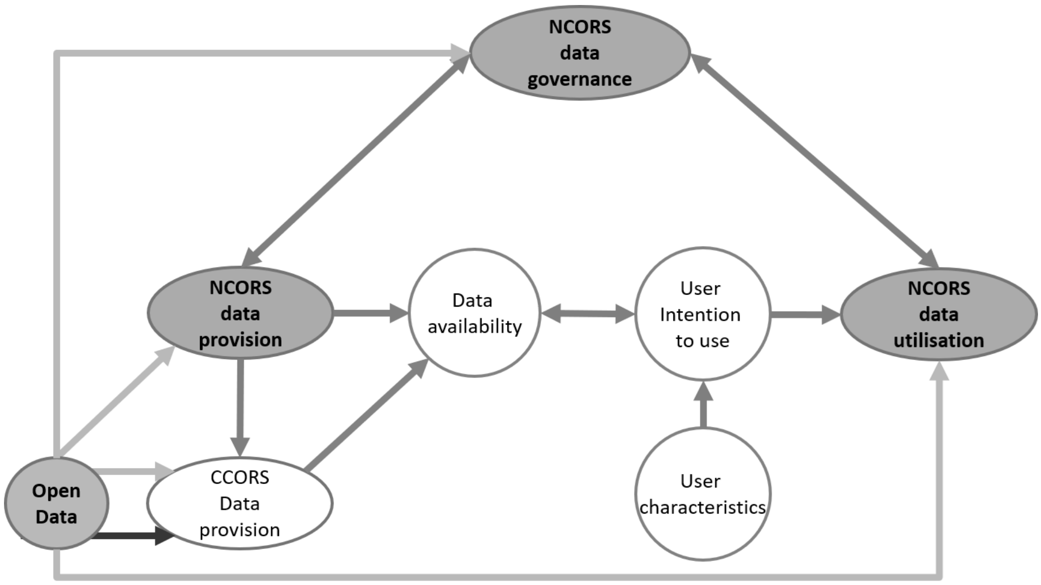


Figure 2. The impacts of OD on the mechanisms of an NCORS ecosystem.

OD's effect on data provision stimulates users to use NCORS data. This way, OD supports the utilisation mechanism in an NCORS ecosystem. Note that user characteristics also contribute to the utilisation.

METHODOLOGY

For this research, we performed a multiple case study approach [32] to investigate the effects and results of OD implementation on NCORS data ecosystems and the factors affecting the implementation. Three cases were selected based on the completeness of NCORS data availability, the availability of the national OD initiative, and the accessibility to the case data.

Desk research was conducted through papers, journals, publications and official reports to acquire the data. The analysis was conducted based on the three mechanisms of the NCORS data ecosystem. Lastly, experts from the cases were asked to review and validate the research findings and provide additional insights about the cases.

OPEN DATA AND NCORS IN DIFFERENT NATIONAL CONTEXTS

The European Union is the political and economic region where member nations have progressively implemented OD. It is also where NCORS and CCORS have been installed and operated. In 2019, the scope of the EU legal framework on open data and the Public Sector Information Directive (PSI) was broadened to real-time data. Consequently, all kinds of NCORS data have become within the scope. However, each Member State has applied OD to NCORS differently.

THE NETHERLANDS

The Dutch NCORS is composed of two networks 1) "Het Actief GNSS Referentie Systeem" (AGRS.NL) – the active GNSS reference System and 2) "De Netherlands Positioning Service" (NETPOS) – the Netherlands Positioning Service. AGRS.NL maintains the Dutch spatial reference frame [34, 35], authorises other GNSS networks, and supports other scientific activities[33]. NETPOS is the second-order network to densify AGRS.NL and to support public works of Kadaster, Rijkswaterstaat, research, education and innovation [34]. There are several CCORS in the Netherlands. The stations from CCORS networks can be determined and certified by AGRS.NL.

The Netherlands is one of the high-rank countries in OD implementation. The OD strategy is part of the National Data Strategy, aligning with the EU Data Strategy [35]. OD has been successfully implemented in many ministries [36].

OD on Dutch NCORS Data Governance

Nederlandse Samenwerking Geodetische Infrastructuur (NSGI) or the Dutch Cooperation Geodetic Infrastructure is the governing body of the NCORS networks. NSGI is a “virtual organisation based on integral management” consisting of three geodetic partners: Dutch Kadaster – horizontal reference system, Rijkswaterstaat – vertical reference system and the Hydrographic Service of the Royal Netherlands Navy – sea level [34]. NSGI facilitates NCORS networks to serve public activities and standardise the positioning services of other CORS networks using the common geospatial reference frame. NSGI assigns the Dutch Kadaster to certify stations of the NCORS and CCORS [37] over the Netherlands and in the neighbouring countries, Figure 3.

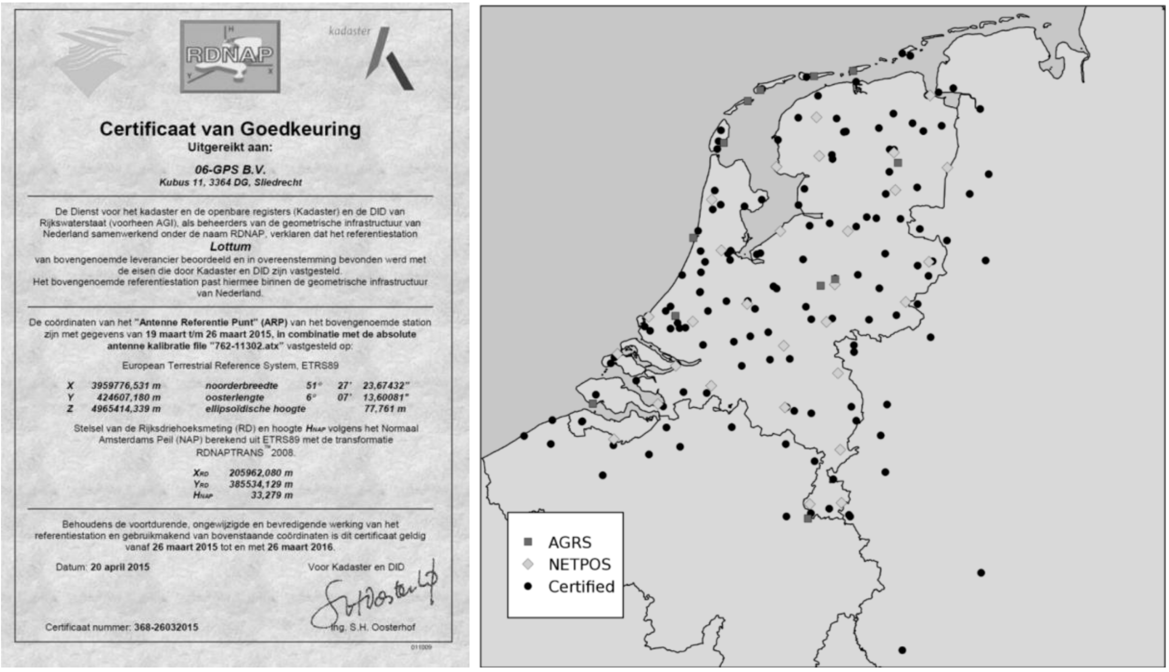


Figure 3. CORS certificate issued by NSGI (left). Public and private CORS stations that are certified (right) [37].

OD is mentioned in the NSGI multi-year plan 2021-2025 [34]. However, NCORS stakeholders share diverse opinions and intentions toward OD-NCORS. An expert of the NCORS provider support OD-NCORS. Since NCORS data are public data, citizens should use them without paying. Meanwhile, a CCORS manager expressed a concern that OD-NCORS might affect their market.

The conditions above may presumably contribute to the decree on further release of raw GNSS data in a file format (Besluit verdere vrijgave van ruwe GNSS-data in bestandsvorm) [38]. Only RINEX has been open since September 7th 2019 [39]. Nevertheless, NSGI seeks to apply OD-NCORS for scientific and primary geodetic tasks [34].

OD on Dutch NCORS Data Provision

NCORS data in the Netherlands is available nationwide. In combing with CCORS, the availability of positioning data can support the use in low-to-high positioning accuracy for any geospatial data activities. OD-NCORS appears only on RINEX from NETPOS [39]. One informant reported that this implementation is to prevent the overlapping of the RTK service of the CCORS. However, the provision of CCORS relies on NCORS since CCORS stations are certified by NCORS.

OD on Dutch NCORS Data Utilisation

OD slightly affects the utilisation of NCORS since it applies only to RINEX data. The Dutch NCORS benefits scientific works, survey and mapping, engineering, navigation and meteorological purposes on national and international scales. Government agencies and academics have been the primary users. According to an expert from NSGI, most users access the data through the Delft University of Technology, the Reference Frame Sub-Commission for Europe (EUREF) and the International GNSS Service (IGS) that stream and share data from some stations. The RTK and PPK mainly were used for cadastral survey and mapping of Kadaster, construction works of Rijkswaterstaat and other government activities, Figures 4 and 5. In addition to the positioning services, NSGI sets a goal of NETPOS "to act as the control service for GNSS product certification" [34].

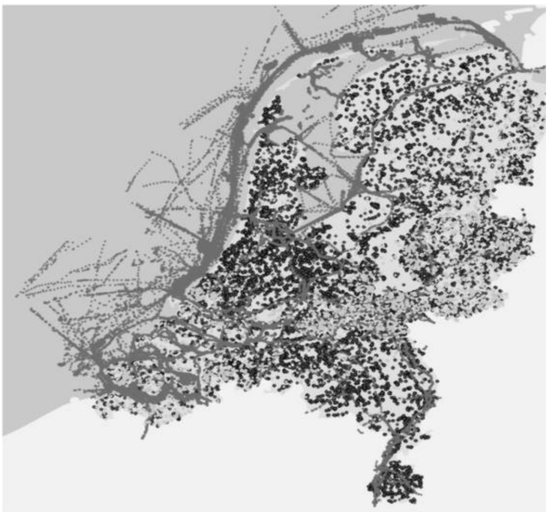


Figure 4. The use of NETPOS RTK in 2017 Black: Kadaster Blue: Rijkswaterstaat, Grey: Other (contractors /education and research) [37].

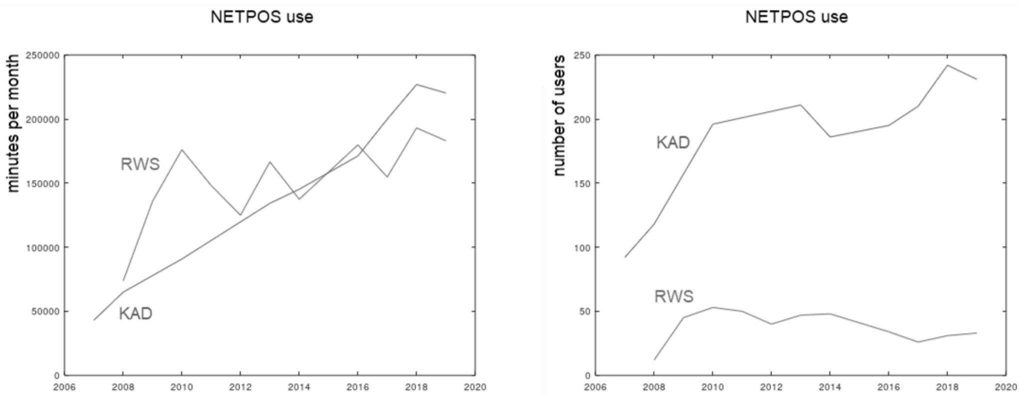


Figure 5. The data use of NETPOS in cadastral and public works. Source: NETPOS based on a September each year with estimation of some missing data.

The availability of CCORS should imply CORS data use beyond NCORS. A CCORS provider explained that a few commercial providers initially serviced RTK in the early 2000s. At that time, some government agencies were also the users of CCORS. Agriculture has been a large sector of CORS data use in the Netherlands since 2010. Half of the Dutch farmers use RTK from the CCORS. NCORS data use in the commercial market is seen through CCORS certified and standardised on the reference frame by the NCORS.

GERMANY

Satellitenpositionierungsdienst der Deutschen Landesvermessung, known as SAPOS is the NCORS network operated by Germany's federal and state governments (Laender). The network was established in 1994 as a prototype and started to service nationwide in 2003 [40] with around 270 stations [41]. The Working Committee of the Surveying Authorities of the Laender of the Federal Republic of Germany (AdV) is the SAPOS governing body. The geospatial related authorities in the federal governments and several authorities in the state governments participate in AdV. SAPOS in the states is networked into the national public network and controlled by the Central Office SAPOS (ZSS). CCORS networks are available in Germany.

The OD policy has been part of the national E-government act passed by the parliament in May 2017. PSI initiative was adopted into the Act on the Reuse of Public Sector Information, the Information Re-Use Act (IWG) [40]. In February 2021, the federal government adopted the second Open Data Act and a Data Use Act (DNG) to extensively expand the availability of open administrative data and improve the possibilities for using publicly financed OD. The OD strategy was formed with stakeholder involvement [35]. For example, data providers and users contributed to selecting high-value datasets at the national level.

OD on SAPOS Data Governance

Since 2014, the OD initiative has contributed to the vision of AdV and SAPOS accordingly. The vision covered the digitisation of administrative procedures to serve e-government and the 'stakeholders' transparency, participation, and cooperation. In the part of AdV, SAPOS shall ensure highly accurate and uniform correction data services in Germany's official three-dimensional reference system (ETRS89 / DREF91) [42, 43].

OD has an impact on SAPOS data policy and financial resources. Before the national OD campaign in 2017, SAPOS generated revenue through the data services based on the AdV guideline of the fees for providing and using geospatial base data [44]. Implementing an OD policy has been at the state with different contexts and approaches. Most states can implement OD on their SAPOS. According to an expert, the OD policy impacts the CCORS market – the public sector competes with the private sector. Compensation has been one of the solutions for such an impact. One example is that a state authority leaves GNSS machine calibration services to the CCORS providers in exchange for OD on SAPOS.

OD on SAPOS Data Provision

SAPOS provides 1) Real-Time Positioning Service (EPS) with decimetre accuracy, 2) High Precision Real-Time Positioning Service (HEPS) with centimetre accuracy, and 3) Geodetic Precision Positioning Service (GPPS) with millimetre accuracy level [45, 46]. AdV seeks to employ PPP-RTK as a SAPOS positioning service [47].

OD has been fully implemented in some states (Figure 6). For example, in North Rhine Wespalean [46], Für Niedersachsen und Bremen [48], and Berlin. Hamburg and Schleswig-Holstein will implement OD in 2022 to support the digital strategy [49]. Some states apply OD for only agriculture. Mecklenburg-Western Pomerania does not implement OD but provides a cost exemption for the public transport authorities and other mutual exchanges.

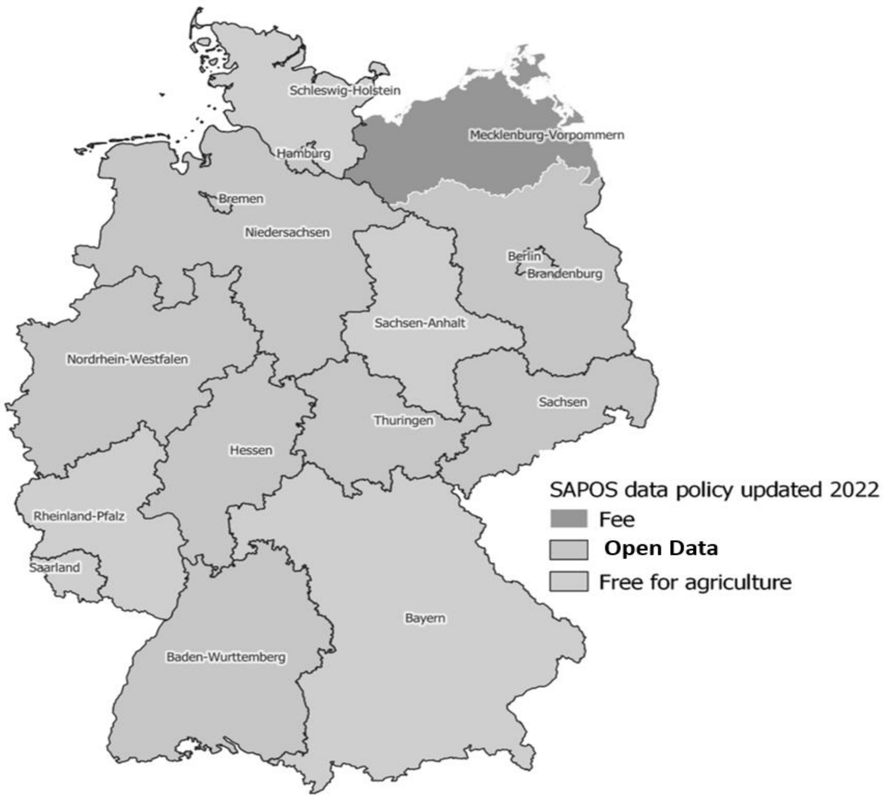


Figure 6. OD-SAPOS in different states in 2022, updated from Riecken [50].

OD on SAPOS Data Utilisation

OD on SAPOS is seen at the state level. An example of an OD state-wide impact is North-Rhine Westphalia (Figure 7). The number of new SAPOS users increased drastically after implementing OD [50]. The growth of user volume also expanded in agriculture [51]. Nonetheless, the SAPOS data benefit many sectors [40].

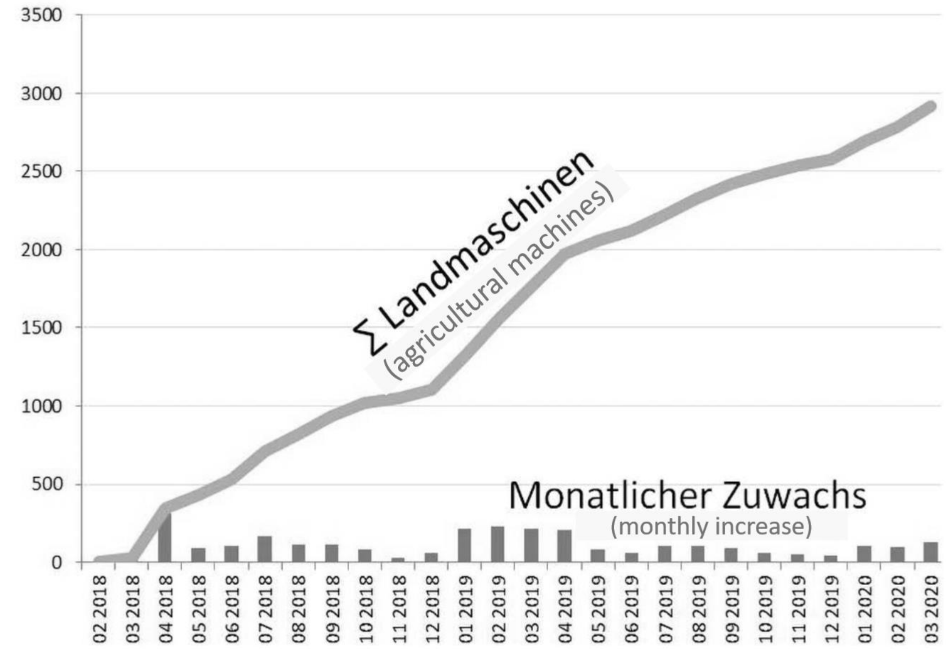


Figure 7. The impact of OD on the use of SAPOS in agriculture of North-Rhine Westphalia, adapted from [47].

SWEDEN

The Swedish public CORS network is SWEPOS™ (ett Svenskt nätverk av fasta referensstationer) established in early 1990s for scientific and public works. In 1992, there was an initiative to utilise SWEPOS for other domains. Later, pilot Network-RTK projects were carried out in many areas [52], together with the densification of SWEPOS to support RTK and other positioning applications in many domains.

Lantmäteriet (the Swedish mapping, cadastral and land registration authority), the leading organisation of SWEPOS, classifies two SWEPOS stations: Class A and B. Class A consists of 21 stations that function in the national spatial reference frame (SWEREF 99). Class B stations are installed to densify and expand the service coverage. Class B stations are accounted for 90% of the SWEPOS. There are other CCORS networks in Sweden [53].

According to Open Data Maturity Report 2021 [35], several OD policies have been implemented in Sweden, such as the national principles for 'Open by 'Default' by the Agency for Digital Government (DIGG), the European Interoperability Framework (EIF), and the adoption of the PSI directive. The Swedish Fundamental Law also states in the principle of public access that "official information is available for reuse". However, the availability of financial resources is a challenge in implementing OD since many public organisations are financed by the revenues from the data they provide. The solution for such issues is politically sensitive and had to reach a verdict in the Swedish parliament [35].

OD on SWEPOS Data Governance

Some SWEPOS settings affect the OD on SWEPOS. Lantmäteriet's geodetic activities 2011-2020 plan states that "to meet Swedish society's needs for a homogeneous, sustainable national geodetic infrastructure and guarantee its availability and use" [54]. Such a vision guides the goal of SWEPOS to genuinely serve society through national positioning services by coordinating with any stakeholders in public and private sectors and international. Densification of the CORS network nationwide is a goal of the SWEPOS [54]. SWEPOS is funded by the government and revenue from data and services. SWEPOS establishment costs are covered by governmental funding. The maintenance, operation and future upgrades costs are covered by the user subscription fees [55]. These settings frame the implementation of OD on SWEPOS. Under the Act on Reuse of Public Administration Documents (2010: 566) and the revision in 2019, the Swedish authorities must publish a government's open data list. Lantmäteriet was assigned to define the valuable datasets that should be made available according to the new OD directive (EU Directive 2019/1024). Even though SWEPOS data were proposed as high-value data [56], OD has not been fully implemented for SWEPOS data.

OD on SWEPOS Data Provision

OD has some extent to the provision of SWEPOS data that are available in the whole of Sweden and along the border with the data exchange collaboration with other Nordic countries [57]. SWEPOS data have been under OD, for instance, DGNSS and RINEX, since 2016 [57]. However, the use of RTK is charged. CCORS providers also stream, process and redistribute SWEPOS data to their customers against fees [29]. The data cooperation with CCORS is to increase the use of SWEPOS data [58]. An expert viewed that the availability of CCORS is necessary to ensure nationwide positioning services, which benefits users in society.

OD on SWEPOS Data Utilisation

The utilisation of SWEPOS data is in many activities, Figure 8. The total use of SWEPOS in all sectors has been exponentially increased since 2003. Note that there were already 200 users at the beginning of SWEPOS RTK in 2004 [29]. Jensen [59] found three main areas of use; 71% in building and construction, 15% in farming and 4% in other activities. Between 2013 and 2018, the use in farming and construction significantly increased from 5% to 15% and 22% to 25%, respectively. It can be said that without full OD implementation, the utilisation of SWEPOS is already significant.

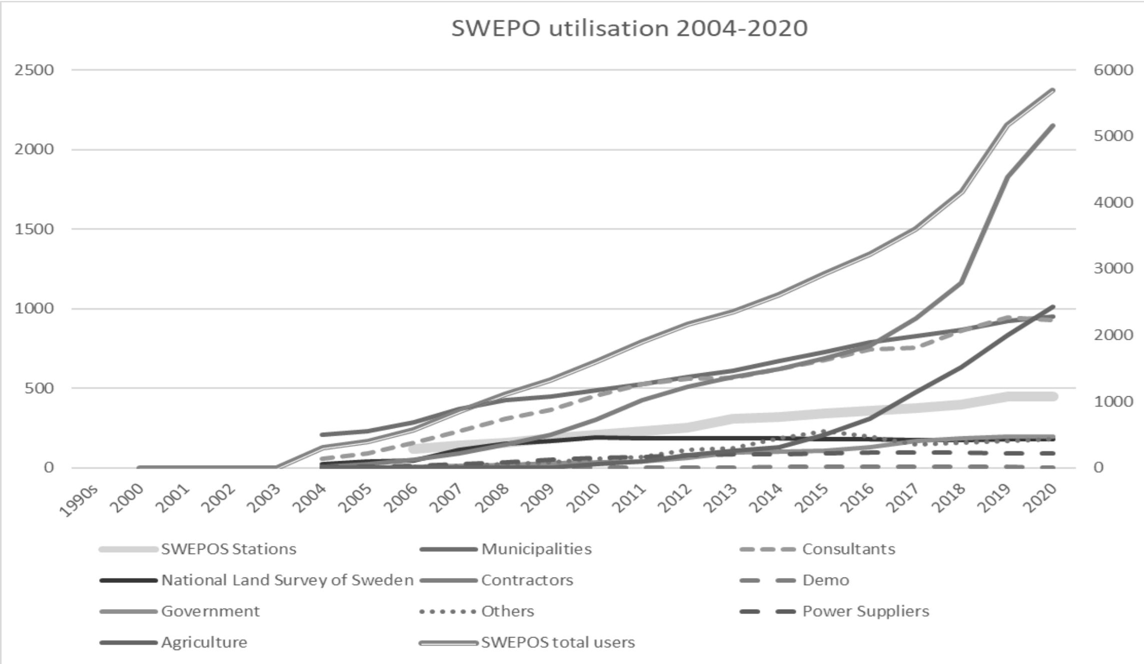


Figure 8. The use of SWEPOS data (based on the 2020 statistics from Lantmäteriet).

DISCUSSION

THE IMPACT OF OD ON THE NCORS DATA ECOSYSTEMS

NCORS data governance is affected by OD. The governance faces a role in dealing with the factors that affect the internal mechanisms of the NCORS data ecosystem, for example, the private sector, political support, financial resources, regulations, and national visions.

In Sweden, OD has been implemented on NCORS alongside the goal of national positioning services. In this way, the charge in some data as the financial support for NCORS and the cooperative competition with CCORS are maintained. In the Netherlands, the implementation of OD has been in conflict with the interests of CCORS. However, the interpretation of the law and the political support preserve the CCORS interest by preventing NCORS from being a competitor. Further relations between NCORS and CCORS are in the standardisation of the services where NCORS certifies CCORS. In Germany, the negotiation between the states and CCORS resulted in the OD-NCORS in most states. A state bartered OD implementation by offering a GNSS equipment calibration service for a CCORS provider.

NCORS data provision is affected by OD through the governance. The perception of CCORS affects the implementation of OD on the NCORS data provision. OD improves the attainability of NCORS data. However, this is not the case for Sweden, where national positioning services are considered on the availability of NCORS and CCORS. OD is applied on only RINEX and DGNSS, which seem not to be the markets of CCORS. In the Netherlands, the RINEX data is free of charge. In Germany, most states adopt OD to NCORS data. The nationwide services are still charged.

NCORS data use, the data use in all cases is considerable. However, all three cases provided an unclear explanation of whether OD impacts NCORS data use. Comparable NCORS data use numbers were found despite different levels of OD implementation.

In the Netherlands, the cost of use in CCORS implies that some users are willing to pay for the services. In Sweden, the use of NCORS is still considerable in many fields. In Germany, the surge in use after OD in the NRW can result from new users starting to use precise positioning services or the mobilisation of users from CCORS to NCORS. Therefore, implementing OD-NCORS might not be the most critical factor in using NCORS data. The impacts of OD-NCORS are in Table 1.

Table 1. The impacts of OD on NCORS data ecosystems in the Netherlands, Sweden and Germany.

Case	Governance	Provision	Utilisation
The Netherlands	Single governing body with multi-public organisations An implication of CCORS influence on OD-NCORS implementation NCORS certifies CCORS.	OD for academic purposes RTK and DGNSS: not in the commercial market RINEX: OD	Serve the original purposes among the government agencies, their contractors, and academics Optimal use in many sectors
Sweden	Single governing organisation A role in defining high-value datasets, including NCORS data Cooperative competition between CCORS and NCORS to achieve the goal of national positioning services	RTK: at a price in the market with CCORS DGNSS and RINEX: OD	Serve the original and other purposes of the public use and commercial use Optimal use in many sectors
Germany	OD is implemented at the state level with different approaches, not nationwide service CCORS influences OD-NCORS implementattion in some states	DGNSS and RINEX: OD OD at the state level varies among the state contexts	Serve the original and other purposes of the public use and commercial use Optimal use in many sectors

THE FACTORS SHAPING OD-NCORS IMPLEMENTATION

We can conclude that the factors contributing to OD-NCORS implementation are 1) the national governing nature and 2) the non-government actors.

Firstly, the national governing nature frames how NCORS governing entities adopt OD. The factor is visible in the national visions and the governing structures in executing them to reach such vision. The three cases share a similar national vision and political awareness in OD. However, the government in charge of NCORS determine the extent of OD implementation. The NCORS governing structures in the Netherlands and Sweden are under the central government. In Germany, the state governments are the governing body to implement OD according to the national strategy. The adoption of OD on NCORS differs among the states due to their contexts.

Secondly, the availability of actors beyond the public sector frames OD-NCORS implementation. The three cases showed different perspectives on the availability of CCORS which affect how OD implements and how NCORS networks position themselves with CCORS. Several interpretations of OD and NCORS might affect OD-NCORS. The Netherlands holds that OD should not affect third parties' interests – i.e., the interests of CCORS providers. Therefore, OD is only applied to the NCORS data that are not competing with CCORS data. CCORS, as an element of a national positioning service, results in a cooperative competition between CCORS and NCORS in Sweden. In Germany, the states adopted their approaches to dealing with CCORS.

IMPLICATIONS

Theoretical implications in this study are found in the comprehension of a data ecosystem through the lens of NCORS data. A data ecosystem is part of the larger ecosystem – a system of systems. The data ecosystem interrelates with other ecosystems, such as business ecosystems and technology ecosystems. Therefore, data governance should be highlighted for the interactions with other elements in other ecosystems. As suggested in SDI, communication channels are essential in gaining political and public support [26].

Measuring data use was still a challenge. The different ways to measure data use can be the number of requests, hits, views, transformation, citations, and the number of users [60-62]. These indicators are applicable for the direct use of NCORS data. However, NCORS data use can be perceived in other ways, as seen in the cases. Data from some NCORS stations are used with CCORS data to provide positioning services, as seen in Sweden. CCORS might be processed or calibrated on the national framework provided by NCORS data, as seen in the Netherlands. These ways of use may imply the necessity for defining NCORS data and other data utilisation.

Practical implications are seen in many aspects. First, the governing body needs to understand the settings of each data and its ecosystem. OD has both positive and negative effects on data availability. Without a clear understanding of the impact of OD on NCORS, the OD implementation might harm the data ecosystem itself.

Second, the practical implications might be a lesson for other data with different stakeholders with diverse perspectives. The role of data governance should follow the so-called informally governance to keep all stakeholders in the loop. The outcome might compromise some initiatives. Still, the compromising must rely on users, the key actors in the utilisation.

Third, the findings imply that user characteristics can be more critical than the data policy in the successful implementation of OD-NCROS. The history of use of the three cases implies the user characteristics: their technical skills, perception about the use, financial capability, and the necessity of use. In the cases, users were capable of using, and it was necessary for them to use despite the relevant costs. Therefore, any OD approaches result in the considerable use of NCORS data. In practice, policymakers should consider user characteristics as part of OD implementation.

CONCLUSION AND RECOMMENDATIONS

This article has applied an NCORS data ecosystem concept to explore Open Data implementation in National Continuously Operating Reference Station Networks (NCORS) in three Member States of the European Union. OD has positive and negative impacts on the NCORS data ecosystem. The common benefit is the availability of data for users. However, OD-NCORS may negatively impact the availability of commercial CORS (CCORS). The national governing nature and CCORS availability shape the OD implementation. Overall, we can conclude that no single approach for OD-NCORS implementation fits every national context. The countries must face the challenges by compromising both the goals of NCORS and OD in a sustainable manner.

This study pertains to limitations on the data to indicate the volume of use from the private sector. Besides, the data ecosystem conception in this study still has a gap in defining data utilisation. Future research is encouraged to apply the NCORS data ecosystem concept to explore NCORS in other contexts where the NCORS elements and mechanisms are not perfectly functioning. The application may provide insights to improve the NCORS data ecosystem conception. The data utilisation defined in this research can be further investigated in terms of the impact to provide a complete picture of data utilisation.

ACKNOWLEDGEMENTS

This article presents the findings of the PhD research “Encouraging the use of national CORS data in Thailand from a user perspective” conducted at the Knowledge Centre Open Data, Delft University of Technology, funded by Her Royal Highness Princess Maha Chakri Sirindhorn of Thailand scholarship. The authors would thank the contributions of the experts from the selected cases that provided feedback and inside information.

This research is part of TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592.

REFERENCES

[1] Rizos, C. and Satirapod, C.: *Contribution of GNSS CORS infrastructure to the mission of modern geodesy and status of GNSS CORS in Thailand.*
Engineering Journal **15**(1), 25-42, 2011,
http://dx.doi.org/10.4186/ej.2011.15.1.25,

[2] Van Buren, J. and Lesparre, J.: *Netherlands Positioning Service, A Real Time Kinematic Network for the Governmental Authorities of the Netherlands.*
EUREF Symposia 2005. EUREF, Vienna, 2005,
http://www.euref.eu/symposia/2005Vienna/3-05.pdf,

[3] Jarroush, J.; Adler, R. and Zeibak, M.: *Cadastre Surveys with Real Time Kinematic GPS (RTK) as a Basis for Future Survey Regulations*
FIG Working Week 2005 and GSDI-8, 2005. Cairo, Egypt,
https://www.oicrf.org/documents/40950/43224/Cadastre+surveys+with+real+time+kinematic+GPS+RTK+as+a+basis+for+future+survey+regulations.pdf/91f82fbd-7245-81b7-679a-7c20622725bc?t=1510174536786,

[4] Van Der Molen P. and Wubbe, M.: *E-Government and E-Land Administration As an example: The Netherlands.*
6th FIG Regional Conference, San José, 2007,
https://www.fig.net/resources/proceedings/fig_proceedings/costarica/papers/ts10/ts10_02_wubbe_vandermolen_2480.pdf,

[5] Lipatnikov, L. and Shevchuk, S.: *Cost Effective Precise Positioning with GNSS*.
FIG report, International Federation of Surveyors Commission 5, Copenhagen, 2019,
https://www.fig.net/resources/publications/figpub/pub74/Figpub74.pdf,

[6] Soler, T.; Snay, R.; Foote, R. and Cline, M.: *Maintaining accurate coordinates for the National CORS network.*
FIG Working Week, 13-17 April 2003. Paris, 2003.
https://www.ngs.noaa.gov/CORS/Articles/FIGParis.pdf,

[7] Snay, R. and Soler, T.: *Continuously operating reference station (CORS): history, applications, and future enhancements.*
Journal of Surveying Engineering **134**(4), 95-104, 2008,
http://dx.doi.org/10.1061/(ASCE)0733-9453(2008)134:4(95),

[8] Hasan, K.; Feng, Y. and Tian, Y.-C.: *GNSS time synchronization in vehicular ad-hoc networks: Benefits and feasibility.*
IEEE Transactions on Intelligent Transportation Systems **19**(12), 3915-3924, 2018,
http://dx.doi.org/10.1109/TITS.2017.2789291,

[9] Minetto, A., et al.: *A Testbed for GNSS-Based Positioning and Navigation Technologies in Smart Cities: The HANSEL Project.*
Smart Cities **3**(4), 1219-1241, 2020,
http://dx.doi.org/10.3390/smartcities3040060,

[10] Rosenthal, G.; Milev, G.; Rokahr, F. and Vassileva, K.: *SAPOS® DGNSS Reference Station System in Germany and its Analogue in Bulgaria*.
United Nations/United States of America Workshop on the Use of Global Navigation Satellite Systems 2001,
https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.542.9866&rep=rep1&type=pdf,

[11] Hicks, A., et al.;*Precision Agriculture in Rice Production, Implementation and Grower Insights*.
https://agrifuturesrice.squarespace.com/s/PA-in-Rice-Production-implementation-and-insights-Andrew-Whitlock.pdf,

[12] Gebbers, R. and Adamchuk, V.: *Precision agriculture and food security.*
Science **327**(5967), 828-831, 2010,
http://dx.doi.org/10.1126/science.1183899,

[13] European Global Navigation Satellite Agency: *Report on agriculture user needs and requirements Outcome of the European GNSS' User Consultation Platform*.
https://www.gsc-europa.eu/sites/default/files/sites/all/files/Report_on_User_Needs_and_Requirements_Agriculture.pdf,

[14] Pini, M., et al.: *Experimental Testbed and Methodology for the Assessment of RTK GNSS Receivers Used in Precision Agriculture.*
IEEE Access **8**, 14690-14703, 2020,
http://dx.doi.org/10.1109/ACCESS.2020.2965741,

[15] Van Loenen, B.: *Developing geographic information infrastructures: the role of information policies.* Ph.D. Thesis.
OTB Research Institute for Housing, Urban and Mobility Studies, Delft University of Technology, Delft, 2006,

[16] Demchenko, Y.; De Laat, C. and Membrey, P.: *Defining architecture components of the Big Data Ecosystem.*
International Conference on Collaboration Technologies and Systems (CTS), 2014,

[17] Schowalter, T.: *Ecosystem Structure and Function.*
Academic Press, 2016,

[18] Welle Donker, F. and Van Loenen, B.: *How to assess the success of the open data ecosystem?*
International Journal of Digital Earth **10**(3), 284-306, 2017,
http://dx.doi.org/10.1080/17538947.2016.1224938,

[19] Oliveira, M.L.; Oliveira, L.E.; Batista, M.R. and Lóscio, B.F.: *Towards a meta-model for data ecosystems.*
The 19th Annual International Conference on Digital Government Research: Governance in the Data Age, 2018,
http://dx.doi.org/10.1145/3209281.3209333,

[20] Otto, B., et al.: *Data Ecosystems*.
Fraunhofer Institute for Software and Systems Engineering, Dortmund, 2019,

[21] Van Loenen, B., et al.: *Towards value-creating and sustainable open data ecosystems: A comparative case study and a research agenda.*
eJournal of eDemocracy and Open Government (JeDEM) **13**(2), 1-27, 2021,

[22] Van Loenen, B.: *Developing geographic information infrastructures: the role of access policies.*
International Journal of Geographical Information Science **23**(2), 195-212, 2009,
http://dx.doi.org/10.1080/13658810701851412,

[23] Supinajaroen, W. and Van Loenen, B.: *Selection of the Elements for a Continuously Operating Reference Station Ecosystem Conceptual Model.*
FIG Working Week, Hanoi, 2019,

[24] Hausler, G.: *National positioning infrastructure: technical, institutional and economic criteria for coordinating access to Australia's GNSS CORS infrastructure.* Ph.D. Thesis.
The University of Melbourne, Victoria,2014,

[25] Janssen, V.; Haasdyk, J. and McElroy, S.: *CORSnet-NSW: A success story.*
Proceedings of Association of Public Authority Surveyors Conference (APAS2016), New South Wales, 2016,
https://www.spatial.nsw.gov.au/__data/assets/pdf_file/0009/209484/2016_Janssen_etal_APAS2016_CORSnet-NSW.pdf,

[26] Kok, B. and van Loenen, B.: *How to assess the success of National Spatial Data Infrastructures?*
Computers, Environment and Urban Systems **29**(6), 699-717, 2005,
http://dx.doi.org/10.1016/j.compenvurbsys.2004.02.001,

[27] Leveson, I.: *Socio-Economic Benefits Study: Scoping the Value of CORS and GRAV-D.*
https://www.ngs.noaa.gov/PUBS_LIB/Socio-EconomicBenefitsofCORSandGRAV-D.pdf,

[28] Rizos, C.; Higgins, M. and Hewitson, S.: *New GNSS Developments and Their Impact on CORS Service Providers.*
4th International Symposium and Exhibition on Geoinformation, Penang, 2005,

[29] Lilje, M.; Wiklund, P. and Hedling, G.: *The use of GNSS in Sweden and the national CORS network SWEPOS.*
FIG XXV International Congress, 2014,

[30] Van Loenen, B.: *Towards a User-Oriented Open Data Strategy.*
T.M.C. Asser Press, The Hague, 2018,

[31] Van Loenen, B.; Vancauwenberghe, G.; Crompvoets, J. and Dalla Corte, L.: *Open data exposed.*
Springer, The Hague, 2018,

[32] Steenhuis, H.-J. and De Bruijn, E.: *Building theories from case study research: the progressive case study.*
The 17th Annual Conference of Production and Operations Management (POMS), Boston, 2006.

[33] Van Der Marel, H. *DEOS Permanent GNSS Array (DPGA).*
http://gnss1.tudelft.nl/dpga/networks.html,

[34] Nederlandse Samenwerking Geodetische Infrastructuur: *Multi-year plan 2021 – 2025 for the Dutch Geodetic Infrastructure Cooperation.* In Dutch.
Infrastructuur, N.S.G., Editor, 2021,

[35] Van Hesteren, D. and Van Knippenberg, L.: *Open Data Maturity Report 2021*.
European Union, Luxembourg, 2021,
https://data.europa.eu/en/news/open-data-maturity-report-2021-out,

[36] Ploeger, H. and Van Loenen, B.: *Open NETPOS nader belicht: Concept rapport*. In Dutch.
Delft University of Technology, Delft, 2016,
https://research.tudelft.nl/en/publications/open-netpos-nader-belicht-concept-rapport,

[37] Huisman, L.: *GNSS RTK and DGNSS in the Netherlands*.
https://eurogeographics.org/wp-content/uploads/2018/04/17_2018PoskenNL.pdf,

[38] Kadaster: *Decide on further release of raw GNSS data in file form.* https://wetten.overheid.nl/BWBR0041329/2018-09-09,
[39] Huisman, L.; Hoentjen, K.; Vet, A. and Vet, P.: *Galileo-Only Cadastral Survey.* FIG Working Week, Amsterdam, 2020, https://www.fig.net/resources/proceedings/fig_proceedings/fig2020/papers/ts06g/TS06G_huisman_vet_et_al_10573.pdf,
[40] The Working Committee of the Surveying Authorities of the Laender of the Federal Republic of Germany (AdV): *National Report 2020/2021*.
[41] Zentrale Stelle SAPOS: *SAPOS® Reference station overview*. In German.
[42] The Working Committee of the Surveying Authorities of the Laender of the Federal Republic of Germany (AdV): *National Report 2016/2017*.
[43] The Working Committee of the Surveying Authorities of the Laender of the Federal Republic of Germany (AdV): *National Report 2013/2014*. 2014,
[44] Engfeldt, A.: *Network RTK in Northern and Central Europe.* LMV-Rapport, Gävle, 2005, https://www.lantmateriet.se/globalassets/kartor-och-geografisk-information/gps-och-geodetisk-matning/rapporter/lmv-rapport_2005_05.pdf,
[45] Working Committee of the Surveying Authorities of the States of the Federal Republic of Germany (AdV): *SAPOS® Precise Positioning in Location and Height*. Mapping, O.G.S.a., Editor. 2015,
[46] Bezirksregierung Köln. *SAPOS® Satellite positioning service.* In German. https://www.bezreg-koeln.nrw.de/brk_internet/geobasis/raumbezug/sapos/index.html,
[47] Riecken, J. and Becker, P.: *A new satellite positioning service for Germany*. In German. Geoinformation und Landmanagement **4** (145), 2020, http://dx.doi.org/10.15576/GLL/2019.4.145,
[48] Landesamt für Geoinformation und Landesvermessung Niedersachsen. *4G: Lower Saxony modernizes satellite positioning with SAPOS®.* In German. https://www.lgln.niedersachsen.de/startseite/wir_uber_uns_amp_organisation/presse_amp_broschuren/4g-niedersachsen-modernisiert-satellitenpositionierung-mit-sapos-r-204284.html,
[49] Behörde für Stadtentwicklung und Wohnen: *SAPOS data available free of charge from the beginning of 2022 Highly precise positioning data can be called up in real time.* In German. https://www.hamburg.de/bsw/landesbetrieb-geoinformation-und-vermessung/15715828/2021-12-17-bsw-lgv-sapos-daten,
[50] Riecken, J., *Enabling smarter farming in Germany with open data satellite positioning service, SAPOS®*, 2020.
[51] The Working Committee of the Surveying Authorities of the Laender of the Federal Republic of Germany (AdV): *National Report 2018/2019*.
[52] Jonsson, B.; Hedling, G. and Wiklund, P.: *Some experiences of Network-RTK in the SWEPOS™ network.* Lantmäteriet, Gävle, 2000, http://www.euref.eu/symposia/book2002/375-379.pdf,
[53] Sedell, D.: *Network-RTK a comparative study of service providers currently active in Sweden.* M.Sc. Thesis. School of Architechture and the Built Environment Royal Institute of Technology (KTH), 2015,
[54] Lantmäteriet: *A strategic plan for Lantmäteriet's geodetic activities 2011 – 2020*. https://www.lantmateriet.se/globalassets/kartor-och-geografisk-information/gps-och-geodetisk-matning/publikationer/geodesy_2010.pdf,
[55] Wiklund, P.: *SWEPOS Network-RTK service – establishment, status and experiences.* UN/Latvia Workshop, Riga, 2012,

[56] Ekström, D. and Johannesson, J.; *The value of open data-Socio-economic benefit analysis of making particularly valuable data available. In Swedish.*
Damvao Analytics, Lantmäteriet, 2020,
https://www.lantmateriet.se/contentassets/e16a59e08cb744149c878776256560e6/bilaga-2_-vardet-av-oppna-data---samhallsekonomisk-nyttoanalys-av-vardefulla-datamangder.pdf,

[57] Lantmäteriet: *National Report of Sweden to the EUREF 2018 Symposium – geodetic activities at Lantmäteriet.*
EUREF Symposium, Amsterdam, 2018,

[58] Norin, D.: *GNSS applications in Sweden and on-going developments of the national CORS network SWEPOS™.*
The Swedish mapping, c.a.l.r.a., Editor, 2014,

[59] Jensen, A.: *GNSS positioning services of the future, future GNSS users, requirements and national geodetic infrastructure*.
AJ Geomatics for Lantmäteriet, Roskilde, 2019,
https://www.lantmateriet.se/globalassets/kartor-och-geografisk-information/gps-och-geodetisk-matning/publikationer/gnss-positioning-services-of-the-future.pdf,

[60] Pashova, L. and Bandrova, T.: *A brief overview of current status of European spatial data infrastructures – relevant developments and perspectives for Bulgaria*.
Geo-spatial Information Science **20**(2), 97-108, 2017,
http://dx.doi.org/10.1080/10095020.2017.1323524,

[61] Kruse, D. and Roes, J.: *How to increase the usage of an NSDI from 1 billion hits to 10 billion hits in 5 years.*
FIG Working Week, Hanoi, 2019,

[62] Welle Donker, F.; Van Loenen, B. and Bregt, A.: *Open Data and Beyond.*
ISPRS International Journal of Geo-Information **5**(4), No. 48, 2016,
http://dx.doi.org/10.3390/ijgi5040048.

Interdisciplinary Description of Complex Systems 20(2), 96-111, 2022

IDENTIFYING AND OVERCOMING THE BARRIERS TOWARDS OPEN DATA OF PUBLIC UNDERTAKINGS*

Frida Boone[1] and Bastiaan van Loenen[2, **]

[1]Utrecht University
Utrecht, the Netherlands

[2]Delft University of Technology
Delft, the Netherlands

DOI: 10.7906/indecs.20.2.3
Regular article

Received: 24 January 2022.
Accepted: 16 March 2022.

ABSTRACT

The creation of open data has seen a series of waves in which every growing resources of data are becoming accessible to a growing number of users from a diversifying number of public entities. The European Commission anticipates this movement by setting a new scope to the re-use of Public Sector Information Directive. Instead of exclusively focussing on Public Sector Information), the new scope of its successor, the Open Data Directive, includes data from public undertakings as well. In order for public undertakings to comply with this future legislation, research into the current openness of public undertakings and the barriers to open data is key. This research presents three different levels of openness of data: (1) data is only open for the own organisation, (2) data is open for the internal organisation and trusted parties can use the data, and (3) open data for all. In this case the public undertakings are Port of Rotterdam and Schiphol Airport. The results showed that the data policy of Port of Rotterdam matches openness level 1. The data policy of Schiphol Airport matches level two. The open data policy of Dutch grid operator, Liander, corresponds with the third level as this organisation provides open data since 2014 for everyone. It can be stated that neither Port of Rotterdam nor Schiphol Airport is ready to comply with the future rules when the Open Data Directive requirements become mandatory. Barriers that are associated with achieving a higher level of open data are related to institutional, financial, legal, and quality and technical aspects. Overcoming these barriers requires, among other things, highly motivated staff to provide open data.

KEY WORDS

open data, public undertakings, open data directive, European Union, barriers

CLASSIFICATION

JEL: C89

*This is the extended version of the abstract published in: Vujić, M. and Šalamon, D., eds.: Book of abstracts of the National Open Data Conference. University of Zagreb, Faculty of Traffic and Transport Sciences, Zagreb, 2021.

**Corresponding author, η: b.vanloenen@tudelft.nl; ++31(0)152782554;
P.O. Box 5030, 2600 GA Delft, The Netherlands

INTRODUCTION

Globally, open data has played an important role in creating social and economic opportunities, solving public problems and empowering citizens to make better decisions [1]. An example of this is the United Kingdom, where heart surgeons of the National Health Service published comparable data on individual clinical outcomes in 2004. In 2011, improvements are reported; the survival rate increased by more than a third [2]. Another example is Nepal where open data regarding aid flows – expressed in geographical information – have contributed to building a transparent and accountable public institution after the civil war [3]. Likewise, within the European Union, open data is considered important for socio-economic developments of the society [4]. Recently, the lack of effective data use to address the COVID-19 virus shows that this important development still requires further work. In April 2020, 500 data practitioners and organisations over the world engaged in the ‘Call for Action’ by GovLab, a big data think tank, to develop an open data infrastructure which is capable of challenge the pandemic and other dynamic threats [1].

The majority of the data which is considered most valuable for tackling dynamic threats in the world is generated and held by the private sector – collected and controlled behind closed doors [1]. Interestingly, most global, regional and national efforts on opening data focus on open government data (or public sector information, PSI). It is expected that the value of open public sector information in Europe will increase from € 52 million in 2018 to € 194 billion in 2030 [5]. However, in order to answer pressing public questions on dynamic threats data publicly obtained needs to be open, central and incorporated into both public and private sector [1, 4]. The growing demand for open data is starting to have an influence on the open data policy of the European Union. The scope in the new open data directive is not limited anymore to public sector organisations, but was extended to other sectors.

In 2019, open data and the re-use of PSI was enacted in a new EU Directive, the Open Data Directive (ODD). The ODD provides a common legal framework for a European market for government-held data [4]. It builds on the Directives of 2003 and 2013, that focused on the re-use of records from public organisations, including national archives and libraries [6]. The new ODD also applies to documents held by public undertakings, research performing organisations and research funding organisations. These are non-government parties that collect, produce, reproduce and disseminate documents to provide services in the general interest [7]. Most often the data policies of public undertakings are restricted, not open data policies. The provisions of the new Directive are not yet mandatory for public undertakings. However, one may expect that new legislation will be more strict in the future. For the Netherlands to comply successfully with future legislation the challenge is to identify the barriers and means of tackling them for public undertakings to achieve an open data policy in the future.

In this article the following research question is central: “How can public undertakings in the Netherlands overcome the barriers to opening their datasets in order to be prepared for expected future legislation towards open data for public undertakings?”

We applied a mixed method research methodology. First, we conducted a comprehensive literature study on various concepts of open data and openness of data, barriers of open data, and the open data directive. This resulted in a first draft openness level model. Then we conducted interviews to review the open data status of three public undertakings and mapped their status on the openness level model and highlighted barriers to be overcome to move to the next level. In this last step, we used the experiences of a best practice open data public undertaking.

In this article we first explain open data and the open data directive, and present three levels of openness. The following section addresses the barriers one may have to tackle when moving from one level of openness to a next. Then, we assess the current level of openness of three Dutch public undertakings and explore the barriers they may experience when opening up their data ultimately by adhering to the requirements of the EU open data directive. The article concludes with the conclusions and recommendations for further research.

OPEN DATA (DIRECTIVE)

Open data is data that does not have any barriers in the (re)use. Open data aims to optimize access, sharing and (re-)using data from a technical, legal, financial, and intellectual perspective [8].

In the European Union, the Directive on the re-use of public sector information (PSI directive 2003/98/EC) was central to the stimulation of open government data. It was, however, only after two revisions that open data was introduced in the Directive on open data and the re-use of public sector information [7]. Re-use of documents held by public sector bodies should be provided in principle free of charge, and not be subject to any conditions in the re-use. High value datasets, documents associated with important socioeconomic benefits having a particular high value for the economy and society, shall be available free of charge, provided va APIs and as a bulk download.

The Directive's focus has been on documents of public sector bodies. However, the scope of the Directive has been extended from solely public sector bodies to educational and research establishments (schools, universities) and cultural establishments (libraries, museums and archives) in 2013 to public undertakings in 2019.

Public undertakings collect, produce, reproduce and disseminate documents to provide services in the general interest. At this moment, the Open data directive applies to public undertakings operating in the transport and utilities sectors only. Organisations operating in these sectors may decide themselves to release their data for re-use. For these data available for re-use a limited set of obligations is applicable, as compared to the general PSI regime. Public undertakings, for instance, can charge above marginal costs for dissemination, and are exempted from the general procedural rules on how to process requests for re-use.

However, the first PSI directive of 2003 exempted documents from educational and research establishments and cultural establishments explicitly, then brought the educational and research establishments and a major part of the cultural ones (libaries, museums, archives) under the scope of the first revision of the PSI Directive 2013/37/EU with very similar voluntary provisions. In the last revision, several of the voluntary provision were replaced by requirements: for example, documents of many educational and research establishments should be available for re-use and in principle be provided free of charge. A similar development can be foreseen for documents of public undertakings.

In order to assess the effort a public undertaking has to undertake to move from its existing data sharing policies towards full open data policies, we developed a framework identifying three levels of openness.

LEVELS OF OPENNESS

The definitions of open data from the literature review were used as an input for the creation of a multi-dimensional model on distinct levels of open data, Figure 1. Three levels of open data were identified: (1) only open for internal use, (2) partly open for external users, and (3) fully open data. To specify the requirements of the three levels, we used the sub-categories of find, play and share from [9]. Find and play are associated with how the data can be found and used, whereas share is associated with the person using the data and how the data can be shared. [9] claims that once data is found and used, it should be possible to share it with others [9]. However, when considering openness level three, 'sharing' was replaced by re-use since sharing does not imply that the data can be re-used by all, which is a requirement of the Open data directive [7].

At the first level data is considered not to be open at all and only accessible for the internal user. Here, the data cannot be found through a general search engine [10]. This makes the data invisible to everyone but the internal user. The absence of an open licence makes it impossible to share the data with external users [9, 10, 12, 13]. This suits an internal regime that is focussed on using the data for internal purposes, limiting the data quality to the purpose of the internal user [14].

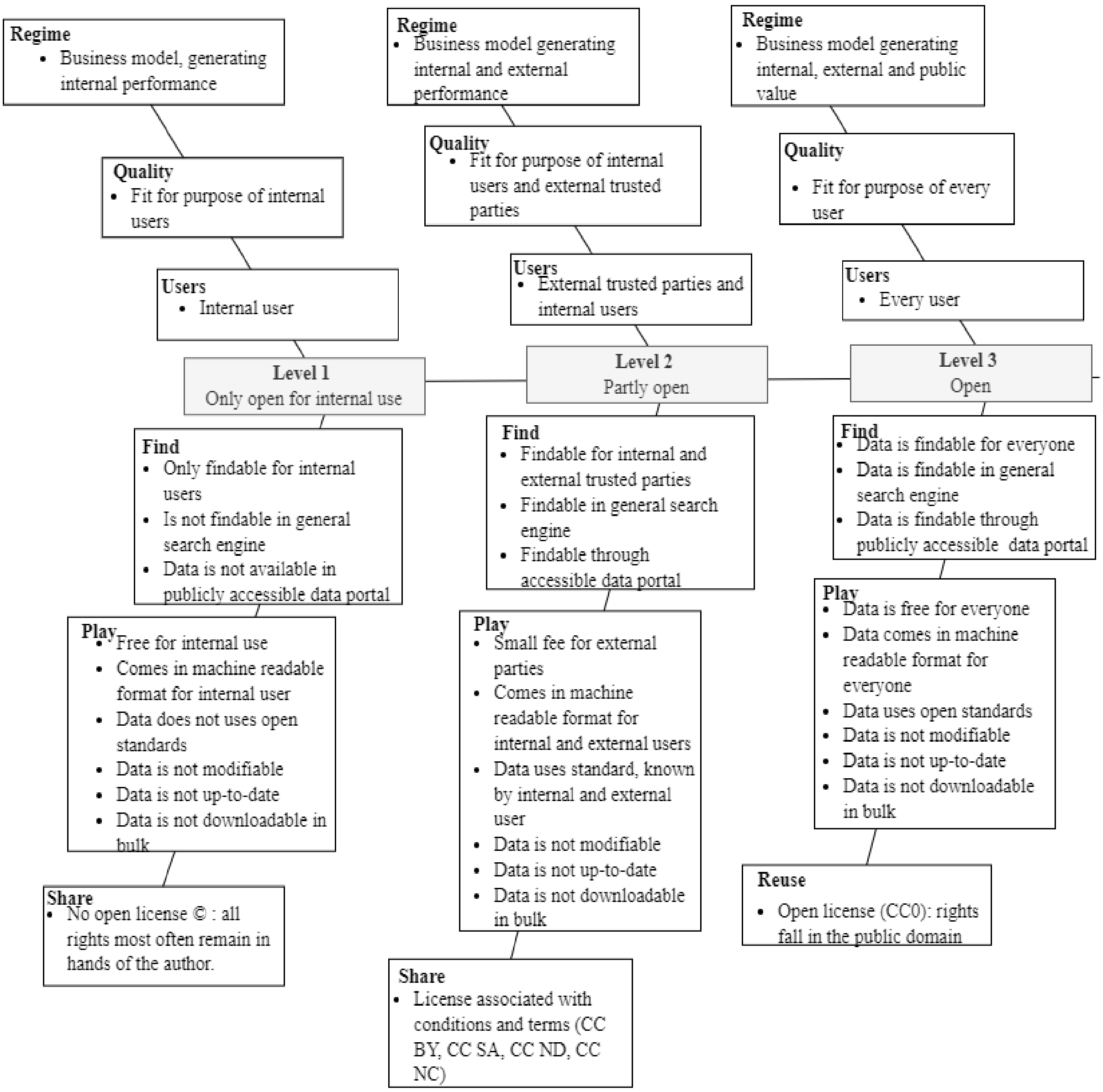


Figure 1. The multi-dimensional model of the three identified levels of open data.

In the second level of openness, partly open for external users, the data is under strict conditions available to external parties: the metadata of the data is published in publicly accessible data portal and/ or search engines, in a machine-readable format [10]. However, fees may be charged and the data can only be shared under certain conditions and terms. This data policy generates both internal and external value.

In the most open level, open data, the data is adhering to the most fundamental principles of open data: free of charge, no conditions in the re-use, data is downloadable in bulk, adhering to open standards, among others. The data is findable through a general search engine and data portal, free of charge, comes in a machine-readable format and with an open licence so that everyone can re-use the data [7, 14, 15]. At this level, internal-, external-, as well as public value are generated. This third level is most closely following the requirements of the Open Data Directive.

MOVING FROM ONE LEVEL OF OPENNESS TOWARDS ANOTHER: ADDRESSING OPEN DATA BARRIERS

While open data can contribute to social and economic benefits, moving from a lower level of openness to full open data will encounter numerous barriers [1, 16]. To achieve open data these barriers need to be identified and overcome. These barriers can be perceived from either the provider's perspective or the user's perspective [10, 11, 16]. We identified three types of barriers: (1) institutional barriers, (3) task complexity barriers, and (3) technical quality barriers.

INSTITUTIONAL BARRIERS

Any unwillingness from data providers in terms of financial and legal risk to make data open available, is known as an institutional barrier [16]. Institutional risks like this make organisations cautious when providing data [17]. Such a risk-averse culture results in organisations preferring not take any risk to change [16, 18].

Perceived financial risks can be divided in two categories: (1) fears for budget deficits due to the loss of income when a cost recovery policy needs to be replaced by an open data policy, and (2) the expected extra costs related to additional human and financial resources both to collect, to maintain, to process the data and finally to distribute it as open data [19]: adaptation costs, infrastructural costs and structural maintenance / operational costs [11, 19].

Legal risks of open data are manifold: higher liability risk due to errors in the data, misuse of the data, disclosure of secured information, such as trade secrets can put an organisation at risk [20], violation of data protection stated by the General Data Protection Regulation (GDPR) or privacy legislation, breaches of existing contracts, and/or open government data unfairly competing with similar datasets sold by a company. An example of unexpected increased liability risk was in Pacific Gas & Electricity (PG&E), an American utility company that published their data without any restrictions toward the use of the data. After a spatial analysis, done with open data from the company on the electricity poles, PG&E were held liable for the cause of the largest and most destructive wildfires in state history. The study showed that the locations of the fires were often in the proximity of the electricity poles from PG&E (energy data request from public datasets from PG&E). Their equipment of electric powerlines across the state evoked sparks that caused wild-fires which took the life of 84 people in 2018 [21]. In 2020, PG&E pleaded guilty and agreed to pay a maximum fine of 25.5 billion dollar for losses from the 2018 wild fire, blamed on the crumbling equipment of PG&E [22]. On the one hand it can be stated that open data is used correctly in this case by directing to the cause of the wildfires in California in 2018. On the other hand, this example highlights that, from a data providers perspective, there are risks associated with open data.

TASK COMPLEXITY BARRIERS

Finding and using data tends to be challenging and often complex for the data user, due to high complexities. These complexities are worsened when there is no explanation of the context of the data or when the data formats and datasets are too complex to handle [11, 16, 18]. For example, complexity becomes a barrier in geographical datasets for an unexperienced user when attempts are made to open an AutoCAD drawing (a detailed 2D or 3D illustration) in a geographical information system (GIS, ArcGIS pro for example). Matching of data formats with information systems can become more challenging and require more user knowledge/experience to manipulate the data.

Therefore, use of data is considered only for those with domain knowledge which allow for opening, using and interpreting the data [16]. So data can only be accessed and used by a user who has the technical skills to download the data, open the data in a GIS and analyse the data through tools. The format and complexity of data may contribute to a digital divide, a barrier, as the use of data might be limited to certain groups; only those with domain knowledge [16]. User skills is a potential barrier that can be tackled by improved data format, structure and utility.

TECHNICAL QUALITY BARRIERS

In order for the data to ensure a valuable return on both user and provider side, the data needs to be fit for use [14, 23]. Because every user may have a different purpose when using data, a guarantee of quality cannot be given [16, 18]. An accuracy check on the data needs to be done before the data can be used for a certain purpose. Such a check can be accomplished through contact with the data creator and by enquiring about the correctness of the data in terms of the completeness of the metadata [16]. Often this is not possible as contact information, if present at all, does not trace back to the actual data creator [11, 16]. Even when the metadata is present sufficient data quality is not guaranteed as there is no single standard for metadata for all usersresulting in heterogeneity of metadata models and different vocabularies [11]. At worst, this could limit or prevent the user from reusing the data [16]. The absence of agreed quality standards, possible lack of a supporting infrastructure (data portal), as well as fragmentation of manipulation software and applications can present technical barriers to data openness.

SUMMARY OF BARRIERS & LEVELS OF OPENNESS

The barriers together with the different levels of open data identified so far are presented in Figure 2. Firstly, it shows the organisational barrier that affect the attainability of the data, which is addressed by the data provider in terms of regime, quality of the data and the type of user (the upper part of the model) [24]. A regime may face institutional, financial and legislative barriers when steps towards an open data policy are made [2-27]. Creating an open data regime requires willingness of the data provider to do so and this includes finding financial funds and applying licenses that allow the user to share and re-use the data [16, 17, 25]. In order to create more openness through improved quality of the data, improving attainability and usability, financial and technical barriers need to be tackled. To modify the quality of the data for external and public users technical skills are required as well as financial resources to adjust the quality to the purpose of external and public users [11]. Legal barriers may be faced when changes in licence are required enabling the sharing of data with external trusted parties whether or not under conditions. This is due to the fact that external parties also have the rights to access and modify the data through a new licence. Therefore, new legal barriers are faced for the data provider to limit data misuse and data fragmentation, which might be caused by external parties as a result of more rights. When legislation does not prevent the re-use of the data for every user anymore, there are liability risks for the data provider when the step towards level three is taken. These risks can be expressed in financial, actual and/or reputational damage from false conclusions drawn from the data by the users, or from publishing private and secure data [28]. Financial barriers are encountered when making the data findable and accessible through search engines and/or data portals for external users. Barriers associated with task complexity are faced when the users shift from being external trusted parties to public users as the data user is unknown to the data provider in level three. The domain knowledge of the user is difficult to assess which makes it is difficult for the data provider to know whether the published data suits the knowledge domain of all the user [16].

As a result of the barriers perceived by the data provider, the ability for the data user to find, play with, and share or re-use the data can decrease. In order to make the data more findable for users other than the internal users, the data providing organisation faces financial and task complexity barriers. The same barriers are faced when it comes down to play. Financial investment by the data provider is required to create the possibility for the user freely to use and modify the data. The additional barrier of technical quality is faced by play since the published data need to be recent, in a machine-readable format and possible to be downloaded in bulk. The barriers faced by the requirements of share/re-use are associated with the application of different types of licenses as this decides whether and under which conditions the data can be shared and re-used. The attainability of the data for the user is determined by the data provider [24, 28-31]. Figure 2 shows the possible barriers between levels of openness, based on the literature review.

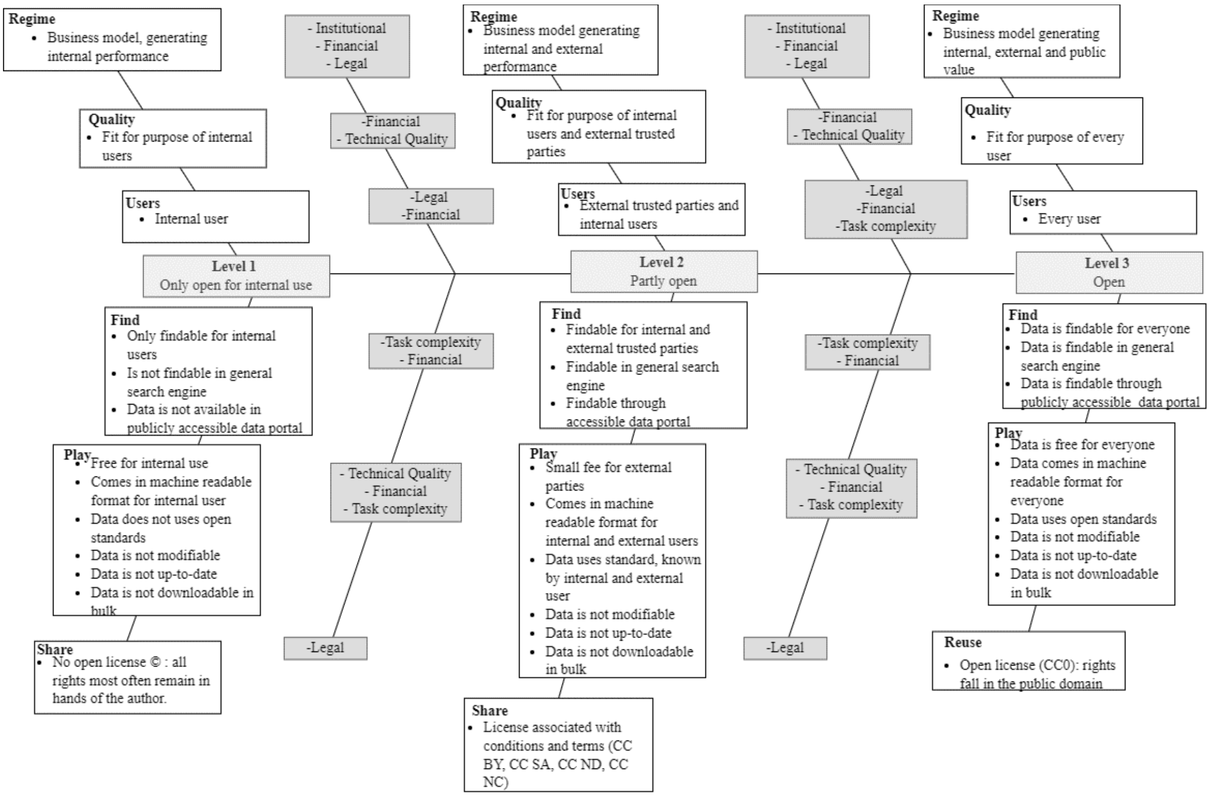


Figure 2. A multi-dimensional model of the three identified levels of open data, including the identified barriers between the levels which are faced when a transfer to a higher level is intended.

OPEN DATA IN THREE PUBLIC UNDERTAKINGS

In this section we show the results of a case study in three Dutch public undertakings. We performed interviews with one Analytics Specialist, two Data Stewards, a GEO-IT solution architect, an Enterprise Data Architect and a Product Developer. We reviewed the level of openness in these organisations, and the perceived barriers to move to a next level of openness.

PORT OF ROTTERDAM

Port of Rotterdam (PoR) is the biggest sea harbour of Europe, situated in the Harbour of Rotterdam. The harbour has deep-sea connections with more than a thousand harbours around the world. The Port Authority has an important role in developing, organising and managing the logistic activities in the Harbour. PoR's shares are held by the Municipality of Rotterdam (70 %) and the Dutch government (30 %). The shares are not listed on stock exchange which makes PoR an unlisted public limited company.

Regarding the different open data levels, Port of Rotterdam can be placed in the situation prior to level 1. Although data is shared with internal users, it is not yet shared with all internal users. Data is collected within departments and typically not shared with others within PoR. Data is only shared with third parties when this is in the interest of PoR's business activities. Until now, data has never been shared with citizens exclusively to generate public value. Awareness of the value of sharing data is growing within the company. This has resulted in 12 data domains which should create an overview of the data that is used by the departments and the impact it has. When this task is completed, a next step will be to create more openness towards third parties to generate both internal and external performance. Sharing data with citizens for the sake of generating public value on its own is not yet on the horizon. This next step will be towards level two of the different open data levels, dealing with, in order of significance to PoR, technical and institutional (including legal) barriers, Figure 3. Legal barriers are not considered to be the biggest issue since PoR controls the conditions and terms that can be determined in the data delivery agreement. Technical issues, however, are considered difficult barriers to deal with since a new technical department needs to be developed to make the data more findable through a portal for third parties (between 'find' and 'play' in Figure 3). The quality needs to be fit for the purpose of third parties which requires additional investments of PoR. The willingness to share data with external parties is growing within the company, but still is not for granted, placing the institutional barrier not on the top of the list. The drive to share data is there but the next step is to find the most suitable technical and financial solution for it. As yet, level three, where data sharing is replaced by data re-use and the user is identified as everyone, is a step too far away for Port of Rotterdam.

SCHIPHOL

Amsterdam Schiphol Airport is the largest airport in the Netherlands and plays an important economic and social role in Europe. It is considered one of the most connected airports in the world and facilitates 332 international connections. Regional airports, international alliances and cooperation enhance this international connection. Schiphol is held by the Royal Schiphol Group, with the Dutch government, the municipality of Amsterdam and Rotterdam and Groupe ADP (an airport operator) as stakeholders.

Schiphol can be placed in level two of the multi-dimensional open data model. The goal of their data governance is to share data with internal, external and public users. Sharing data with the public user is, however, only executed when there is no interference with the commercial interests of Schiphol. Sharing data with external trusted parties is possible and comes with an agreement that covers liability issues regarding misuse of the data. For the public user the available datasets are presented on the open data portal of Schiphol. Even though the data is

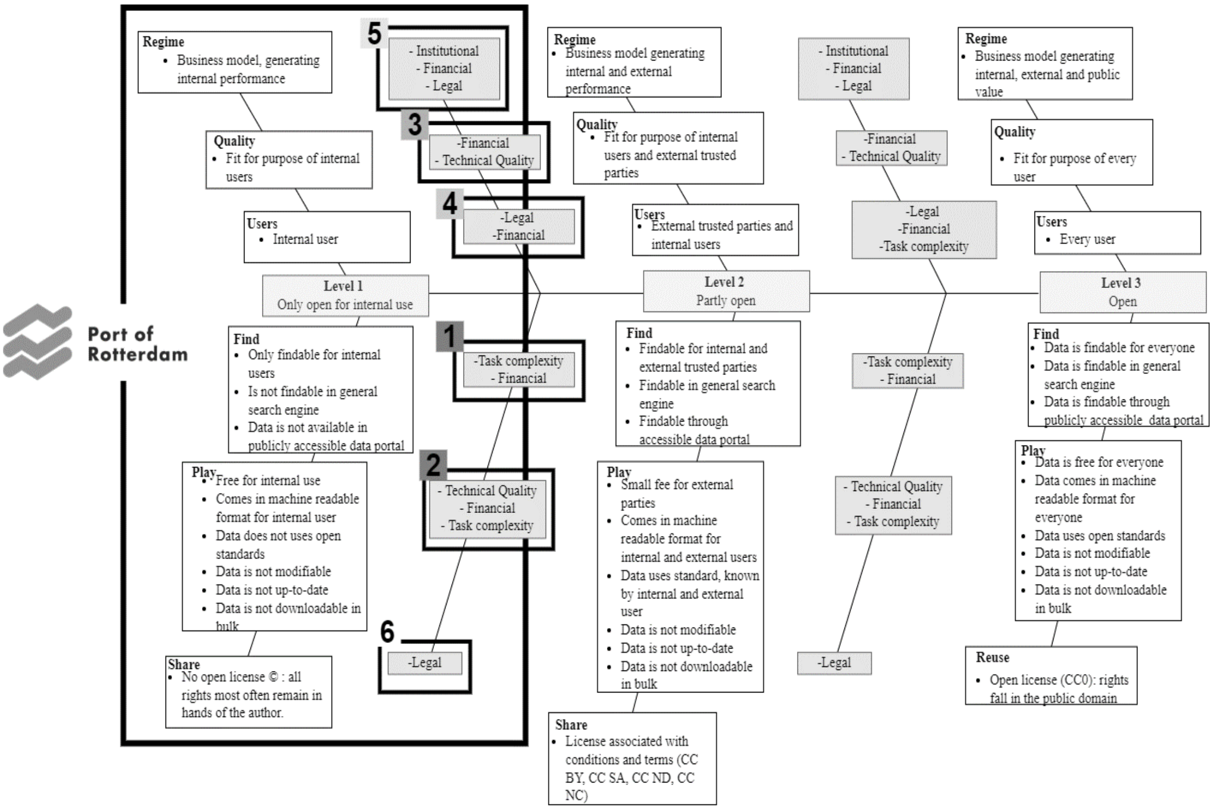


Figure 3. The level at which Port of Rotterdam can be placed in (level 1) and the barriers which are faced, numbered from most significant (1) to least significant (6).

available for the external and public user, it is not directly accessible and so it cannot be 're-used'. To control the data used by the public user a data request – subsequently a registration – is needed from the user. After a request, the data is provided to the user with less meta data and a lower level of detail than the source data, controlling the sensitivity and amount of data provided to external users.

Perceived barriers that Schiphol associates with the next step towards open data concern security and privacy barriers, confidentiality barriers and institutional barriers (see Figure 4). The interviewees highlighted that the main issue that causes the privacy and security barriers is the level of detail of the data. This applies first of all to the security barrier. The available data reveals too much detail, such as the location of the armoury, that could assist a terrorist attack. Secondly, too much detail can reveal private data about individuals at Schiphol which places these data under the scope of the GDPR, which does not allow for open data (level three in the model). Confidential agreements with third parties cause the third legal barrier. Schiphol cannot share data that is retrieved from third parties if re-use is only allowed by internal users of Schiphol; this data cannot be shared with others. Lastly, due to fear of false conclusions drawn from the open data of Schiphol, not all data is made openly available. This is an institutional barrier. Both interviewees state that Schiphol has already experienced reputational damage as a result of false conclusions drawn by users and, as a result of that, they are not willing to adapt to a fully open data regime. However, it could be argued that publishing open data could prevent reputational damage. By publishing open data, Schiphol creates the opportunity to provide good and correct data, which can prevent the risk of false conclusion drawn by the user. So instead of fearing open data, it could also be considered a solution.

In contrast to the Port of Rotterdam, financial, technical, and quality issues are not considered to be the main causes for the barriers faced by Schiphol. These barriers are listed as numbers four, five and six – associated with 'quality', 'find' and 'play', in Figure 4. Financial issues due to development and maintenance costs of open data are not considered since costs for developing and distributing data for public use are already made and not considered a great issue. A technical quality barrier will also not be the main problem since Schiphol already succeeded in setting up a data portal for the users (developer.schiphol.nl). Technical quality barriers are not faced in the sense that modification of the data for public use is not possible; it is possible but does take some time and effort.

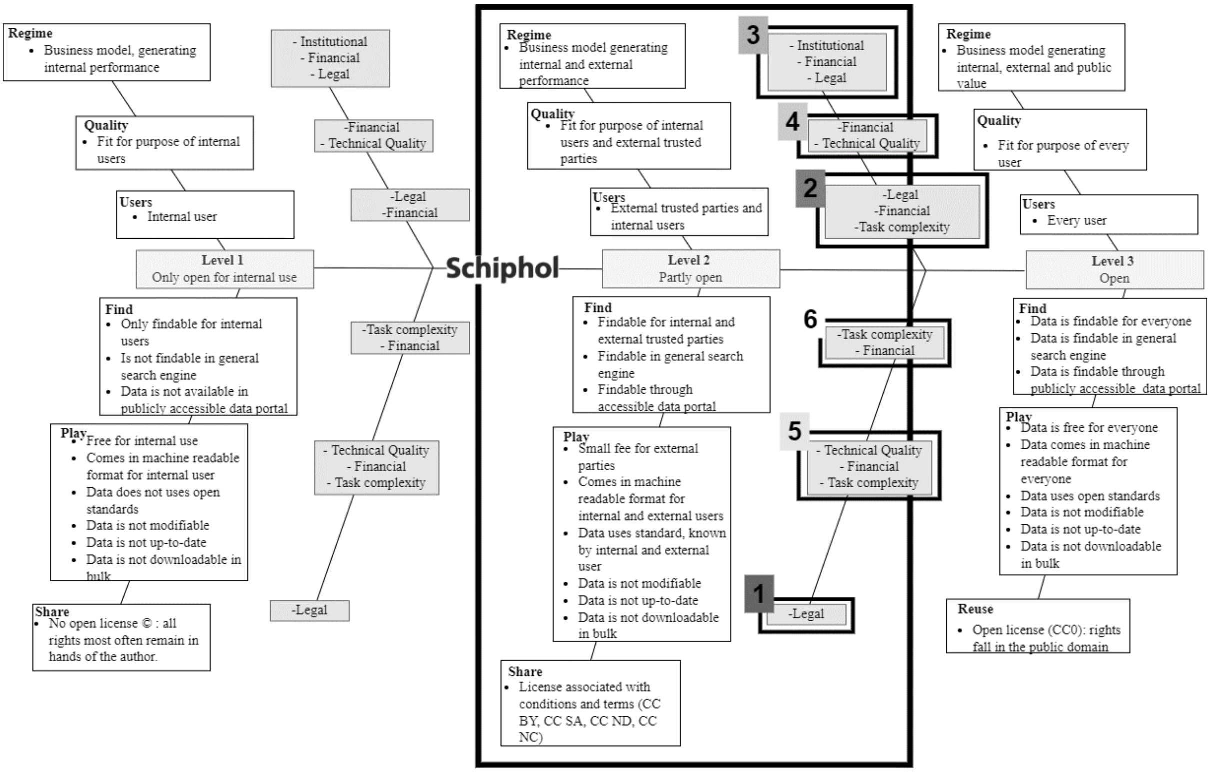


Figure 4. The level at which Schiphol can be placed in (level two). The barriers which are faced are numbered from most significant (1) to least significant (6).

LIANDER

At present, for most organisations it is clear what open data is, and which social and economic values it can offer [4]. However, 'open data' is not a sort of package that organisations can buy in a shop that comes with instructions on how to apply it. This raises the question how organisations can best share their data or even provide it as open data. Since data sharing is relatively new, hard facts and figures are not yet available to indicate the best way of data sharing. It is often best to learn from the success of other organisations by learning how they overcame the barriers to open data. This was acknowledged by the European data organisations who focussed on data sharing for both governments as well as private organisations (Data.overheid.nl, user meeting, 2020).

One public undertaking that has successfully opened their data is Liander, which is a Dutch utility company and subsidiary of Alliander, which develops and manages energy networks. Liander is allowed to operate as an independent grid operator, however, the activities cannot be in conflict with the national grid management. The company is providing open data since 2014 and define open data as digital data that is made available for everyone through the internet [32].

The motivation to provide open data needs to be very clear. Liander wanted to contribute to a better collaboration with the regions within their area of operation [33]. Furthermore, it was motivated by the social benefits open data brings to the society. Liander often received individual questions regarding the utility usage of their network and the bottlenecks within their network. These requests needed to be answered one by one which resulted in a time consuming activity for Liander. By publishing their most sought-after data they wanted to create more opportunities for the users (such as municipalities) to work with their information without having to ask Liander for input every time. With data readily accessible, the user can work with Liander's information and substantiate their own plans with meaningful data. In turn, such plans could benefit to Liander's network. This potential societal benefit was the main motivation for Liander. Saving time on individual data questions was also considered a motivation to continue to provide open data. The other motivation arose from the fact that Liander wants to contribute to energy transition. Providing open data on for example consumption per year, per type of house can give an understanding on the best possible manner to lower the energy consumption. Since Liander mainly provides their own collected data and not data from external parties, they do not face the same issue of sharing confidential data from clients as Port of Rotterdam and Schiphol.

The first question that needs to be answered is whether sharing of data is allowed under the law. Similar to Port of Rotterdam and Schiphol, the fear of liability issues was present. Although Liander has experienced significant objection to data sharing from different Dutch legal authorities, liability issues were never experienced. The disclaimer used for the open data (Creative Commons BY) may be an explanation for this.

Similar to Schiphol Airport, the fear of terrorism also affects Liander's data. In order to get approval for such issues was discussed with the Dutch General Intelligence and Security Service (AIVD). After years of discussion, the AIVD decided which of Liander's data could be published and which data could not be published. The AIVD imposed aggregation of the level of detail that was publicly available. The AIVD decided that the electricity cable network could be published, whereas the location of the gas pipes was considered too sensitive to publish in terms of explosion risk. Although it took years for Liander to satisfy the requirements of AIVD, the willingness to provide open data never gave way to fear for legal or terrorism issues according to the interviewee.

Another legal barrier was in potential unfair competition. In the Netherlands, the ACM (consumer association authority) ensures a fair balance between companies and protects consumer interest. Initially, the ACM considered the E-Atlas of Liander as a distortion of competition: unfair competition. The fear was that other companies would be disadvantaged in their business if Liander put a similar business to the market, financed by public funds. In practice, this was not the case as other businesses were not allowed to access this source data on electricity and gas usage due to data protection legislation. Due to market barriers, not related to data, other companies could not start a similar business. In this case unfair competition could not result from data issues.

Technical barriers were not an issue for Liander when setting up an open data portal. "Setting up the open data portal is done by internal employees so no extra, external costs are made". Moreover, the data portal was developed by the internal employees to reduce the time spent on previous data requests. Therefore, the internal time spent on the development is an investment to win time in the future. The internal expenses were estimated at 0,5 FTE. Ignoring opportunity costs and since these costs were made already, no additional financial issues were faced by Liander.

Feedback on the quality of the data is actively encouraged by Liander as it gives information on the quality requirements of the user. Moreover, the feedback given by the users can be used to improve the quality of the data so that other users will not face the same issue.

The case of Liander proves that consistent determination to provide open data is key to achieving it. Liander faced mainly legal barriers associated with the level of detail of the data they could provide. The initial level of detail of the data interfered with both the guidelines of the AIVD and the GDPR. Aggregation of the data was key for the organisation to ensure open data without breaking the legal protection and privacy guidelines. In their action plan towards open data they dealt with legal, technical and quality issues that were challenged with data aggregation and legal discussions. By opening up Liander's data, the company experienced benefits in time and money saved on individual data request. Providing open data contributed to the energy transition because Liander's data informed on possible manners to lower energy consumption. Liander considered providing of open data a social benefit and which was a key motivating point. The next chapter assesses whether Port of Rotterdam and Schiphol could apply this working method as well in order to achieve open data.

DISCUSSION

The Liander experience indicates that aggregation is an important tool to achieve open data that complies with legal requirements in terms of privacy, security and confidentiality of data. A discussion point could be whether open data can still be achieved without the option of aggregation. For PoR this could apply to the less sensitive datasets classified as 'public' instead of 'internal' or 'confidential'. This could, for instance, be the case for the datasets on traffic signs as often used by PoR for internal purposes. Within this dataset, all attributes such as model number, location, and year of placement are classified as public; no aggregation is needed to provide this data as open data. Other datasets which seemingly do not hold confidential data, such as road networks, may prove otherwise. The dataset on the road network holds several attributes which are classified as public such as the road type, function, length, and hardening layer, but this dataset also holds confidential information, such as inspection results and level of ambition (the desired maintenance level of the asset). For Schiphol, the same could be considered. Is it possible to open up datasets which presumably do not hold sensitive data? Of all the datasets on which Schiphol is currently working it may be possible for the dataset on flying birds. The approach could be that this information has no potential provoke terrorism or breach the requirements of the EU General Data Protection Regulation. This approach could be applied to more datasets than the flying birds. Moreover, without the possibility of aggregation it might be possible for both companies to provide open data through the CC-BY licence, as done by Liander. Since this licence allows re-users to distribute adapt, and build upon the data if attribution is given to the creator, it could remain a sense of control for the companies. This way, it becomes clear for what purposes the data is used and by whom.

The lessons from Liander show that at the heart of open data sits an open data mindset; a fundamental belief in the concept that openness of data is desirable and a service to the common good. PoR and Schiphol, as mostly publicly owned organisations, could reasonable be expected to adopt a more socially responsible approach to their data. Liander's experience showed a more holistic approach to data by considering the cost of responding to questions in combination with the cost of data openness. They showed that the added expense was limited and outweighed by commercial as well as social benefits.

Liander also showed that careful management of accessibility of the data, for instance by aggregating, mitigates the risk to reputation and could be off-set by the benefit of being regarded as a transparent, accountable and socially responsible organisation. As a grid operator, Liander is a ready target for potential criticism, for instance on climate impact. PoR and Schiphol, being less open with their data, but still significant potential targets for criticism, could benefit a more pro-active approach to open data. A pro-active approach to open data could anticipate such accusations and potential reputational damage. Hence, they could point at the readily available data; this is often sufficient to deflect more detailed investigation. For example, for Schiphol this could be done by providing correct data on road networks to avoid false conclusions drawn by the users, subsequently the media. The cost effort in responding to media or legal challenges, both in direct financial terms but also in reputational terms, should be included in the equation when considering data openness.

Finally, the pressure to open up data in public and semi-public organisations will continue to accumulate. Both PoR and Schiphol would recognise the unavoidability of moving toward open data. Early recognition of the inevitability investment requirement would still give them an opportunity to plan, schedule, implement and finance their open data programme at their own tempo. Once regulation overtakes their effort, the tempo will be set from outside and may be less optimal.

CONCLUSION

In this research the following research question was addressed: "How can public undertakings in the Netherlands overcome the barriers to opening their datasets in order to be prepared for expected future legislation towards open data for public undertakings?"

It can be stated that public undertakings, such as Port of Rotterdam (PoR) and Schiphol Airport, can overcome barriers towards open data to be prepared for the foreseen future legislation of the Open Data Directive. However, to do so changes need to be implemented. The multi-dimensional model in this research identified three different levels of open data for a public undertaking to reference it's data policy: (1) only open for internal use, (2) partly open for external users, and (3) fully open data. In this model the requirements of open data are interpreted from the data provider's perspective in order to make the data more open for the end-user. At the first level data is only accessible for the internal user, using the data for internal performance; such data cannot be found through a general search engine. At the second level data openness is improved as it is findable and accessible through a general search engine or data portal, available for the external data user as well as to the internal data user. Data is used for generating internal and external performance. At the third level data can be considered most open. The data is findable for the internal, external and public user, through a general search engine and data portal, free of charge and with an open licence for everyone to re-use the data.

Neither PoR nor Schiphol are ready to comply with the future rules when the Open Data Directive requirements become mandatory. Barriers still need to be overcome, but Liander has shown that this can be achieved with prolonged leadership. PoR is placed in level one of the model as the collected data is used by and shared with the internal user mostly. Schiphol can be placed in level two since it shares data with internal, external and public users. Sharing data with the public user is, as yet, only executed when there is no interference with the Schiphol's commercial interests. The main goal of both public undertakings is to generate internal performance with their collected data. In between the levels, barriers are identified which are faced when a higher level is pursued. Identified barriers to be overcome are financial, institutional, task complexity, legal, and technical. For Liander, that provides open data since 2014, similar barriers were encountered and defeated on their path to open data. The Liander case shows that achieving open data starts with the institutional motivation to do so. A commitment to open data must stem from the top level in the corporate body to gain sufficient traction.

RECOMMENDATIONS

For future research it is recommended to take this research as a motive and reset the scope to the outcomes of this research. An interesting feature presented in the results was the use of aggregation by Liander. Aggregation was considered to be the key method to use for the achievement of open data in terms of legal requirements concerning security privacy and confidentiality. One proposed action would be to focus on the level of aggregation, suitable for the current data policy of both PoR and Schiphol. The question to consider would be: to what extent can the level of the datasets be aggregated and still contribute to the internal performance of the companies? This question interprets the level of detail from the data provider. However, the same question could be asked from the perspective of the users: how valuable is aggregated data for users?

Another recommendation derives from the action plan used by Liander to achieve open data. The different legal and technical steps taken in this action plan could also be taken by PoR and Schiphol. Liander’s action plan helped the company to map the different steps and actions needed to achieve open data; it is recommended to set up a similar action plan for PoR and Schiphol. Future research could develop a similar and suitable action plan for PoR and Schiphol that gives insights in the detailed actions needed to achieve open data for these public undertakings.

A final future research question is associated with the three levels of openness. The aim would be to reach the third level of openness since then an organisation meets the requirements of the Open Data Directive. However, for organisations in beginning stages of open data, is it possible to simply ignore the second stage and leapfrog directly to the most open level? Or should organisations first experience an intermediate level of openness in order to be fully equipped and ready for the final stage?

ACKNOWLEDGEMENTS

This article is based on the research findings of the Master of Science thesis “Third wave open data in the Netherlands: Identifying and overcoming the barriers towards open geographical data of public undertakings”, MSc. Geographical Information Management and Applications (GIMA).

This research is part of TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592.

REFERENCES

[1] Verhulst, S.G., et al.: *The Emergence of a Third Wave of Open Data How To Accelerate the Re-Use of Data for Public Interest Purposes While Ensuring Data Rights and Community Flourishing.* http://dx.doi.org/10.2139/ssrn.3937638,

[2] Rogers, K.: *Open data reduces mortality rate in UK hospitals.* http://opendatahandbook.org/value-stories/en/uk-mortality,

[3] Open Data Institute.: *Making aid more effective in Nepal. Open Data handbook.* http://opendatahandbook.org/value-stories/en/effective-aid-in-nepal/,

[4] European Commission: *From the Public Sector Information (PSI) Directive to the open data Directive.* https://ec.europa.eu/digital-single-market/en/public-sector-information-psi-directive-open-data-directive,

[5] Barbero, M., et al.: *Study to support the review of Directive 2003/98/EC on the re-use of public sector information.*
http://dx.doi.org/10.2759/373622,

[6] The European parliament and the council of the European Union: *Directive 2013/37/EU of the European parliament and of the council.*
https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32013L0037&from=EN,

[7] The European parliament and the council of the European Union: *Directive (EU) 2019/1024 of the European parliament and of the council.*
https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32019L1024&from=EN,

[8] Van Loenen, B.; Vancauwenberghe, G. and Crompvoets, J., eds.: *Open Data Exposed.*
Information Technology and Laws Series **30**(1), 2018,
http://dx.doi.org/10.1007/978-94-6265-261-3,

[9] Eaves, D.: *Three Laws of Open Data (International Edition).*
https://eaves.ca/2009/11/29/three-laws-of-open-data-international-edition/,

[10] Welle Donker, F. and van Loenen, B.: *How to assess the success of the open data ecosystem?*
International Journal of Digital Earth **10**(3), 284-306, 2016,
http://dx.doi.org/10.1080/17538947.2016.1224938,

[11] Martin, S., et al.:*Risk analysis to overcome barriers to open data.*
Electronic Journal of E-Government **11**(1), 348-359, 2013,

[12] The Organisation for Economic Co-operation and Development: *Open Government Data.*
https://www.oecd.org/gov/digital-government/open-government-data.htm,

[13] World Wide Web Foundation.: *The Open Data Barometer.*
https://opendatabarometer.org/barometer/,

[14] International Organisation for Standardization: *ISO 9001:2015 Quality management systems – Requirements.*
https://www.iso.org/standard/62085.html,

[15] Berners-Lee, T.: *Linked Data.*
https://www.w3.org/DesignIssues/LinkedData.html,

[16] Janssen, M.; Charalabidis, Y. and Zuiderwijk, A.: *Benefits, Adoption Barriers and Myths of Open Data and Open Government.*
Information Systems Management **29**(4), 258-268, 2012,
http://dx.doi.org/10.1080/10580530.2012.716740,

[17] Van Loenen, B., et al.: *Databeleid Rijkswaterstaat: Deel: Hoofdwatersysteem; een overzicht van de juridische kaders omtrent het omgaan met data.*
http://repository.tudelft.nl/view/ir/uuid:540a04e6-d4ee-4d78-a526-7d6a8300009b,

[18] Barry, E. and Bannister, F.: *Barriers to open data release: A view from the top.*
Information Polity **19**(1-2), 129-152, 2014,
http://dx.doi.org/10.3233/IP-140327,

[19] Welle Donker, F.: *Funding Open Data.*
Information Technology and Law Series **30**, 55-78, 2018,
http://dx.doi.org/10.1007/978-94-6265-261-3_4,

[20] Tankard, C.: *Big data security*.
Network Security **2012**(7), 5-8, 2012,
http://dx.doi.org/10.1016/S1353-4858(12)70063-6,

[21] Kasler, D.: *PG&E says it will begin shutting down power lines when fire risk is extreme*.
https://www.sacbee.com/article206369044.html,

[22] CNBC: *PG&E pleads guilty to 84 deaths in 2018 California wildfire.*
https://www.cnbc.com/2020/06/16/pge-to-plead-guilty-to-deaths-from-california-wildfire.html,

[23] International Organisation for Standardization: *ISO 19157:2013 Geographic information – Data quality*.
https://www.iso.org/standard/32575.html,

[24] Welle Donker, F, Braggaar, R.C. and Van Loenen, B.: *Hergebruikers van open data in beeld.*
https://kennisopenbaarbestuur.nl/media/256297/hergebruikers-van-open-data-in-beeld.pdf,
[25] Afuah, A. and Tucci, C. L.: *Internet Business Models and Strategies: Text and Cases*. McGraw, Hill, 2001,
[26] Janssen, M. and Kuk, G.: E-Government business Models for Public service networks. In: Anttiroiko, A., ed.: *Electronic Government: Concepts, Methodologies, Tools, and Applications.* IGI Global, pp.898-916, 2008,
http://dx.doi.org/10.4018/978-1-59904-947-2.ch071,
[27] Zuiderwijk, A., and Janssen, M.: *Open data policies, their implementation and impact: A framework for comparison.*
Government Information Quarterly **31**(1), 17-29, 2014,
http://dx.doi.org/10.1016/j.giq.2013.04.003,
[28] Creative Commons.: *About CC Licenses - Creative Commons.*
https://creativecommons.org/about/cclicenses,
[29] Miller, P.; Styles, R. and Heath, T.: *Open data commons, a licence for open data.*
http://events.linkeddata.org/ldow2008/papers/08-miller-styles-open-data-commons.pdf,
[30] Kalshoven, L.; Schep, T.; Zeinstra, M. and Brinkerink, M.: *Open Data Reader 2016.*
Nederlands Instituut voor Beeld en Geluid, 2016,
https://www.kl.nl/wp-content/uploads/2016/04/Open-Data-Reader-2016-1.pdf,
[31] Open Knowledge Foundation.: *Open Data Commons: legal tools for open data.*
https://opendatacommons.org,
[32] Juffermans, P.: *Open data Liander.*
https://www.slideshare.net/lisettevanbeusekom/open-data-liander-paul-juffermans,
[33] Van Loenen, B. And Bregt, A.K.: *Open data and beyond: Deelrapport: Instrumenten voor de monitoring van de impact van open data (bij Alliander).*
https://research.tudelft.nl/en/publications/open-data-and-beyond-exploring-existing-open-data-projects-to-pre-2.

Interdisciplinary Description of Complex Systems 20(2), 112-124, 2022

FIREFLY OCCURRENCES IN CROATIA – ONE STEP CLOSER FROM CITIZEN SCIENCE TO OPEN DATA*

Helena Virić Gašparić[1], Katarina M. Mikac[2], Ivana Pajač Živković[1, **], Bruno Krehula[1], Matej Orešković[1], Marija A. Galešić[1], Pave Ninčević[1], Filip Varga[3, 4] and Darija Lemić[1]

[1]University of Zagreb, Faculty of Agriculture,
Division of Phytomedicine, Department for Agricultural Zoology
Zagreb, Croatia
[2]University of Wollongong, School of Earth, Atmospheric and Life Sciences,
Centre for Sustainable Ecosystem Solutions
Wollongong, Australia
[3]University of Zagreb, Faculty of Agriculture,
Division of Plant Science Department of Seed Science and Technology
Zagreb, Croatia
[4]Centre of Excellence for Biodiversity and Molecular Plant Breeding
Zagreb, Croatia

DOI: 10.7906/indecs.20.2.4
Regular article

Received: 14 January 2022.
Accepted: 14 March 2022.

ABSTRACT

Fireflies (Coleoptera: Lampyridae), with more than 2 000 species in 100 genera worldwide, are a charismatic nocturnal species. Although popular in different cultures because of their association with warm summer evenings in childhood, fireflies are an under-researched insect. Like numerous other insects worldwide, fireflies have experienced declines in their distribution and abundance. Anthropogenic impacts and climate change are likely to influence their development, reproduction, and survival. A project called "Krešo Krijesnica" (eng. "Krešo the Firefly"), used a Citizen Science model of data collection, to determine where are the fireflies located and how abundant are they throughout Croatia. Citizen Science involves the participation of the general or non-scientific public in data collection so determining the basic demographic profile of the citizen scientists involved was also one of the project goals. During the first phase of the project (2019-2021), data on fireflies were provided by citizen scientists through a formal survey on social media (Facebook, Instagram). Phase two aims to open the fireflies' datasets to the public through various open data portals. In the three years of the project, more than 16 000 records of fireflies were collected and analysed from over 1800 sightings. Descriptive statistics showed that the highest firefly population density was found in central Croatia, which is consistent with the greater number of people living in this area and thus a greater chance of firefly detection. Higher number pf female reporters were noted during the project. The dataset collected in this Citizen Science project presents a valuable source of information to the scientific community, especially in the field of entomology, conservation biology and ecology.

KEY WORDS

biodiversity conservation, citizen science campaign, fireflies, Krešo the Firefly, Lampyridae, Coleoptera

CLASSIFICATION

JEL: Q57

*This is the extended version of the abstract published in: Vujić, M. and Šalamon, D., eds.: Book of abstracts of the National Open Data Conference. University of Zagreb, Faculty of Traffic and Transport Sciences, Zagreb, 2021.

**Corresponding author, η: ipajac@agr.hr; +385 1 239 3948;
University of Zagreb, Faculty of Agriculture, Svetošimunska 23, HR – 10 000 Zagreb, Croatia

INTRODUCTION

Insects are among the groups of organisms most likely to be affected by anthropogenic impacts and climate change, because of the direct influence these factors have on their development, reproduction, and survival [1-3]. Fireflies (Coleoptera: Lampyridae) are among the most charismatic of all insects, and their spectacular courtship lit flights have inspired ordinary citizens, artists, poets and scientists alike. With more than 2 000 species in 100 genera, the worldwide diversity of lampyrids is impressive and includes both diurnal and nocturnal species [4-6]. They are true beetles that exhibit a distinct sexual dimorphism. For example, in *Lamprohiza splendidula* L. males are winged and fly while females mostly have only wing stumps incapable of flying [7]. Light organs are located on the underside of the body and light organ patterns differ depending on sex (Fig. 1). Fireflies, like numerous other insects, have experienced declines in their distribution and abundance worldwide [8]. Although they are widely known in society, especially due to the folkloric legends and association with warm summer evenings in childhood, fireflies continue to be loved and appreciated, despite it being an under-researched insect from a scientific viewpoint worldwide. Courtship in European lampyrids is simple and involves bioluminescent displays in which flying males are attracted to sedentary females that emit an uninterrupted bioluminescent glow [8]. Once females have mated, they generally cease to glow [8]. Both males and females of lampyrids are active for about an hour after sunset or until mating ends. There is variation in the glow patterns (i.e., continuous or intermittent) of *Lampyris,* Geoffroy, 1762 species males, with variation noted within species depending on the pattern and timing of display [9].

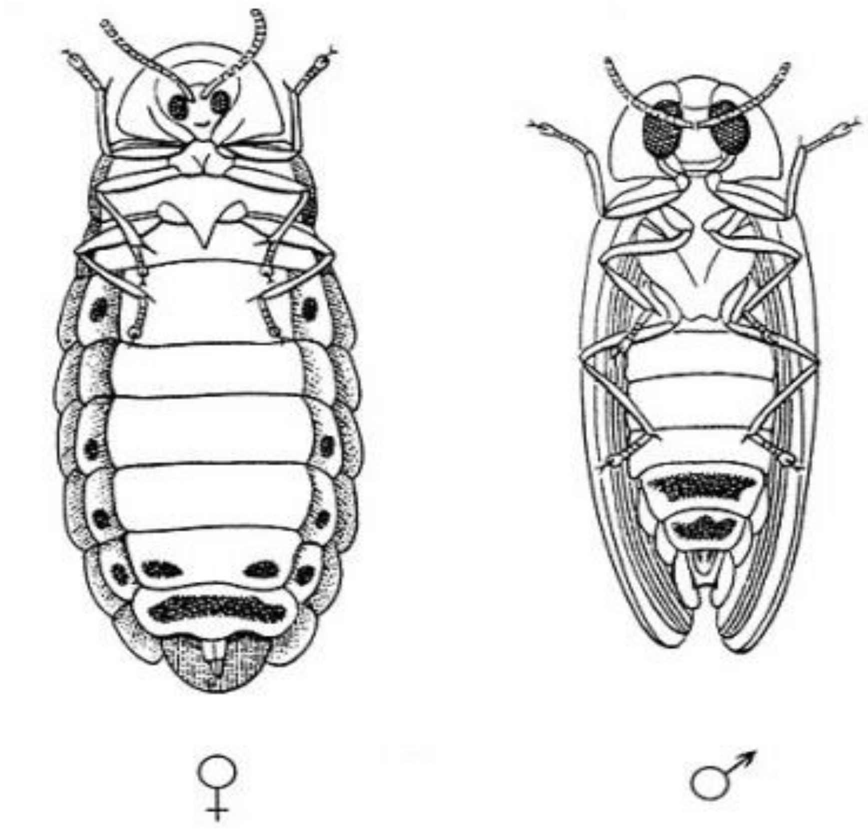


Figure 1. Male and female light organ patterns indicated as dark areas on the abdomen [9].

FIREFLY SPECIES OVERVIEW

Currently, 64 species of fireflies are known to occur in continental Europe [9; p.164-165]. The Catalogue of Palaearctic Coleoptera [10] mentions the occurrence of eight species of lampyrids in Croatia: *Lampyris germariensis* Jacquelin du Val, 1860; *Lampyris noctiluca* Linnaeus, 1767; *Lampyris zenkeri* Germar, 1817; *Lamprohiza germari* Küster, 1844; *Lamprohiza splendidula* Linnaeus, 1767; *Phosphaenus hemipterus* Geoffroy, 1762; *Luciola italica* Linnaeus, 1767 and *Luciola lusitanica* Charpentier, 1825. Voucher specimens of the listed species are found in the Croatian Natural History Museum in Zagreb and are part of the collections of Đ. Koča, I. Novak, P. Novak, R. Weingartner, R. Mikšić, K. I. Igalfy, V. Redenšek and F. Perović. To date, there are no empirical data in Croatia, as no official surveys have been conducted or published for species present, and there is no checklist of Croatian lampyrids. Only recently have European Lampyridae experts gained access to these collections and reviewed them for accuracy according to current taxonomic standards and naming conventions [11-13].

Detailed empirical data on lampyrids in Europe are limited to a few recent studies detailing the basic biology and ecology of even the most common fireflies, primarily by a few notable authors in the field (e.g. [10, 14] and more recently [9, 11-13, 15-18]. Moreover, there is a lack of information on the distribution and species diversity of lampyrids not only for Croatia, but for the whole of southeastern Europe [11]. According to De Cock [9], four species of lampyrids are most common and widespread in Europe: *L. noctiluca*, *L. splendidula*, *P. hemipterus*, and *L. lusitanica*. Therefore, it is reasonable to assume that these species are also the most common in Croatia.

CITIZEN SCIENCE PROJECT AND INITIATIVE "KREŠO THE FIREFLY"

Citizen science (CS) has become a popular means to address large scale scientific questions because it allows for the collection of basic scientific data that an individual or small group of scientists would not be able to conduct due to limited research funds, time, and personnel. CS is the participation of the public in scientific data collection on a particular topic with a specific goal [19]. CS is a popular means of not only informing the general, non-scientific public about important issues in science, but also educating the public and generating data and new knowledge without the usual high costs associated with conducting primary research, e.g. [20]. Well-known citizen science campaigns in entomology include the studies of Pocock and Evans [21], Kampen et al. [22] and Dennis et al. [23]. For fireflies in Europe, there are a number of successful citizen science campaigns targeting firefly's occurrences in Spain [16], Portugal [18] and Italy [24]. The data retained by the researchers and presented in scientific papers addresses the basic ecology of fireflies and provide information on the distribution and abundance of various European firefly species.

It is suspected that most emerging issues in biodiversity conservation is caused by new biological and digital technologies, new pollutants, and invasive species, climate change and potential human responses to it [25]. Light pollution affects the reproduction and migration of insects, amphibians, fish, birds, bats, and other animals, and can disrupt plants by distorting their natural day-night cycle [26]. Also, there is a growing awareness of light pollution, presence of artificial (anthropogenic) light, that is known to harm nocturnal species [27]. Over 30 % of all vertebrates and more than 60 % of all invertebrates in the world's biodiversity are nocturnal species. Artificial light threatens biodiversity by altering the nocturnal behaviour of organisms, such as insects attracted to streetlights. In Germany, each streetlight kills about 6.8 million insects every night in summer [26]. Fireflies live in damp and warm places, such as forested lakes, rivers or wetlands, by the sea or forest paths. Habitat loss has always been highlighted as one of the major threats, especially as a consequence of deforestation and urban expansion [28, 29]. Sexual communication in fireflies involves the emission and perception of the combination of bioluminescent cues and/or pheromones and is finely tuned depending on the species, environment, and time of day in which they occur [30, 31]. It is known that artificial light makes it difficult for fireflies to communicate and thus reduces the chances of mating, which affects future generations of fireflies [27].

Concerned students and scientists of University of Zagreb Faculty of Agriculture (Croatia) and University of Wollongong (Australia) launched a citizen science initiative called "Krešo the Firefly". The aim of the campaign was to collect as much data as possible to answer three specific questions about Lampyrids in Croatia: i) where are they located; ii) how abundant are they throughout Croatia and iii) what is the basic demographic profile of the citizen scientists.

FROM CITIZEN SCIENCE TO OPEN DATA

The concept of open data has started to become more and more prevalent in both government and private sector context, but also in science and citizen initiatives. In the last 10-15 years a number of open data initiatives have been launched with the aim of increasing the efficiency and transparency of governments, solving simple questions raised by citizens and local communities to name a few [32]. Open data should adhere to several principles in order to be considered open such as being complete, accessible in machine-readable format, license-free and free of charge [33, 34].

As principles of open data are being introduced into the Croatian scientific community [35] and public institutions and government departments [36], this presents an opportunity for increasing the availability and reuse of occurrence data of fireflies. To be able to analyse the datasets generated in the project, making the data available open, accessible and in machine-readable formats would increase usability of the collected data as well as assist scientists, natural resource managers and policy makers protect and restore firefly diversity in Croatia. As stated by Bonney et al. [37] those who seek to build capacity in the citizen science field can help by developing and improving open-source data management technologies to those already available in other fields of science.

There are some data on fireflies' form various publishers, e.g. [38-44] available in the Global Biodiversity Information Facility (GBIF), an international network and data infrastructure funded by the world's governments. Its goal is to provide open access to data on all species of life on Earth. Of Lampyridae species, there are a total of 34 records on the territory of Croatia, Austria, Slovenia and Bosnia and Herzegovina, including *L. noctiluca* (24), *L. zenkeri* (5), *L. germariensis* (2) and three unidentified species. The base of records presented on GBIF shows either human observations, occurrences or preservation/material specimens. No data are available on the number of fireflies in a given area or on the individual who made the taxonomic identification [45].

METHODOLOGY

DATA COLLECTION AND PROCESSING

The model for the development and implementation of a CS project was developed by a group of individuals with expertise in education, entomology, population biology, conservation biology, and information science. In order to involve as many people in Croatia as possible project pages were opened in social media (http://www.facebook.com/kreso.krijesnice.5, https://instagram.com/kresokrijesnica). In addition to social media, the public was also informed of the project through newspaper articles, radio broadcasts and television programmes and interviews (examples provided in Appendices A and B). Through the Croatian news agency Hina, the call for citizen participation was distributed to all media agencies. Also, a simple paper flyer (Appendix C) was used to advertise the project, this flyer was dropped in public places in dozens of Croatian cities.

Phase one of the project was conducted in Croatia between 2019 and 2021. The data on fireflies were provided by citizen scientists through a formal survey on Facebook and Instagram, based on sightings and records of date, location, and number of individuals and associated lampyrid data (pictures and videos). Effort and intensity for the media campaign was equal in all of the years of the project. Phase two aims to open the fireflies' datasets to the public through open data sources such as BIOLOGER, interactive free content platform aimed at collecting and digitizing knowledge about biodiversity in Eastern Europe (https://bdj.pensoft.net/article/53014) for reuse as well as to automate the process of data collection using mobile apps.

DESCRIPTIVE ANALYSIS

Descriptive statistics on collected data was performed using R, a free software environment for statistical analyses [46]. Geographic maps were produced using QGIS v 3.10.7 [47].

RESULTS

During all three years of "Krešo the Firefly" over 16 000 fireflies located all over Croatia were observed by more than 1 520 citizen scientists involved in the project. In 2019, over 1100 records with 10 421 fireflies and over 470 photos were collected. In 2020 and 2021, the number of sightings was lower due to the COVID-19 pandemic and the earthquake that hit the Capital, Zagreb and Sisak-Moslavina County. More than 500 records were collected and 5 369 fireflies were counted in 2020. The lowest number of reports, just over 180, was in 2021 with 1 606 fireflies recorded.

As shown in Figure 2, the number of fireflies reported ranged from 1 to over 100 per report throughout the project. The lowest number of fireflies per report was one, and this was the most common case (825 reports). The highest number of fireflies per report (over 100) was reported only five times.

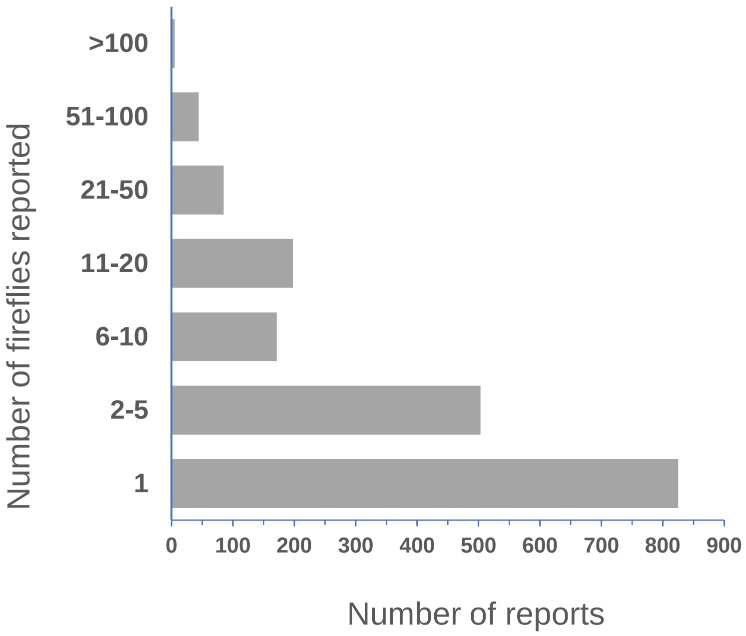


Figure 2. Number of reports regarding to reported number of fireflies collected in the citizen science campaign "Krešo the Firefly" 2019-2021.

Most sightings and reports (Fig. 3) came from central Croatia. Zagreb County, Primorje-Gorski Kotar County and Zadar County had the most reports in all three years of the project. In 2019, there were over 100 notifications from these counties, while in 2020 the number of notifications dropped below 100. In 2021, these same counties were still the most active, but the number of notifications decreased to between 17 and 21. In 2019 and 2020, citizen scientists from 16 other counties were consistently active, reporting between 10 and 100 firefly occurrences. Virovitica-Podravina County and Vukovar-Syrmia County participated with less than 10 reports in the "Krešo the Firefly" project during all three years of the project. The largest number of fireflies (Fig. 4) was reported from central Croatia (capital and Zagreb County, Karlovac County), which is connected with the higher number of reports from these locations. In the other counties, the number of fireflies is similar and varies between

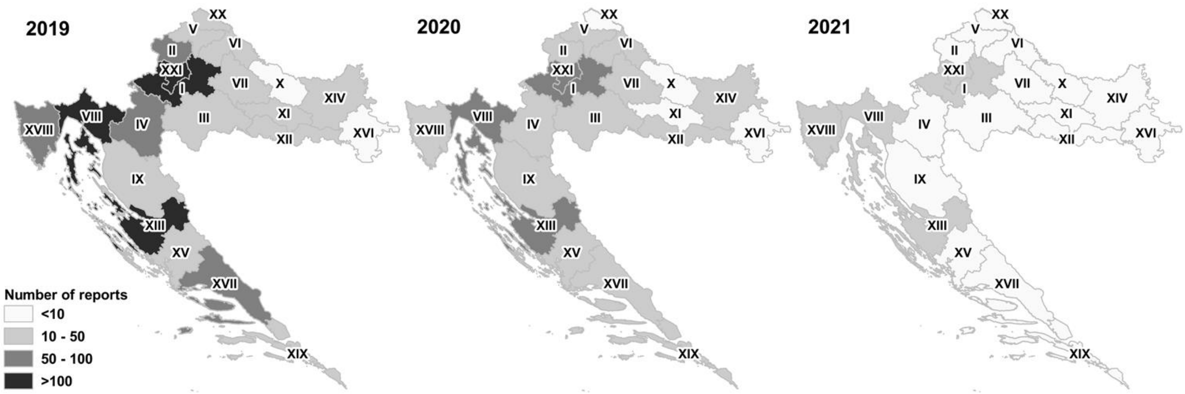


Figure 3. Number of firefly report density sorted by counties of Croatia in period from 2019 to 2021. Counties: I – Zagreb, II – Krapina-Zagorje, III – Sisak-Moslavina, IV – Karlovac, V – Varaždin, VI – Koprivnica-Križevci, VII – Bjelovar-Bilogora, VIII – Primorje-Gorski Kotar, IX – Lika-Senj, X – Virovitica-Podravina, XI – Požega-Slavonia, XII – Brod-Posavina, XIII – Zadar, XIV – Osijek-Baranja, XV – Šibenik-Knin, XVI – Vukovar-Syrmia, XVII – Split-Dalmatia, XVIII – Istria, XIX – Dubrovnik-Neretva, XX – Međimurje, XXI – City of Zagreb.

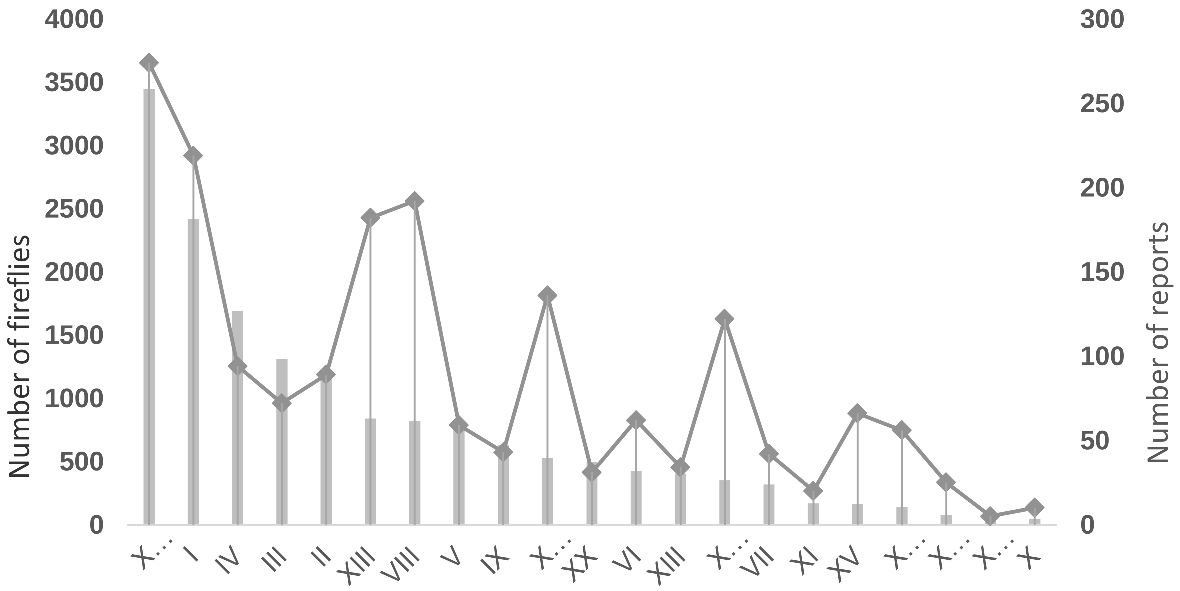


Figure 4. Number of firefly reports (blue line) and number of observed fireflies (light blue column) fireflies collected in the citizen science campaign "Krešo the Firefly" 2019-2021. Counties: Numbers denote Counties as in Figure 3.

500 and 1300 and the number of reports between 70 and 130. On average, there were 1,2 reports per person (or citizen scientist), which means that in most reports a person reported the occurrence of fireflies only once. The maximum number of reports per person was six.

Citizen scientists classified by gender are shown in Figure 5. In all three years, the number of female citizen scientists was significantly higher than the number of male citizen scientists, ranging from 50 % to 100 % of participants. The proportion of male citizen scientists averaged less than 25 % throughout the study period.

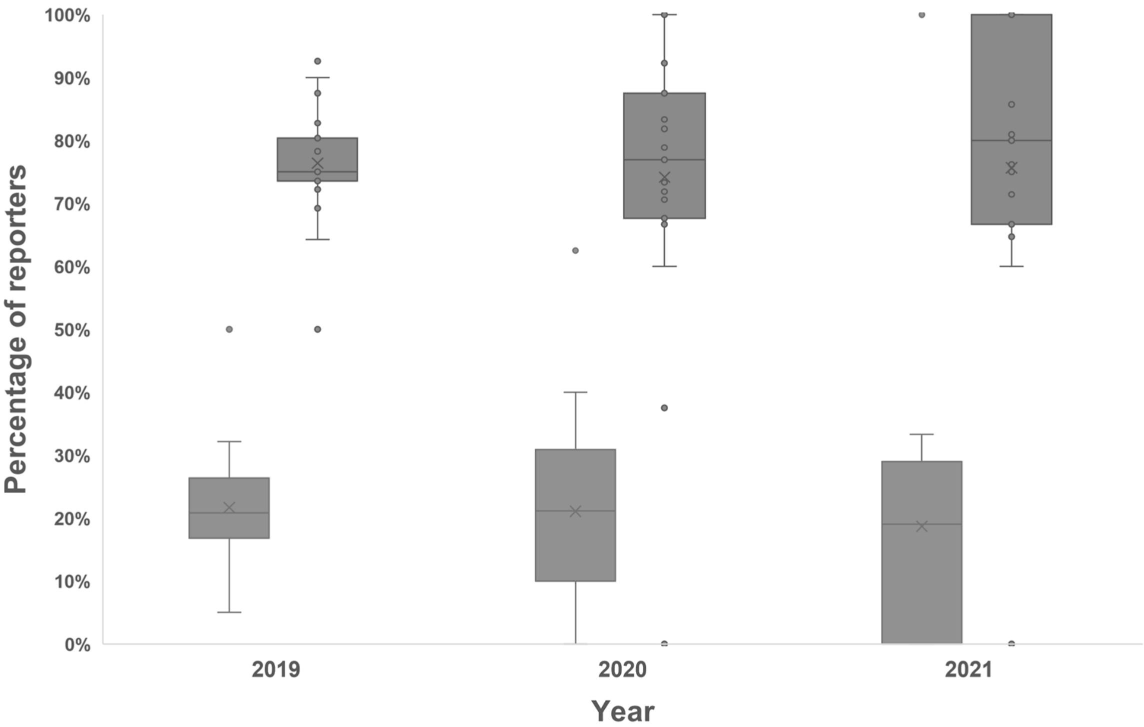


Figure 5. Percentage of female (light grey) and male (dark grey) citizen scientists who participated in the citizen science campaign "Krešo the Firefly" 2019-2021.

DISSCUSION

Our study documented four different Lampyridae species and their occurrence and abundance throughout Croatia using a CS approach. Such data makes an important contribution to the conservation biology of the species documented, as shown by previous research [48-52]. While GBIF provides only 34 records of Lampyridae species on the territory of Croatia, Austria, Slovenia and Bosnia and Herzegovina [45], we reported a considerable number of firefly occurrences (n = 1833) for all 21 counties of Croatia over a three-year period. The higher number of reports from central Croatia is consistent with the larger number of people living in this area. In 2020 and 2021, the number of sightings was lower due to the COVID-19 pandemic and the earthquake that hit the Capital, Zagreb and Sisak-Moslavina County. The COVID-19 pandemic was a setback for our project, as the movement of the Croatian population was restricted for a while and media had different focus priorities. The search for fireflies was not a priority for most people at the time. Nevertheless, we received positive feedback from several citizen scientists that this activity helped them feel better, be active in the fresh air and be useful in such unfortunate times. The same positive effect of participating in CS and feeling like they are contributing to society as a whole was also noted by Holden [53]. According to Groffman at al. [54] most CS projects are generally considered necessary to raise awareness of environmental issues, democratize science, and (re)establish trust between science and society. Our project succeeded in involving Croatia as a whole and the media in the search for fireflies, as most people have fond memories of balmy summer nights filled with firefly light displays.

Most of engaged volunteers were female reporters. The same trend, with a greater proportion of female reporters, was noted in the invasive species case study by Crall et al. [55]. Studies by Kwak and Radler [56]; Underwood et al. [57], report that, similar to traditional survey modes, women respond in greater proportions in online surveys than men. During the project, we found that the campaign had little impact in some counties (Virovitica-Podravina, Vukovar-Syrmia, Požega-Slavonia, and Osijek-Baranja), which may have affected the results. The most likely reason is the older population in rural areas as well as the fact that these counties are affected by drastic depopulation trends. Limited access to the Internet and social networks is most probable reason why the CS campaign and media surrounding did not reach these counties sufficiently. A similar example of lower participation of rural areas in a CS projects was highlighted by Holden [53]. Research by Kwak, and Radler, Underwood et al. and Vaske et al. [56-58] has shown that demographic factors, including place of residence, influence people's attitudes and values toward scientific research. According to Loenen et al. [35], word of mouth and newspapers have been shown to work best for recruiting CS in rural areas. In our study, the primary focus for CS recruitment was the use of social media, while personal contacts and newspapers were limited sources. Researchers often use visual estimates to infer species frequency, and these estimates typically vary among observers [59-61]. In our study, most CS reported counting "the firefly" as only one specimen, while abundance was likely higher, which is consistent with the study by Crall et al. [55], who found that volunteers and professionals tended to under- or overestimate the number of different species differently. The highest abundance in the whole of Croatia was recorded in the most densely populated areas with the highest number of young people actively involved in social networks. According to Stohlgren [62], standardised monitoring protocols established by professionals and tested under realistic field conditions can improve data quality and analyses. In our study, photographs submitted by volunteer reporters were often of low quality and locations were poorly described. Some material could not be used to identify lampyrids and some locations could not be plotted on the map produced. According to Crall et al. [55], photographing geolocated specimens may be better for verifying the accuracy of species identifications.

CONCLUSION

The dataset documented in this study is a valuable source of information for the scientific community, especially in the fields of entomology, conservation biology, and ecology. New communication tools and social science research can help scientists interact more effectively with the public. In this study, we were able to collect a large amount of data on individual firefly locations in each of the 21 counties and abundance per site. The basic demographic profile showed higher female activity in more populated areas (capital and larger cities), which provides good insight into preparation of future strategies for targeting the project to more rural areas and the male portion of the population. More focus on diverse and specific habitats, as well as monitoring of environmental factors, would further complete the dataset. The data collected as part of "Krešo the Firefly", including the list of species occurring in Croatia, their abundance, locations, and original photographs, represent a high- quality dataset. Making these data available in open and machine-readable formats would certainly increase their reusability by other scientists and policy makers in Croatia, and ultimately help us protect and restore firefly diversity in Croatia. A collaboration with a team of experts from Biologer has already been established, and the transfer of the data into open databases according to the given specifications of their portal is underway. A data collection strategy for future research was also agreed upon. This research is a first attempt, at least in Croatia, to focus on this sensitive, endangered, poorly distributed and low abundant taxa. Although further collections in a wider range of habitats would be of further benefit, these data have provided a better understanding of Croatian fireflies and a good foundation for future work on firefly biodiversity and conservation in this region.

ACKNOWLEDGMENTS

We would like to thank all the volunteers and colleagues who contributed greatly to the development of the Citizen Science project “Krešo the Firefly” and to the data collection for the preparation of this article. A special thanks goes to all the media that were interested in the topic and helped in disseminating the project details.

This research is part of TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592.

APPENDIX A

Scientific television show “Znanstveni krugovi”, shown on June 6, 2019, https://hrti.hrt.hr/api/api/ott/socialshare?target=copy&referenceId=438788155&channelReferenceId=null&mobile=false&serie=false&operatorReferenceId=hrt.

APPENDIX B

Scientific television show “Prometej” shown on April 4, 2021, https://hrti.hrt.hr/api/api/ott/socialshare?target=copy&referenceId=f3e18e43-6c07-8261-0f87-6a2744b8f8d5&channelReferenceId=null&mobile=false&serie=false&operatorReferenceId=hrt.

APPENDIX C

Figure 6. Poster “Krešo Krijesnica”.

REFERENCES

[1] Bale, J.S., et al.: *Herbivory in global climate change research: Direct effects of rising temperature on insect herbivores.*
Global Change Biology **8**, 1-16, 2002,
http://dx.doi.org/10.1046/j.1365-2486.2002.00451.x,

[2] Cornelissen, T.: *Climate Change and Its Effects on Terrestrial Insects and Herbivory Patterns.*
Neotropical Entomology **40**(2), 155-163, 2011,
http://dx.doi.org/10.1590/S1519-566X2011000200001,

[3] Scherber, C., et al.: *Multi-factor climate change effects on insect herbivore performance.*
Ecology and Evolution **3**(6), 1449-1460, 2013,
http://dx.doi.org/10.1002/ece3.564,

[4] Lloyd, J.E.: *Insect Bioluminescence*.
In: Herring, P.S., ed.: *Bioluminescence in Action.* Academic Press, New York, 1978,

[5] Lloyd, J.E.: *Sexual Selection in Luminescent Beetles*.
In: Blum, M.S. and Blum, N.A., eds.: *Sexual Selection and Reproductive Competition in Insects.* Academic Press, New York, 1979,

[6] Viviani, V.R.: *Fireflies (Coleoptera: Lampyridae) from south eastern Brazil: Habitats, Life History, and Bioluminescence*.
Annals of the Entomological Society of America **94**(1), 129-45, 2001,
http://dx.doi.org/10.1603/0013-8746(2001)094[0129:FCLFSB]2.0.CO;2,

[7] Zahradnik, J.: *A Filed guide in colour to Insects.*
Aventinum Publishing House, Prague, 1990,

[8] Menéndez, R.: *How are insects responding to global warming?*
Tijdschrift voor Entomologie **150**(2), 355-365, 2007,

[9] De Cock, R.: *Biology and behaviour of European lampyrids*.
In: Meyer-Rochow, V.B., ed.: *Bioluminescence in Focus-A collection of Illuminating Essays.* Research signpost, Kerala, 2009,

[10] Geisthardt, M.: *Lampyridae*.
In: Löbl, I. and Smetana, A., eds.: *Catalogue of Palaearctic Coleoptera, 4.* Apollo Books, Stenstrup, 2007,

[11] Novák, M. and De Cock, R.: *Light in the darkness I: Some insights in Lampyridae from the Balkans from the Zagreb Museum of Croatia: Species of the genus Luciola Laporte, 1833.*
In: Proceedings of 2017 International Firefly Symposium, Taipei, 2017,

[12] De Cock, R. and Novák, M.: *Light in the darkness II: Some insights in Lampyridae from the Balkans from the Croatian Natural History Museum: Species of the genera Lampyris Geoffr., 1762, Phosphaenus Cast., 1833 and Lamprohiza Motsch., 1858.*
In: Proceedings of 2017 International Firefly Symposium. Taipei, 2017.

[13] Novák, M.; De Cock, R. and Day, J.: *Light in the darkness III: Some insights in Lampyridae from the Balkans from the Croatian Natural History Museum: Genetic results and notes on the Montenegrinian endemic Luciola novaki Müller, 1946*.
In: Proceedings of 2017 International Firefly Symposium, Taipei, 2017,

[14] Geisthardt, M.; Figueira, G.; Day, J.C. and De Cock, R.: *A review of Portuguese fireflies with a description of a new species, Lampyris iberica, sp. nov. (Coleoptera: Lampyridae).*
Heteropterus Revista de Entomología **8**(2), 147-154, 2008,

[15] De Cock, R.: *Rare or simply overlooked? Practical notes for survey and monitoring of the small glow-worm Phosphaenus hemipterus (Coleoptera: Lampyridae).*
Belgian Journal of Zoology **130**(2), 93-101, 2000,

[16] Guzmán Alvarez, J.R., and De Cock, R.: *The biology and distribution of glow-worms (Coleoptera: Lampyridae) in Spain*.
The Lampyrid **1**, 22-31, 2011,
http://dx.doi.org/10.1016/j.diapre.2011.06.001,

[17] De Cock, R. and Guzmán Alvarez, J.R.: *Methods and recommendations for surveying firefly glow-worms (Coleoptera: Lampyridae): a practical example from Southern Spain*.
The Lampyrid **3**, 49-95, 2013,
http://dx.doi.org/10.3901/JME.2013.24.095,
[18] De Cock, R.; Alves, H.N.; Oliveira, N.G. and Gomes, J.: *Fireflies and Glow-worms of Portugal*.
Aguas e Parque Bioológico de Gaia, Câmara Municipal de Vila Nova de Gaia, 2015,
[19] Silvertown, J.: *A new dawn for citizen science*.
Trends in Ecology & Evolution **24**(9),467-471, 2009,
http://dx.doi.org/10.1016/j.tree.2009.03.017,
[20] Acorn, J.H.: *Entomological citizen science in Canada*.
The Canadian Entomologist **149**(6), 774-785, 2017,
http://dx.doi.org/10.4039/tce.2017.48,
[21] Pocock, M.J.O. and Evans, D.M.: *The Success of the Horse-Chestnut Leaf-Miner, Cameraria ohridella, in the UK Revealed with Hypothesis-Led Citizen Science*.
PLoS ONE **9**(1), e86226, 2014,
http://dx.doi.org/10.1371/journal.pone.0086226,
[22] Kampen, H., et al.: *Approaches to passive mosquito surveillance in the EU*.
Parasites Vectors **8**(9), 1-13, 2015,
http://dx.doi.org/10.1186/s13071-014-0604-5,
[23] Dennis, E.B., et al.: *Using citizen science butterfly counts to predict species population trends*.
Conservation Biology **31**(6), 1350-1361, 2017,
http://dx.doi.org/10.1111/cobi.12956,
[24] Picchi, M.S., et al.: *Fireflies and land use in an urban landscape: the case of Luciola italica L. (Coleoptera: Lampyridae) in the city of Turin*.
Journal of Insect Conservation **17**(2), 797-805, 2013,
http://dx.doi.org/10.1007/s10841-013-9562-z,
[25] Sutherland, W. J., et al.: *A horizon scan of global conservation issues for 2014*.
Trends in Ecology & Evolution **29**(1), 15-22, 2014,
http://dx.doi.org/10.1016/j.tree.2013.11.004,
[26] Hölker, F.; Wolter, C.; Perkin, E.K. and Tockner, K.: *Light pollution as a biodiversity threat*.
Trends in Ecology & Evolution **25**(12), 681-682, 2010,
http://dx.doi.org/10.1016/j.tree.2010.09.007,
[27] Avalon, C.S.; Owens, S. and Lewis, M.: *The impact of artificial light at night on nocturnal insects: A review and synthesis.*
Ecology and Evolution **8**, 11337-11358, 2018,
http://dx.doi.org/10.1002/ece3.4557,
[28] Barlow, J., et al.: A*nthropogenic disturbance in tropical forests can double biodiversity loss from deforestation*.
Nature **535**, 144-147, 2016,
http://dx.doi.org/10.1038/nature18326,
[29] Gonçalves-Souza, D.; Verburg, P.H. and Dobrovolski, R.: *Habitat loss, extinction predictability and conservation efforts in the terrestrial ecoregions*.
Biological Conservation **246**, No. 108579, 2020,
http://dx.doi.org/10.1016/j.biocon.2020.108579,
[30] Cronin, T.W.; Järvilehto, M.; Weckström, M. and Lall, A.B.: *Tuning of photoreceptor spectral sensitivity in fireflies (Coleoptera: Lampyridae).*
Journal of Comparative Physiology A 186, 1-12, 2000,
http://dx.doi.org/10.1007/s003590050001,
[31] Stanger-Hall, K.F., et al.: *The evolution of sexual signal modes and associated sensor morphology in fireflies (Lampyridae, Coleoptera).*
Proceedings of the Royal Society B: Biological Sciences **285**(1871), No. 20172384, 2018,
http://dx.doi.org/10.1098/rspb.2017.2384,

[32] Welle Donker, F. and van Loenen, B.: *How to assess the success of the open data ecosystem?* International Journal of Digital Earth **10**(3), 284-306, 2017, http://dx.doi.org/10.1080/17538947.2016.1224938,

[33] Sunlight Foundation: *Ten Principles for Opening Up Government Information* https://sunlightfoundation.com/policy/documents/ten-open-data-principles, accessed 14th February 2022,

[34] Welle Donker, F.; van Loenen, B. and Bregt, A.K.: *Open Data and Beyond*. ISPRS International Journal of Geo-Information **5**(4), 48, 2016, http://dx.doi.org/10.3390/ijgi5040048van,

[35] Loenen, B. et al.: *Towards an open data research ecosystem in Croatia*. In: Virkar, S., et al., eds.: InternationalConference EGOV-CeDEM-ePart, CEUR, 2020,

[36] Zdjelar, R.; Musa, A. and Žajdela Hrustek, N.: *Open data availability in Croatian local government: Improving the quality of life.* The Smart Cities and Regional Development Journal **5**(3), 21-40, 2021,

[37] Bonney, R., et al.: *Next steps for citizen science*. Science **343**, 1436-1437, 2014, http://dx.doi.org/10.1126/science.1251554,

[38] iNaturalist.org: *iNaturalist Research-grade Observations.* http://dx.doi.org/10.15468/ab3s5x,

[39] The International Barcode of Life Consortium: *International Barcode of Life project.* http://dx.doi.org/10.15468/inygc6,

[40] de Vries, H. and Lemmens M.: *Nature data from around the World.* http://dx.doi.org/10.15468/5nilie,

[41] naturgucker.de: *naturgucker*. http://dx.doi.org/10.15468/uc1apo,

[42] European Bioinformatics Institute: *INSDC Sequences*. European Nucleotide Archive (EMBL-EBI), 2021, http://dx.doi.org/10.15468/sbmztx,

[43] Slieker, F.J.A.; van der Es.H.; Andeweg, R. and Langeveld, B.W.: *Specimens*. Natural History Museum Rotterdam, 2022 http://dx.doi.org/10.15468/kwqaay,

[44] Biologiezentrum Linz Oberoesterreich: *Biologiezentrum Linz*. http://dx.doi.org/10.15468/ynjblx,

[45] Global Biodiversity Information Facility: *Free and open access to biodiversity data.* https://www.gbif.org, accessed 18th February 2022,

[46] R Core Team: *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, 2020, https://www.R-project.org,

[47] QGIS Development Team: *QGIS Geographic Information System 3.10.7.* Open Source Geospatial Foundation Project 2021, http://qgis.osgeo.org,

[48] Hames, R.S., et al.: *Effects of forest fragmentation on tanager and thrush species in eastern and western North America*. In: George, L. and Dobkins, D.S., eds.: *The Effects of Habitat Fragmentation on Birds in Western Landscapes: Contrasts with Paradigms from the Eastern United States*. Cooper Ornithological Society, 2002,

[49] Hames, R.S., et al.: *Adverse effects of acid rain on the distribution of the wood thrush Hylocichla mustelina in North America*. Proceedings of the National Academy of Sciences **99**: 11235-11240, 2002, http://dx.doi.org/10.1073/pnas.172700199,

[50] Cooper, C.B.; Hochachka, W.M.; Butcher, G. and Dhondt, A.A.: *Season and latitudinal trends in clutch size: Thermal constraints during laying and incubation.* Ecology **86**, 2018-2031, 2005, http://dx.doi.org/10.1890/03-8028,

[51] Cooper, C.B.; Hochachka, W.M. and Dhondt, A.A.: *Latitudinal trends in within-year reoccupation of nest boxes and their implications*.
Journal of Avian Biology **36**, 31-39, 2005,
http://dx.doi.org/10.1111/j.0908-8857.2005.03319.x,

[52] Rosenberg, K.V.; Lowe, J.D. and Dhondt, A.A.: *Effects of forest fragmentation on breeding tanagers: A continental perspective*.
Conservation Biology **13**, 568-583, 1999,
http://dx.doi.org/10.1046/j.1523-1739.1999.98020.x,

[53] Holden, A.D.: *Organizing Rural Communities for Effective Citizen Science Programs.*
Graduate Student Theses, Dissertations, & Professional Papers 724, 2007,

[54] Groffman, P.M. et al.: *Restarting the conversation: Challenges at the interface between ecology and society.*
Frontiers in Ecology and the Environment **8**(6), 284-291, 2010,
http://dx.doi.org/10.1890/090160,

[55] Crall, A.W., et al.: *Assessing citizen science data quality: an invasive species case study*.
Conservation Letters **4**, 433-442, 2011,
http://dx.doi.org/10.1111/j.1755-263X.2011.00196.x,

[56] Kwak, N. and Radler, B.: *A comparison between mail and web surveys: response pattern, respondent profile, and Data quality).*
Journal of Official Statistics 1**8**(2), 257-273, 2002,

[57] Underwood, D., Kim, H., and Matier, M.: *To mail or to Web: Comparisons of survey response rates and respondent characteristics.*
40th Annual Forum of the Association for Institutional Research, Cincinnati, 2000,

[58] Vaske, J.; Donnelly, M.; Williams, D. and Jonker, S.: *Demographic Influences on Environmental Value Orientations and Normative Beliefs About National Forest Management.*
Society and Natural Resources **14**(9), 761-776, 2001,
http://dx.doi.org/10.1080/089419201753210585,

[59] Williams, C.; Ericsson, G. and Heberlein, T.: *A quantitative summary of attitudes toward wolves and their reintroduction (1972-2000).*
Wildlife Society Bulletin **30**(2), 575-584, 2002,
https://www.jstor.org/stable/3784518,

[60] Kellert, S.: *Attitudes and Characteristics of Hunters and Antihunters*.
Transactions of the 43rd North American Wildlife and Natural Resources Conference, 1978,

[61] Kennedy, K.A. and Addison, P.A.: *Some considerations for the use of visual estimates of plant cover in biomonitoring.*
Journal of Ecology **75**(1), 151-157, 1987,
http://dx.doi.org/10.2307/2260541,

[62] Stohlgren, T.J.: *Measuring plant diversity: lessons from the field*.
Oxford University Press, New York, 2007.

Interdisciplinary Description of Complex Systems 20(2), 125-135, 2022

URBAN DOG SPACES: THE OPENNESS OF DOG-RELATED GOVERNMENT DATA IN THE CITY OF ZAGREB, CROATIA*

Filip Varga[1, 2, **], Larisa Hrustek[3], Karlo Kević[4], Frederika Welle Donker[5] and Dragica Šalamon[1]

[1]University of Zagreb, Faculty of Agriculture
Zagreb, Croatia
[2]Centre of Excellence for Biodiversity and Molecular Plant Breeding
Zagreb, Croatia
[3]University of Zagreb, Faculty of Organization and Informatics
Varaždin, Croatia
[4]University of Zagreb, Faculty of Geodesy
Zagreb, Croatia
[5]Delft University of Technology Faculty of Architecture and the Built Environment
Delft, Netherlands

DOI: 10.7906/indecs.20.2.5
Regular article

Received: 15 January 2022.
Accepted: 3 March 2022.

ABSTRACT

Dogs and dog owners are increasingly present in modern urban spaces, and the construction and maintenance of urban infrastructure that includes places for them has become standard in most major cities. We wanted to investigate the extent to which the City of Zagreb is adhering to open data principles when it comes to dog-related data it makes available to citizens.The openness and quality of dog-related data was analysed in three steps. First, dog-related data was searched on various official websites and portals of the city and the data formats were ranked according to the five-star system for open data. In the second step, based on the available data, a field survey was conducted in 2020 to verify the found datasets and geocode them using a GPS device. In addition, the locations obtained from the local community of dog owners through social media were reviewed. Finally, data obtained from the survey was cross-checked with the government data to assess their quality. Government data on the locations of 300 dog waste bins and 72 green areas where dogs can be walked off-leash were available in Croatian from one or more government sources. All data sets found received the lowest score in terms of open data formats. Field survey revealed differences between the data and reality. The location of 40 dog waste bins could not be confirmed, and additional 53 bins were found that were not mentioned in the data. As for green areas, there were reportedly 10 dog parks in the city of Zagreb. The survey confirmed all locations and discovered 12 more, five of which were mentioned in the data but not designated as dog parks. The results suggest that the municipality needs to update the already open datasets more frequently. Improved implementation of these datasets into existing city data portals or the creation of a separate hub for dog owners would greatly improve the availability and reuse of this data by citizens.

KEY WORDS

open data, *Canis familiaris*, city infrastructure, local government data

CLASSIFICATION

JEL: O18

*This is the extended version of the abstract published in: Vujić, M. and Šalamon, D., eds.: Book of abstracts of the National Open Data Conference. University of Zagreb, Faculty of Traffic and Transport Sciences, Zagreb, 2021.

**Corresponding author: fvarga@agr.hr; +385 1 239 3845;
University of Zagreb, Faculty of Agriculture, Svetošimunska cesta 25, HR – 10 000 Zagreb, Croatia

INTRODUCTION

Dogs (*Canis lupus familiaris* Linnaeus, 1758) have been among man's most loyal animal companions since they were domesticated from their wild wolf ancestors 15 millennia ago [1]. As social structures have evolved throughout history, their popularity has continued to increase, and in recent years, dog ownership has peaked worldwide [2]. In addition, the notion that a dog's primary role is to guard the household and accompany its owners on hunts is now largely a thing of the past. Other benefits of dog ownership, such as companionship, social contact (with other dog owners), and time spent outdoors, are becoming more common, and the number of dogs in urban areas is increasing [3]. With this in mind, the need for a well-developed dog-friendly infrastructure in modern cities arises. Dog walking and using public spaces are the main activities of dog owners in urban areas [4]. With some exceptions, there is surprisingly little research on the development of urban infrastructure for dog owners. Most of the research to date has been aimed at examining the effects of dog ownership on social interactions [5].

In recent years, there has been a growing trend for cities to publish data according to Open Data standards, link data from multiple sources, allow citizens to generate and update data using mobile apps and application programming interfaces (APIs), and use this data in policy making [6]. Governments (local, regional, and national) have a great responsibility to inform their citizens about existing and developing infrastructures and services in an appropriate and user-friendly manner, and to allow citizens to express their needs [7]. As more and more citizens become computer literate, governments (both local and national) are under pressure to disseminate and collect important information via mobile apps [8], social media [9] or APIs [10] with more or less success.

As Open Data concepts that have been slowly but steadily implemented in Croatia over the last 10 years through policy making, development of data portals and formal obligation to publish data of public institutions, we finally see the positive trend of opening government data in various sectors. In terms of local governments, we can see that the larger cities are keeping up with the national trend, while other cities are significantly lagging behind in publishing their data [11]. Rijeka and Zagreb are definitely ahead in terms of opening data with 175 and 75 datasets opened, respectively from various categories such as locations of parking spots for cars and bicycles, bus stops, and educational facilities, to name a few [12, 13].

Dog-related government data in cities, presented in a simple and user-friendly manner, is of great benefit not only to existing dog owners, informing them of changes in dog infrastructure, but also to out-of-town visitors (e.g., tourists) navigating the city with their dogs. Another important stakeholder group that data relating to dog content would be of great interest is the bussiness sector aimed at dog owners, such as pet supplies stores, veterinary practices and dog grooming services in identifying city areas with great potential for expansion of their bussiness. The aim of this research was to determine the extent to which government information about dog-related content is available to end users, in what form, and the quality of that data. The research was based on a simple social question, "What dog-related content is available in the city of Zagreb?" in order to measure the maturity of the dog owner-focused open data ecosystem in Zagreb.

MATERIALS AND METHODS

To assess the openness of dog-related government data in the city of Zagreb, a mixed approach was used. In the first phase, various potential government data sources were identified and accessed. For the City of Zagreb, there are two open data portals through which municipal data can be accessed. One of them is the Open Data Portal of the City of Zagreb [13], where open data can be searched based on categories, tags, data formats and licence types. The second portal, ZG Geoportal – Zagreb Spatial Data Infrastructure [14], provides geospatial data for the territory of the City of Zagreb in interactive map and WMS (Web Map Service) formats with accompanying metadata catalogue. In addition, the City website [15], the City Services website [16], and the City of Zagreb Statistical Yearbook for 2020 [17] were reviewed for dog-related datasets. All available datasets were downloaded regardless of the format in which they were available and where possible, ranked according to the five-star open data system [18]. Data from government sources were translated into a machine-readable table format (where necessary) and used as a starting point for the field survey.

In the second phase of this research, a field survey was conducted in 2020 (Fig. 1). To geocode open data from government sources, each location from the original datasets was visited and geographic coordinates were recorded using the Garmin Etrex Vista GPS device if the content was actually present at said location. Additionally, a Facebook community page Dog City Zg was created to build a community of dog owners [19]. Users were asked to send either geocoded photos, geographic coordinates, or descriptions of dog content locations in their neighbourhoods. The information received through the Facebook community page was additionally verified in the field and geocoded using the aforementioned GPS device in combination with satellite imagery, Google Maps, and OpenStreetMap platforms. Finally, the data from government sources and the data from the field survey were cross-checked with QGIS 3.10.7 [20] to verify how accurate and up-to-date the data from open sources are for the city of Zagreb.

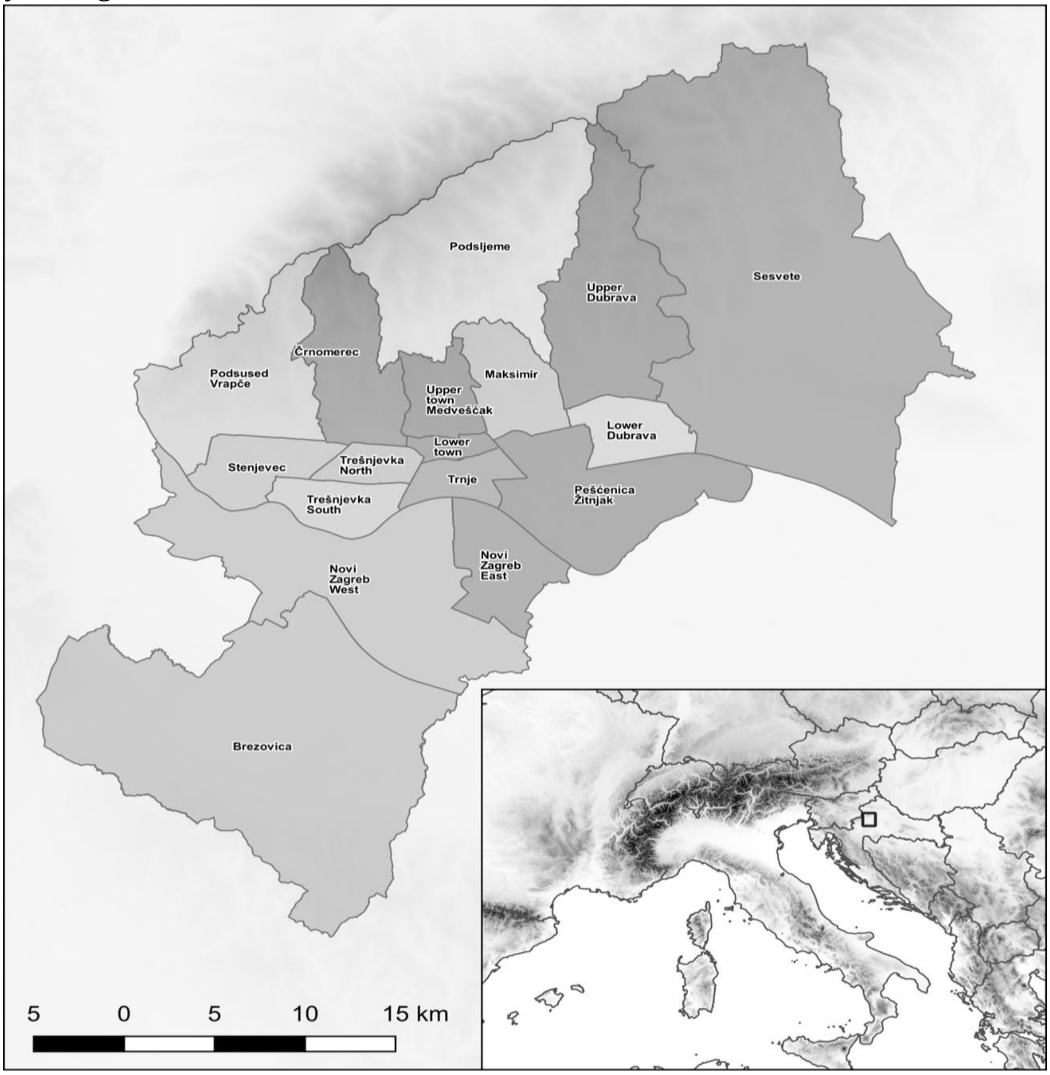


Figure 1. Location of the field survey area in the City of Zagreb.

RESULTS AND DISCUSSION

The first phase of the research identified data on three types of dog-related content from various government data sources: Locations of dog waste bins, locations of dog parks, and locations of green areas where dogs can be walked off-leash.

The data on locations of dog waste bins were found on the website of the main service of the City of Zagreb responsible for waste management – Zagreb Holding, subsidiary Čistoća – under the heading Cleaning of public traffic areas [21]. The dataset was available in tabular form as a text-based PDF file without metadata, which was rated with one star on the five-star deployment scheme. The dataset contained the following information: Street name and number (if applicable), site description, district, and the number of bins per site. Although the dataset was available on both the Croatian and English versions of the website (and referenced in both languages depending on the preferred language), the data were presented in only Croatian. The dataset contained 293 records with information on 300 locations of dog waste bins (in some cases with multiple bins at one location) in 16 of 17 city districts. Several records indicated that there were no dog waste bins at a particular location. The district with the highest number of dog waste bins was Upper town-Medvešćak (36 bins), while no bins were recorded in government data for the Brezovica district, with an average of 17.7 bins per city district. When reviewing the ZG Geoportal interactive map, under the Green Space Cadastre group, the same data were observed in the Parks equipment layer [14], marked on the map as “other facilities”. When looking at more detailed information for the specific location of the bin, the type of equipment was defined as “dog equipment”, with no difference between dog waste bins and equipment elements such as platforms or crossbars. The layers in this group can be downloaded through the WMS service using various GIS (Geographic Information System) software, but only in the raster format without background data.

Data on the locations of dog parks and green areas where dogs can be walked off-leash came from three different sources in different formats. Although they are clearly defined as different concepts in the Statistical Yearbook of the City of Zagreb for 2020 [17], the information about them was always available in a common dataset. The first source was the Decision of the City Assembly from December 2018 on the conditions and manner of keeping domestic animals and the manner of treatment of abandoned and lost animals and wild animals [22]. Article 17 of this decision lists 72 locations where dogs may be walked off-leash, including 10 dog parks. The decision was available as a text-based PDF file without metadata, rated with one star on the five-star deployment scheme, and was written exclusively in Croatian. At the end of the article, it states that a mapped representation of said locations will be published on the city's website. In fact, 22 static maps in PDF format named after the respective city area are published on the city website under the group Agriculture and Forestry, section Animal protection, subsection Domestic Animal Husbandry [23]. The maps show 66 localities, which are numerically identified on the maps. According to the information on the website, the maps were last updated in April 2017, which may explain the discrepancy in the number of locations between the two government sources. The data can also be found on ZG Geoportal under the Animal Welfare group in the Public Areas for Dogs layer [14]. A total of 72 locations are marked, but no detailed site information is available. The layer is marked as a new layer on the Geoportal and is currently not available for download via the WMS service.

All locations (300 dog waste bins and 72 off-leash dog walking areas) from available government data were reviewed during the field survey. The locations for which it was indicated that no bins were present were also reviewed. Of the 300 entries for dog waste

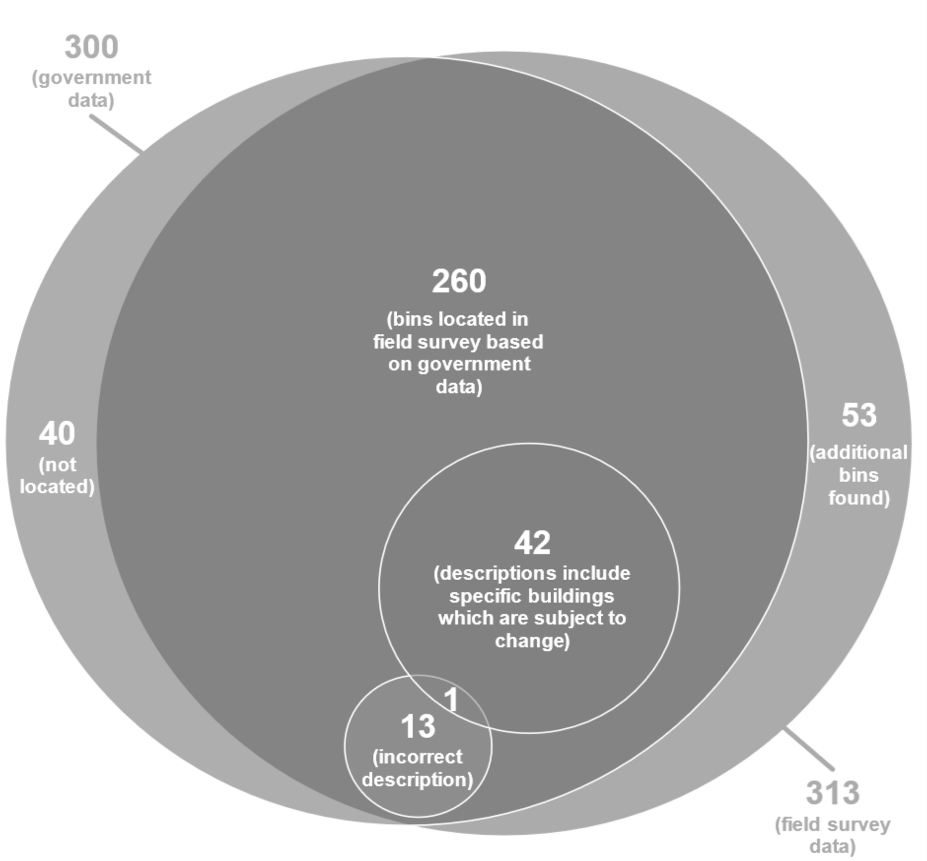


Figure 2. Comparison of latest government and 2020 field survey data on dog waste bin in the City of Zagreb.

bins, 260 were present at the described locations (Fig. 2). A large number of the descriptions (42) included information about proximity to specific buildings such as grocery stores**,** phone booths, schools, daycares, etc. Only a few records (13) contained incorrect information about locations, such as incorrect or outdated street names, incorrect house numbers, or an incorrect number of bins at the location.

Records that could not be found based on the given description (40) include five waste bins that, according to government data, are located in a dog park that does not exist and five waste bins that, according to government data, are located in green areas around the fountains across from the National and University Library in Zagreb. Not only were the waste bins not present at this location, but there are also signs reading "Dogs not allowed" in several places around the green area. Thanks to the continuous reports of dog owners collected on the Facebook page Dog City Zg, another 53 dog waste bins were found 313 in total (Fig. 3).

Figure 3. Various types of dog waste bins present in the City of Zagreb.

The highest number of bins recorded in the survey was in Trnje district (38). The Brezovica district was not surveyed as there was no data to indicate that dog waste bins were located there and no information from the dog owners' community to indicate otherwise. The Lower Dubrava district was the only district where all dog waste bin locations could be successfully located from government data (2) and where no other dog waste bins were present during the survey (current data). Discrepancies between data and actual numbers were noted for other districts. For five districts, all locations were confirmed using government data, with additional waste bins found, indicating an improvement in infrastructure in the districts, but also that the data on dog waste bins had not been updated. In most districts, the number of dog waste bins has stagnated or increased, with the exception of the districts of Črnomerec, Lower Town and Sesvete, where the number has decreased since the last update (Fig. 4). The City of Zagreb provides a mobile application called "Razvrstaj MojZG" to inform citizens how to dispose of different types of waste, at what times waste is collected, etc. [24]. Within the application there is an interactive map with the locations of recycling centres and so-called green islands (areas where large recycling bins for glass, plastic and cardboard waste are located). The locations of dog waste bins would be a very welcome addition in one of the future updates to the application.

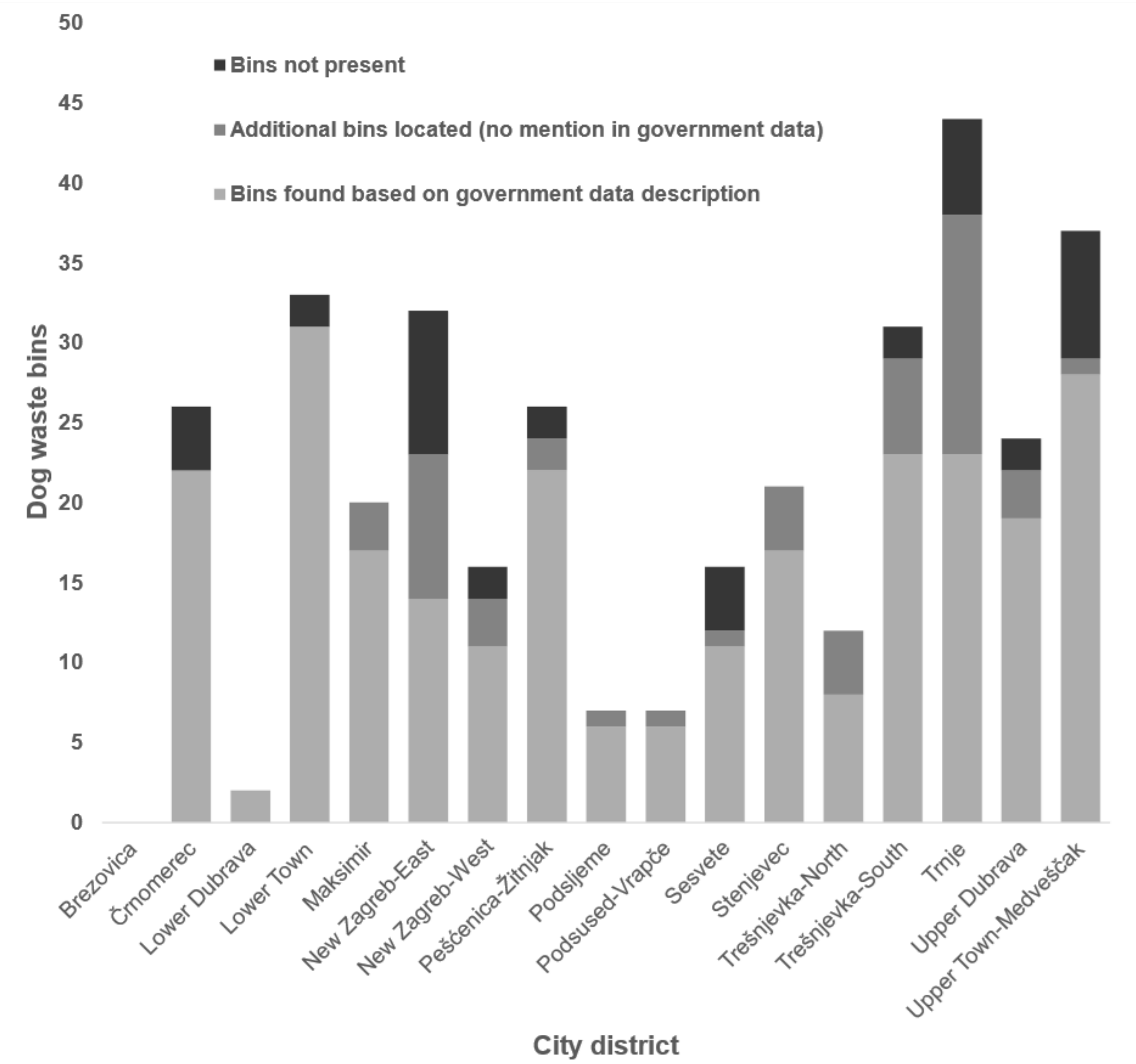


Figure 4. Dostribution of dog waste bins in the districts of Zagreb.

Field surveying revealed that there were 22 dog parks and 53 green areas where dogs can be walked off-leash. Of the 22 dog parks, 10 were listed in available government data (Fig. 5). An additional five dog parks were listed in the available government data as green spaces where dogs can be walked off-leash, rather than dog parks. The other 53 areas found were all present in the most recently found government data. Four areas from the list were clearly repurposed and the green areas were either non-existent (a school was built on one such area), fenced (Brodarski Institute), or the area was marked with "dogs prohibited" signs (the same area across from the National and University Library in Zagreb, where there were no dog waste bins). In addition, the descriptions of 13 green areas where dogs are allowed to be walked off-leash are not clear, making it difficult to easily find and accurately map them. For four areas, the descriptions are incorrect or outdated (wrong street address or the building near the area has changed its purpose and the entry has not been updated).

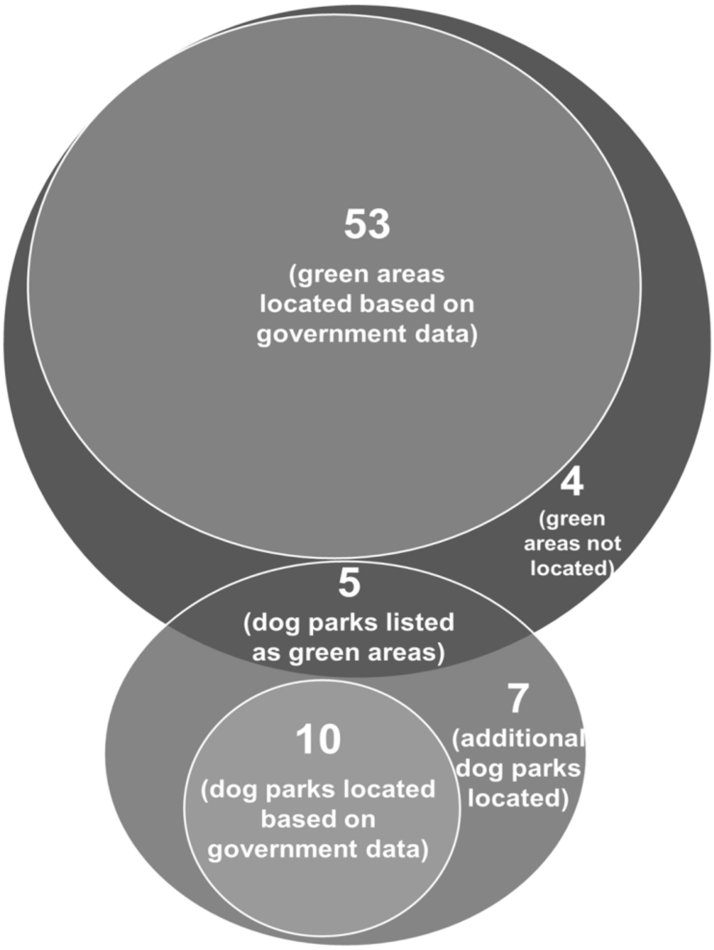


Figure 5. Comparison of government and field survey data on green spaces for dogs in the City of Zagreb.

The field survey revealed that seven out of 17 districts of Zagreb do not have a dog park, while some districts have more than one (Pešćenica-Žitnjak has a whopping 6). On the other hand, all districts except Brezovica have access to one or more green areas where dogs can be walked off-leash, which is especially important for districts with few green areas like the Lower Town district. Another major drawback of the City of Zagreb's infrastructure for dogs is that the green areas where dogs can be walked off-leash are not marked in any way by a sign (with the exception of an area in the Maksimir Forest Park, which is under the jurisdiction of the Maksimir Public Institution), and violation of the city ordinance can be punished by a fine ranging from 80 to 270 euros [22]. According to McCormack et al, the proximity of dog parks and other green areas where dogs do not have to be leashed reduces the distance dog walkers travel, suggesting that they prefer them [25]. Publicly available data on this content would direct dog walkers' attention to content that is specifically targeted towards them. Comparing government data on green areas for dogs in Zagreb with Dutch cities such as Utrecht [26], The Hague [27], or Amersfoort [28], one can see the drastic difference not only in the way the data are available and presented to citizens, but also that some local governments recognise the value of these data and classify them as high value data [29]. What Zagreb and the Dutch cities have in common is that they provide the data and interactive maps only in the local language, which was to be expected since the providers are local governments. But with increasing globalisation, local governments should also reconsider this practise, considering how many foreigners visit their cities.

Although the initial results of this research indicate that dog infrastructure in the city of Zagreb is growing and developing, the pace at which the government is updating the data available to the public about this infrastructure is both slow and insufficient (poor description of locations, general description of content present at locations). A clear evidence of this is the sharp increase from 43 green spaces designated for dogs in 2016 [30] to 72 in 2018 [22]. Another positive indication is the list of 25 projects that have been or are currently being implemented to improve urban infrastructure for dogs [31], but according to our findings, publicly available data on dog-related content in the city of Zagreb is very outdated. Several media outlets report that in some dog parks in Zagreb there are constant attempts to poison or injure the dogs by inserting nails or glass inside of meat [32-34]. Information about such activities is crucial for dog owners when choosing their dog walking routes and the parks they visit, and it would be of great benefit to the dog owners community to make this information publicly available through the city portal.

The information collected through the survey is publicly available in the form of a Google thematic map [35]. These data are not only of high value to the city of Zagreb and the dog owners living there, but also show that local government must constantly evolve to deal with dynamic, open data sets that can change frequently. Apart from that, the data presented in this study is already outdated. At the time of this study, new dog waste bins were recorded in several neighbourhoods without updated government data available online. The Google thematic map created as part of this study is constantly updated based on contributions from the public via social media and provides more up-to-date information on dog-related content in the city of Zagreb.

CONCLUSION

The results suggest that the City of Zagreb urgently needs to change the way it manages data on dog-related content so that it is more useful to its citizens. The results suggest that there is a need to update the already open datasets more frequently, enrich them with new data and missing metadata, and potentially open additional datasets that would be of great benefit to the community of dog owners in Zagreb. Improved implementation of these datasets into existing city data portals would improve the availability and reuse of this data and increase the already high value potential of this data. A mobile or web application covering all dog-related content in the city area and offering the possibility to collect data from users would also be a step towards a smart city approach to city management.

ACKNOWLEDGMENTS

This research is part of TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592.

REFERENCES

[1] Frantz, L.A.F.; Bradley, D.G.; Larson, G. and Orlando, L.: *Animal domestication in the era of ancient genomics.*
Nature Reviews Genetics **21**(8), 449-460, 2020,
http://dx.doi.org/10.1038/s41576-020-0225-0,

[2] Banks, P.B. and Bryant, J.V.: *Four-legged friend or foe? Dog walking displaces native birds from natural areas.*
Biology Letters **3**(6), 611-613, 2007,
http://dx.doi.org/10.1098/RSBL.2007.0374,

[3] Carter, S.: *Planning for dogs in urban environments*. Ph.D. Thesis.
University of Melbourne, Faculty of Architecture, Building and Planning, Melbourne, 2016,

[4] Brown, S.G. and Rhodes, R.E.: *Relationships among dog ownership and leisure-time walking in Western Canadian adults.*
American Journal of Preventive Medicine **30**(2), 131-136, 2006,
http://dx.doi.org/10.1016/J.AMEPRE.2005.10.007,

[5] Wood, L.L. and Christian, H.E.: *Dog walking as a catalyst for strengthening the social fabric of the community.*
In: Johnson, R.A., Beck, A.M. and McCune, S., eds.: *The Health Benefits of Dog Walking for Pets and People*. Purdue University Press, West Lafayette, pp.51-74, 2011,
http://dx.doi.org/10.2307/J.CTT6WQ5K0.11,

[6] Fox, M.S.: *City Data: Big, Open and Linked.*
Enterprise Integration Lab, University of Toronto, Toronto, 1-20, 2013,
http://www.eil.utoronto.ca/wp-content/uploads/smartcities/papers/City-Data-v5.pdf, accessed 15th November 2021,

[7] Gorica, K.; Kripa, D. and Zenelaj, E.: *The Role of Local Government in Sustainable Development.*
Acta Universitatis Danubius. Œconomica **8**(2), 139-155, 2012,

[8] Lev-On, A. and Rosenberg, H. *Mobile applications in local government*.
Electronic Government **17**(2), 183-198, 2021,
http://dx.doi.org/10.1504/EG.2021.114578,

[9] Graham, M. and Johnson Avery, E. *Government Public Relations and Social Media: An Analysis of the Perceptions and Trends of Social Media Use at the Local Government Level.*
The Public Relations Journal **7**(4), 1-21, 2013,

[10] Vaccari, L.; Posada, M.; Boyd, M. and Santoro, M.: *APIs for EU Governments: A Landscape Analysis on Policy Instruments, Standards, Strategies and Best Practices.*
Data **6**(6), No. 59, 2021,
http://dx.doi.org/10.3390/DATA6060059,

[11] Zdjelar, R.; Musa, A. and Žajdela Hrustek, N.: *Open data availability in Croatian local government: Improving the quality of life.*
Smart Cities and Regional Development (SCRD) Journal **5**(3), 21-40, 2021,

[12] City of Rijeka: *Open data portal of the City of Rijeka.* In Croatian.
http://data.rijeka.hr, accessed 28th February 2022,

[13] City of Zagreb: *Open data portal of the City of Zagreb.* In Croatian.
http://data.zagreb.hr, accessed 12th October 2021,

[14] City Office for the Strategic Planning and Development of the City of Zagreb: *ZG Geoportal.* In Croatian.
https://geoportal.zagreb.hr/Karta, accessed 12th November 2021,

[15] City of Zagreb: *City of Zagreb Official website*. In Croatian.
https://www.zagreb.hr, accessed 12th November 2021,

[16] Zagreb Holding: *Holding centre.* In Croatian.
https://holdingcentar.zgh.hr, accessed 12th November 2021,

[17] City Office for the Strategic Planning and Development of the City: *Statistical yearbook of the City of Zagreb for 2020*.
https://www.zagreb.hr/UserDocsImages/1/SYCZ 2020_web.pdf, accessed 12th November 2021,
[18] Hyland, B.; Atemezing, G.A.; Pendleton, M. and Srivastava, B.: *Linked Data Glossary*.
https://www.w3.org/TR/ld-glossary, accessed 9th August 2021,
[19] Varga, F.: *Dog City Zg*.
https://www.facebook.com/dogcityzg, accessed 7th September 2021,
[20] QGIS Development Team: *QGIS Geographic Information System 3.10.7*.
http://qgis.osgeo.org, accessed 1st February 2021,
[21] Zagreb Holding: *Holding centre – Subsidiary Čistoća*. In Croatian.
https://www.zgh.hr/UserDocsImages/pdf/Financijska%20izvje%C5%A1%C4%87a/ZGH_godisnje%20izvjesce_2018_www.pdf, accessed 15th December 2020,
[22] City Assembly of the City of Zagreb: *Decision of the City Assembly on the conditions and manner of keeping domestic animals and the manner and treatment of abandoned and lost animals and wild animals.* In Croatian.
http://www1.zagreb.hr/zagreb/slglasnik.nsf/7ffe63e8e69827b5c1257e1900276647/1ddf5b568d06eda0c125836a0023593e/$FILE/Odluka o uvjetima i načinu držanja kućnih ljubimaca.pdf, accessed 12th September 2020,
[23] City of Zagreb: *Pet keeping.* In Croatian.
https://www.zagreb.hr/drzanje-kucnih-ljubimaca/84089, accessed 13th January 2020,
[24] City of Zagreb: *Razvrstaj MojZG mobile application v 1.4*.
https://play.google.com/store/apps/details?id=com.razvrstajmojzg&hl=en&gl=US, accessed 3rd January 2022,
[25] McCormack, G.R.; Rock, M.; Sandalack, B. and Uribe, F.A.: *Access to off-leash parks, street pattern and dog walking among adults*.
Public Health **125**(8), 540-546, 2011,
http://dx.doi.org/10.1016/J.PUHE.2011.04.008,
[26] Municipality of Utrecht: *Maaikaart.* In Dutch.
https://gemu.maps.arcgis.com/apps/webappviewer/index.html?id=ee6544813e0447c6b3346cfa947339c0, accessed 13th December 2021,
[27] Municipality of The Hague: *Dog off-leash areas in The Hague.* In Dutch.
https://www.arcgis.com/apps/webappviewer/index.html?id=3e25e98c0e954c4fbb3978b26bf31d20&extent=71159.7306%2C449404.1941%2C89562.0674%2C459335.6139%2C28992, accessed 12th November 2021,
[28] Municipality of Amersfoort: *Amersfoort – dog card.* In Dutch.
https://ckan.dataplatform.nl/dataset/amersfoort-hondenkaart/resource/e467231b-e0c8-4c59-8ad2-c8681004f191?view_id=16d3c112-5f1a-4340-a654-d8b4c3806083, accessed 13th September 2021,
[29] Dutch government: *National data portal.* In Dutch.
https://www.data.overheid.nl, accessed 12th November 2021,
[30] Skuliber, K.: *Use of public spaces and spatial practices of dog walkers: example of Trešnjevka (Zagreb)*. M.Sc. Thesis. In Croatian.
University of Zagreb, Faculty of Science, Department of Geography, pp.1-54, 2018,
[31] City of Zagreb – local government: *City of Zagreb activities.*In Croatian.
https://aktivnosti.zagreb.hr, accessed 12th December 2021,
[32] Korljan, Z.: *Dog owners, beware! The poison was found again, but this time it was hidden in pieces of bread.* In Croatian.
https://www.jutarnji.hr/vaumijau/novosti/vlasnici-pasa-oprez-opet-je-pronaden-otrov-ali-ovoga-puta-skriven-je-u-komade-kruha-15035296?fbclid=IwAR1Af39MHgFCjQNBrSEVrNnSJd9kPB6yfPoBZHLZXkGmuqus1qR5mnx2qMI, accessed 12th December 2021,
[33] Kekuš, S.: *CAUTION Nails found in food in the dog park in Švarcova street.* In Croatian.
https://direktno.hr/zagreb/oprez-pronadeni-cavli-hrani-psecem-parku-svarcovoj-205213/?fbclid=IwAR2PWTVTbLhku8yc_jEdAR2OnmCkN-BbK1MKq3sRIyhzI5M19HQkjwtWFoY, accessed 12th December 2021,

[34] Index News: *Warning to dog owners in Zagreb: Someone intentionally leaves food with nails around.* In Croatian. https://www.index.hr/vijesti/clanak/upozorenje-vlasnicima-pasa-u-zagrebu-netko-namjerno-ostavlja-okolo-hranu-s-cavlima/2229960.aspx, accessed 12th December 2021,

[35] Varga, F.: *Dog content in the City of Zagreb*. In Croatian. https://www.google.com/maps/d/u/0/viewer?fbclid=IwAR2uJlGgmlvDVi0v_eFA7y9cS6eMmJq_kR3ysHGBSyxgZZPx3s6G8qGR95U&mid=1RvoFXSvfTwB3tqzkQvmbGo_SfAn3jIGg&ll=45.79826259494657%2C15.998241586381349&z=12, accessed 12th December 2021.

Interdisciplinary Description of Complex Systems 20(2), 136-148, 2022

THE IMPORTANCE OF OPEN DATA ACCESSIBILITY FOR MULTIMODAL TRAVEL IMPROVEMENT*

Bia Mandžuka[1,] **, Krešimir Vidović[2],
Miroslav Vujić[1] and Charalampos Alexopoulos[3]

[1]University of Zagreb, Faculty of Transport and Traffic Sciences
Zagreb, Croatia

[2]Ericsson Nikola Tesla
Zagreb, Croatia

[3]University of Aegean
Athens, Greece

DOI: 10.7906/indecs.20.2.6
Regular article

Received: 20 January 2022.
Accepted: 3 March 2022.

ABSTRACT

The dynamic urban network continues to face a number of problems caused by traffic. One of the main problems is the increasing use of personal vehicles (especially for shorter journeys) and an unattractive alternative – public transport. In this context, *Intelligent Transport Systems* can be defined as a holistic, management and information communication upgrade of the classic transport and traffic system. From the passengers' point of view, the usage of personal vehicles is still more pronounced compared to public transport. The main reason is that the public transport service quality needs to be improved if compared to the personal vehicles. The concept of *multimodal travel* is not new, but with the usage of adequate Intelligent Transport Systems services, it is possible to support and encourage modal shift, optimise the use of public space and influence passengers' behaviour patterns. *Multimodal Journey Planners* provide travellers with better and more complete information when choosing a mode of transport so they can select the most suitable option for their needs. The *open data* approach is crucial for defining a system that responds to the end-users' actual needs and aspirations (personalisation of the service). Another major challenge in providing a high-quality multimodal journey planning service is the availability and accessibility of data. EU directives require each Member State to establish a *National Access Point*. The National Access Point is a digital interface, a single/unique access point providing all information regarding travel and traffic. In this article, the importance of traffic data collection, acquisition and distribution according to the open data concept is described.

KEY WORDS

intelligent transport systems, multimodal travel, multimodal journey planners, national access point, open data

CLASSIFICATION

JEL: O18

*This is the extended version of the abstract published in: Vujić, M. and Šalamon, D., eds.: Book of abstracts of the National Open Data Conference. University of Zagreb, Faculty of Traffic and Transport Sciences, Zagreb, 2021.

**Corresponding author, η: bmandzuka@fpz.unizg.hr; +385 98 908 6171;
Faculty of Transport and Traffic Sciences, Vukelićeva 4, HR -10 000 Zagreb, Croatia

INTRODUCTION

The lifestyle of urban commuters is changing significantly due to rapid urbanisation and economic globalisation, as well as the continuous development of information and communication technologies. According to projections [1], 68 % of the world's population will live in urban areas, placing a significant burden on environmental management and existing infrastructure. One of the most important aspects of the social ecosystem is urban mobility.

Urban mobility is essential for the survival of urban agglomerations. The need for mobility in urban areas is constantly growing and is met by the use of personal vehicles. The spread of personal vehicles has increased people's mobility, but it has also had unforeseen effects on the sustainability of urban ecosystems. With the increasing level of motorisation, a number of problems are emerging that threaten the long-term viability and mobility of the population [2, 3].

Inadequate transport infrastructure, lack of space for people, limited mobility and accessibility, and negative environmental impacts are just some of the problems that need to be addressed comprehensively. The EU Urban Mobility initiatives and policies [4, 5] emphasise seamless mobility for all user groups and offer a range of attractive, efficient and environmentally friendly public transport options. Transforming urban transport towards innovative mobility solutions and services that meet the real needs of residents could be a response to existing challenges. However, it is important to stress out that the transition should start at a personal level, at the level of the users of such a system (of course, if the quality of public transport system is equal or better than the usage of personal vehicles). According to the literature, this transition has already begun and is referred to as the "socio-technical" transition [6].

By applying modern Information and Communication Technologies (ICT) in transport – specifically Intelligent Transport Systems (ITS) services, it is possible to achieve a sustainable, clean, and energy-efficient transport. ITS offers new approaches, models and technologies to solve a variety of traffic and transport problems [7].

"Multimodal mobility" is one of the possible solutions for establishing sustainable public transport and, as a result, liveable urban spaces. The approach combines different modes of transport into everyday transport routines. In other words, the concept represents the transition from owning a vehicle to using the public transport system. In this regard, the transition to sustainable mobility is seen as a major challenge in the upcoming decades, aiming to eliminate or at least mitigate the negative effects currently caused by traffic [6, 8, 9].

In order to simplify multimodal journeys and overcome barriers (waiting time, transfers, switching to another mode of transport), it is necessary to offer users a service that will cover the journey from point A to point B. This is especially important for modern users who are more inclined to use technology (technophiles) [10]. To achieve the quality and efficient functioning of such systems, modern technologies (smartphones and other digital gadgets) need to be used. In addition, data and interoperability for cross-border journey planners play an important role in the quality provision of travel planning services. In this sense, specific data quality requirements and appropriate data exchange protocols are prerequisites for implementing travel planning services [11, 12].

The aim of the LinkingDanube project (Interreg – Danube Transnational Program) was to encourage the provision of transnational, interoperable and Multimodal Journey Planners (MJPs) covering urban and rural areas. The project "Coordination mechanisms for multimodal cross-border traveller information network based on OJP for Danube Region (OJP4DANUBE) explores the possibility of personalised cross-border travel choices based on a distributed architecture [13, 14].

In recent years, opening data has become easier due to sensor technology integrated into our urban environment. In that matter, opening up is cheaper and easier than building a new transport infrastructure. The project "TODO – Twinning Open Data Operational", approved by the Horizon2020 programme – Twinning, is important in terms of interdisciplinarity and raising awareness of the importance of open data approach [15].

According to EU Directives, each Member State is required to create a National Access Point (NAP), which will be a single/unique access point based on open data [16]. Furthermore, data security and privacy concerns are prevalent, particularly when collecting data and providing personalised multimodal information for multimodal journeys [17].

The purpose of this article is to emphasise and elaborate the importance of open data in the implementation of advanced ITS solutions such as multimodal journey planning services. This article aims to provide an analysis of the state- of- the- art research on multimodal travel, multimodal journey planners, and the open data concept. The key steps for the realisation of multimodal journey planners based on open data are presented in the final section.

THE CONCEPT AND NEED OF MULTIMODAL TRAVEL

There are many strategies that focus on shifting from personal vehicles to environmentally friendly travel solutions. The Avoid-Shift-Improve strategy (Germany) is a holistic approach to creating a sustainable transport system. The strategy is based on three pillars: "Avoid/ Reduce", "Shift/Maintain" and "Improve". According to [18], the first pillar refers to improving the overall efficiency of the transport system, minimising the length of trips and the need for personal vehicle. The second pillar is based on the idea of improving individual travel by focusing on the use of alternative transport modes (from personal vehicle to public transport). This is particularly important in promoting multimodal travel. This includes walking and cycling, the most environmentally friendly options, but also public transport (lower CO_2 emissions per passenger-kilometre compared to personal vehicles). The third pillar, "Improve", means, among other things, increasing the attractiveness of public transport.

The question of establishing sustainable urban mobility is so complex because a rethinking and a new understanding of sustainable urban mobility must be initiated at the individual level. When it comes to choosing a means of transport, the personal vehicle still dominates because it meets the need for flexibility and independence.

In order to achieve sustainable travel behaviour, public transport must meet the needs of the individual and work towards modal shift, i.e., the transition from the personal vehicle to sustainable modes of transport. In this sense, it is crucial to know users' reasons and triggers (at the individual level) for switching to more sustainable transport.

In [19], it is stated that the increasing number of passengers in urban areas has encouraged new travel options (in addition to traditional ones), with multimodal travel being a promising option. This research was conducted to address the research gap between the obvious/standard criteria (that guide passengers in choosing a trip) and the actual needs of passengers. In addition to some key determinants of travel mode choice that are consistent with the findings of previous analyses, five new key findings for urban transport were identified. Particular attention is paid to the perception of sustainability, which passengers perceive as a cost-related element of urban travelling.

In other words, urban travellers increasingly perceive the pollution associated with the use of certain modes of transport as a cost that impacts the environment. Furthermore, the newly identified perceptual dimension - individualisation - is particularly highlighted. This perceptual dimension indicates that the urban traveller has to adapt to the daily changing situations during his journey. Technological possibilities allow travellers to adapt to different travel scenarios. This is particularly important for new forms of urban transport, such as the multimodal travel concept. Through the use of advanced information and communication technologies, it is possible to achieve a high degree of autonomy and individualisation [19].

The concept of multimodal travel is one of the potential solutions to mitigate the negative impact of transport on the environment. The approach is not new, but it is gaining new characteristics and possibilities especially due to the fast-growing digitalisation, new mobility models and business models ("sharing culture" - car sharing, bike sharing). In the literature, the concept is defined as a door-to-door journey that provides continuous connectivity throughout the journey and uses a combination of transport modes (walking, tram, bus, e-scooter, etc.), Figure 1. Due to its complexity, such a travel concept must ensure easy switching between modes and flexibility in combining modes. Only in this way it can compete with the personal vehicle [20-22].

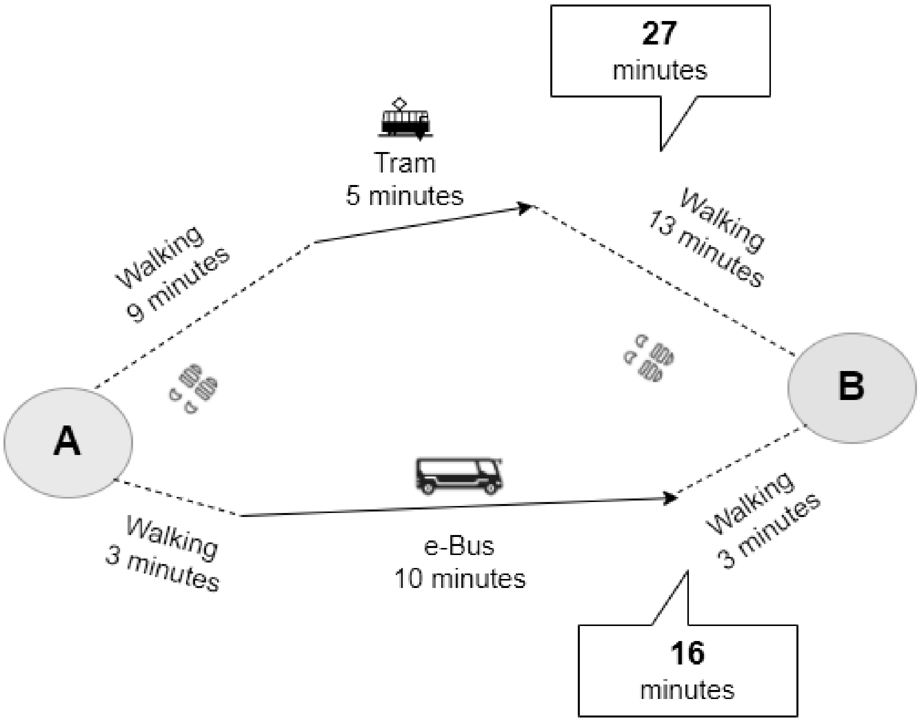


Figure 1. Example of Multimodal Journey Routes (edited by authors) [23].

According to the authors [19], one of the main advantages of the multimodal travel concept is the possibility of individualisation and autonomy compared to unimodal travel (by personal vehicle), where one chooses one mode of transport to get from point A to point B. This hybrid concept can also partially cover the need to adapt to the travel scenario and thus compete with the personal vehicle (in terms of flexibility and autonomy).

To optimise the operations of urban areas, the paradigm of “smart cities” requires sustainable solutions to ensure economic prosperity and social well-being. In this context, the concept of “Mobility-as-a-Service” (MaaS) is a new way of thinking and a smart solution for implementing effective and sustainable urban mobility. It provides a transport solution based on user preferences. The platform combines different types of mobility (public transport, car-sharing, car and bike rental, taxi and a combination of the above) on one platform. This concept offers a number of opportunities that are beneficial to the urban mobility ecosystem. In addition to the ticketless model, MaaS generates vast amount of traffic data that can help transport planners eliminate bottlenecks and determine the actual needs of passengers. In other words, it is possible to identify and understand passenger behaviour, which will ultimately help to implement transport services that meet passengers' needs. Open data is the basis for establishing user-centric services such as MaaS. This is the only way to empower travellers to change their travel behaviour and improve the overall travel experience and quality of life in general [24-26].

MULTIMODAL JOURNEY PLANNERS

In specific scenarios of multimodal travel, there are some obstacles, such as lack of information for certain modes of transport (transfers, waiting times), lack of personalisation etc. To achieve the quality and efficient functioning of such systems, it is necessary to use advanced ITS services. This area includes static and dynamic information about the transport network, pre-trip and on-trip information services, and support for services that collect, store and manage data for planning transport activities. Pre-trip information services are of great importance to users as they allow them to plan a journey from home or any other location where internet access is available. In addition to pre-trip information, access to information during the trip is crucial for users to make a timely decision about their trip [7]. Specifically, the Advanced Traveller Information System (ATIS), as an integral part of the above-mentioned functional area, aims to support the traveller in multimodal travel planning. Because of the increasingly complex requirements of modern users, ATIS plays a significant role in intelligent mobility [27-29].

On-trip information includes real-time travel information, estimated travel time depending on existing traffic conditions, parking availability, traffic accidents, etc. MJPs represent the integration of these services and, as such, are a key component of implementing ITS in cities today. They offer a comprehensive range of relevant information to assist the user in planning a door-to-door journey [7].

MJPs are defined as guidance and navigation services by the ITS architecture and the ISO 14813-1 specification [7]. MJPs enable travel planning for end-users based on the collected and processed traffic/transport data. The user can combine different (available) travel options depending on the current state of the transport network and his own needs. MJPs provide better modal integration and more sustainability by allowing travellers to choose the most appropriate combination of transport modes. This could lead to increased use of public transport, cycling or walking in urban areas [30]. Figure 2 shows an example of a multimodal journey planner for a selected destination, suggesting four routes with different combinations of travel modes.

The basic task of the multimodal planning service is to answer the user's question: "How do I get from point A to point B at a certain departure/arrival time and under what conditions?" [31]. Today, MJPs are mainly based on algorithms generating the shortest route/travel time. The complexity of creating a model that generates options tailored to passengers is reflected in the diversity of their preferences.

In urban areas, multimodal information is of great importance as it increases the use of public transport and other 'healthy' ways such as walking and cycling. Multimodal travel information promotes mobility for all user groups, especially users with disabilities or reduced mobility, by providing information about facilities and assistance at transport interchanges. A special type of user of MJPs are tourists. They may be guided by additional criteria when choosing a travel route, e.g., choosing a route that includes different Points of Interests (POIs) [30, 31, 33].

Citymapper is one of London's best public transport planning tools [34] and is generally applicable to many European and American cities. The service offers multimodal travel planning

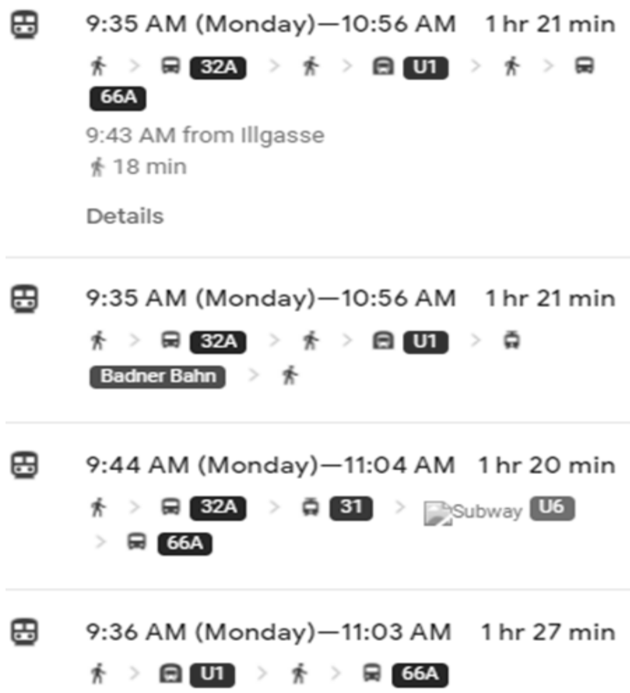


Figure 2. Example of multimodal routes generated by MJP [32].

with a variety of features that enhance the user's experience and satisfaction. The service provides step-by-step instructions during the trips, which is especially important for multimodal trips due to the complexity of the system (switching to another transport mode, waiting, etc.). The travel planning service displays tariffs for available modes of transportation, and includes platforms such as Uber, Lyft, City Bike, GREENBike, and others. In addition, the application allows users to personalise their profiles by storing favourite routes, stops, and other information. The user can subscribe to a specific mode to receive information for selected modes. The location can be shared via social networks and selected profiles/contacts.

TripGo is a travel planning service developed under the MaaS concept. The planner has been highly rated by experts and it is widely used in the EU. As an additional, partially personalised feature, the service offers the possibility to plan a daily schedule of activities including trips for the day. The system suggests the most convenient routes for the user based on that plan [35].

Personalised MJPs are a step forward in the field of travel information. In a dynamic urban context, such a multifunctional system integrates a variety of functions required for planning journeys that meet travellers' needs. The end-user is at the centre of such an advanced information system. There are a number of active multimodal planners in the EU with different functionalities. Most include basic functions/criteria for selecting a multimodal route, and only a few offer some degree of personalisation [35-37].

Travel data is also personal data because it reveals much more than just the user's movement [17]. In recent years, various technologies have provided efficient and cost-effective sources of data collection. The number of smartphone users in the world has been on the rise, resulting in the largest source of data on human mobility. MJPs, which help users plan their trips, are also sources for collecting large amounts of travel data. For each user and travel scenario, such a system tracks mobility patterns based on preferences. Additionally, it contains data that is entered into the system when setting up a user profile (if the platform offers this option). These are usually some of the basic preferences of end-users.

THE IMPORTANCE OF OPEN DATA IN THE IMPLEMENTATION OF ADVANCED ICT SERVICES

Open data (OD) could be described as data being available with no financial and legal restrictions [38]. OD represents available resources that can be integrated into innovative digital solutions [39]. The resources provided at the national level need to comply with the OD principles (e.g., G8 Open Data Charter) and to satisfy the principles of maturity [40]. Furthermore, the OD ecosystem is a broader concept including many interrelated perspectives with closely connected interdependencies and reflecting on different aspects of open data [41].

The concept of OD is strongly associated with innovative capacity and transformative power [42]. It is increasingly recognized that proactively opening public data can create considerable benefits for several stakeholders, such as firms and individuals interested in the development of value-added e-services or mobile applications, by combining various types of OGD, and possibly other private data, or scientists, journalists and active citizens who want to understand better various public problems and policies through advanced data processing and production of analytics [43].

The benefits of OD are common in all initiatives considering a socio-economic focus for both, private and public sector, and, according to [44] four different types of values can be distinguished: (1) efficiency values, related to the higher efficiency of public sector bodies as a direct result of the availability of data coming from other public organisations, (2) innovation values, where the use of OGD (Open Government Data) can create new economic profit for the public sector, (3) transparency values, which contribute to the growth of public trust in government and reduction of corruption, and (4) participation values, where private sector gains social benefits through cooperation with the public sector. Pfenniger et al. [45] derived with OD benefits for science: (1) Improved quality of science in order to meet fundamental scientific principles like traceability, reproducibility and transparency, (2) More effective collaboration across the science-policy boundary in terms of better and more transparent results through data quality and data validation processes and (3) Increased productivity through collaborative burden-sharing avoiding unnecessary duplication and learning from each other.

The above benefits are almost impossible to implement without ensuring availability and then proper accessibility to the data. Only providing data will not necessarily lead to a generation of new value [46]. To achieve this, some efforts must be made in data quality, a rise of awareness among users about the benefits of open data and development of mechanisms to exploit data potential. At last, OGD success does not depend only on data itself but on a functioning ecosystem where two sides, data availability and data demand, are building on each other [38]. During the past 10 years, several OD assessment methods have been created. However, the outcome of such assessments is no guarantee whatsoever that a simple question a citizen may have can be answered by the open data. By checking the availability and proper accessibility of data for a specific domain is the first step towards the realisation of innovative services.

In establishing open data, technical (interoperability, inconsistent open standards, data silos) and social barriers (open data ethics, equity, engagement) are still present [17]. The benefits of open travel data should be seen as a win-win situation. By opening and reusing data, the quality, efficiency and transparency of public transport services increase. It helps operators to organise passenger transport effectively and to offer innovative mobility services that meet the needs of modern users [47]. On the other side, users can be provided with a personalised travel information service. Discovering preferences from large (open) data sets plays a key role in user profiling and in establishing personalisation.

For example, if the discovery of user preferences for the student population were included, then more credible data on their travel behaviour would be obtained (preferred modes of transport to certain locations, POIs, etc.). The analysis can encourage travel planners and other experts to understand better and predict travel demand (e.g., where, when and how to travel) and assess the impact on the transport network and society.

For this reason, it is necessary to act in the direction of raising awareness and highlighting the benefits of open travel data. With open access, it is possible to increase the availability and attractiveness of public transport. It also allows operators to understand better the real needs of users based on their mobility patterns. However, there is still some mistrust in the ethical use of data. When data is shared and used to combine openness and innovation while based on ethical principles, it has significant global potential for society and the economy [17, 47].

NATIONAL ACCESS POINT FOR MULTIMODAL INFORMATION

Without the openness and availability of data, MJPs would not be relevant to end-users (especially when it comes to personalising the service). According to the EU Directive, each Member State will set up its own NAP, which will be a single/unique access point based on open data.

National Access Point for Multimodal Information (NAPMM) is defined as a digital interface where the static travel and historical traffic data, together with the corresponding metadata, are made accessible for reuse or where the sources and metadata of these data are made accessible for reuse [48]. Delegated Regulation 2017/1926 [49] defines that multimodal travel information services are based on both static and dynamic travel and traffic data, where ‘dynamic travel and traffic data’ stands for data related to different transport modes that often change on a regular basis, and ‘static travel and traffic data’ stands for data related to different transport modes that do not change at all or change seldom and on a regular basis. The document also recommends that travel information services should use the European Technical Specification entitled ‘Intelligent Transport Systems – Public Transport – Open API for distributed journey planning 00278420’ currently under finalisation when performing distributed journey planning. When service providers establish handover points for distributed journey planning, such handover points should be listed in the national access point. In technical terms, NAPMM is based on an information-communication system (ICT solution).

NAPMM is intended to be the main open data repository for multimodal travel planning, where relevant stakeholders will deposit their static and dynamic data. NAPMM involves several groups of stakeholders, as presented in Figure 3. Stakeholders either provide data, corresponding metadata information on the quality of the data accessible to users and end users. Stakeholders providing data include transport authorities, transport operators, transport on-demand service providers and infrastructure managers. Users and end-users can use this data as an enabler for transportation-related services and to make travel decisions.

NAPMM covers various modes of transport, including air, rail including high-speed rail, conventional rail, light rail, long-distance coach, maritime including ferry, metro, tram, bus, trolleybus, shuttle bus, shuttle ferry, taxi, car-sharing, car-pooling, car-hire, bike-sharing, bike-hire, and personal modes. The static travel data, among others, include location search data of origins and destinations and access nodes, trip plans, access nodes and trip plan computations for both scheduled modes of transport and for road transport, special fare query for all scheduled modes, information service for all modes, trip plans and trip plans computation.

Dynamic travel and traffic data includes passing times, trip plans and auxiliary information. Information service regarding publicly accessible charging stations for electric vehicles and refuelling points, and availability and future predicted road link travel times. The opening of

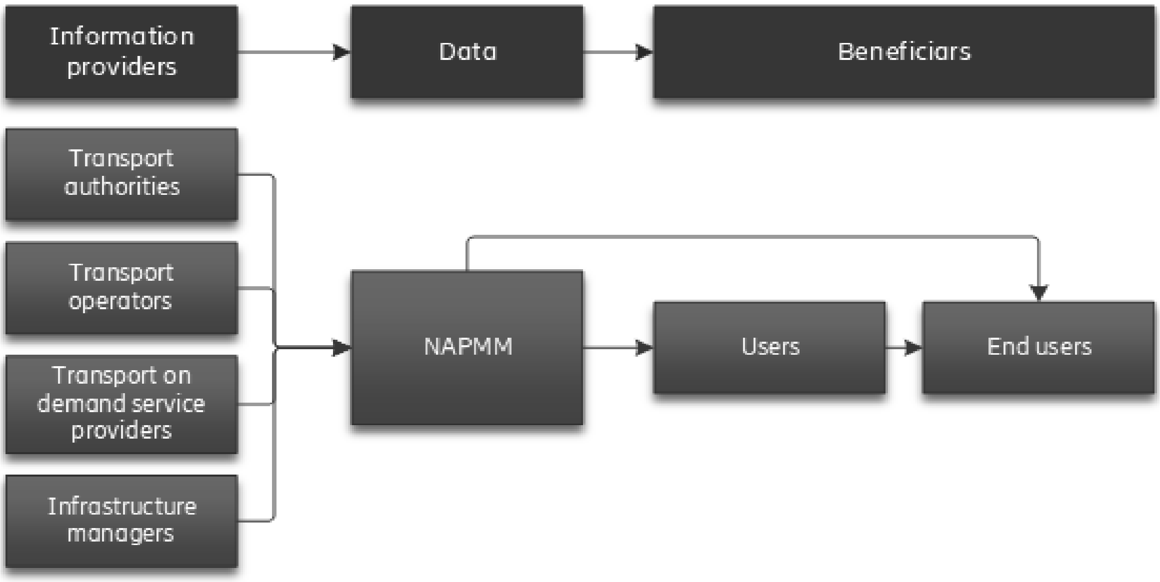


Figure 3. Data flow and stakeholders of NPTMM.

transport data is an ongoing process that has been implemented for the last decade with the goal to enable access to better information that would offer an improved service to users. NAPMM helps mobility players publish their data and make them accessible on a single platform, which facilitates their reuse by travel information services [11].

CONCLUDING REMARKS

Urban mobility, as a leading concept in urban traffic sustainability, is one of the main concerns of the development of modern traffic and transport networks. With the rapid increase of personal vehicles and the growing population in urban areas, the existing traffic network cannot fulfil the user needs for daily migrations within the urban traffic network. A logical alternative should be encouraging the usage of the public transport system and/or other sustainable modes of transport. In order to achieve that, it is necessary to improve the quality of the public transport system, or at least, some specific and crucial parts of it. Traveller information is the best and easiest ITS service that end-users see and use in their daily migration travels.

When choosing a mode of transport, end-users mostly rely on the travel time from origin to a destination, not taking into account the other benefits of the modal transition. Multimodal travel as a concept is the most suitable alternative for personal vehicle replacement because it satisfies all the end-user requirements. In this article, the importance of opening raw traffic data is described, which allows service developers to broaden their offer of products to end-users.

Multimodal journey planners are a complete service that provides all travel information to end-users in one place, so the openness of raw traffic data is essential. The first step is to define the essential traffic data that can be used for multimodal journey planners service development. The second step is to point out the importance of opening those datasets and to provide a platform (both for end-users and other stakeholders) for data collection, analysis, and dissemination/distribution. The final step should be the creation of one complete multimodal journey planner service, which should be developed in real-time according to the present traffic situation and the most suitable mode of travel according to user needs and preferences.

To achieve a better quality of traveller information distributed to users, the implementation of the open data concept is essential. Since the urban traffic system generates a large amount of traffic data (which must be collected, analysed and distributed), the openness of such data is essential not only for end-users but also for other stakeholders (traffic and transport operators, value-added service operators and distributors, etc.). Currently, multimodal information services across Europe are generally not interoperable and are fragmented in terms of what they offer, including modal and geographical coverage, real-time information, and quality levels. Therefore, a next step and focus on the future period would be on harmonisation of data and exchange of data among NAPMM from various countries and regions.

ACKNOWLEDGMENT

This research is part of TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592. In addition, part of the research activities is related to the OJP4Danube project “Coordination mechanisms for multimodal cross-border traveller information network based on OJP for Danube Region” within the Danube Transnational Program.

REFERENCES

[1] United Nations: *World Urbanization Prospects: The 2018 Revision.*
https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf,

[2] Tomaszewska, E.J. and Florea, A.: *Urban Smart Mobility in the Scientific Literature – Bibliometric Analysis.*
Engineering Management in Production and Services **10**(2), 41-56, 2018,
http://dx.doi.org/10.2478/emj-2018-0010,

[3] Ribeiro, P.; Dias, G. and Pereira, P.: *Transport Systems and Mobility for Smart Cities.*
Applied System Innovation **4**(3), No. 61, 2021,
http://dx.doi.org/10.3390/asi4030061,

[4] European Commission (Directorate-General for Mobility and Transport): *Sustainable urban mobility: European policy, practice and solutions.*
http://dx.doi.org/10.2832/51274,

[5] European Commission (European Platform on Sustainable Urban Mobility Plans): *Guidelines for Developing and Implementing a Sustainable Urban Mobility Plan 2019.*
https://www.eltis.org/sites/default/files/guidelines_for_developing_and_implementing_a_sustainable_urban_mobility_plan_2nd_edition.pdf, accessed 18th December 2021,

[6] Spickermann, A.; Grienitz, V. and vod der Gracht, H. A.: *Heading towards a Multimodal City of the Future?*
Technological Forecasting and Social Change **89**, 201-221, 2014,
http://dx.doi.org/10.1016/j.techfore.2013.08.036,

[7] Bošnjak, I.: *Intelligent Transport Systems 1.* In Croatian.
University of Zagreb, Faculty of Transport and Traffic Sciences, Zagreb, 2006,

[8] Van Nes, R.: *Design of Multimodal Transport Networks, A Hierarchical Approach.* Ph.D. Thesis.
Delft University Press, Delft, 2002,

[9] Schuppen, J.; Kettner, S.; Delatte, A. and Schwedes, O.: *Urban Multimodal Travel Behaviour: Towards Mobility without a Private Car.*
Transportation Research Procedia **4**, 553-556, 2014,
http://dx.doi.org/10.1016/j.trpro.2014.11.042,

[10] Nikolaidou, A. and Papaioannou P.: *Social media and transport choices: How social media can affect tripmakers’ choices.*
WIT Transactions on Ecology and The Environment **226**, 605-616, 2017,
http://dx.doi.org/10.2495/SDP170531,

[11] Vidović, K.; Mandžuka, S. and Šoštarić, M., eds.: *National Access Point for Provision of EU-wide Multimodal Travel Information Services.*
MIPRO 2019 – 42nd International Convention. Rijeka, 2019,

[12] Vidović, K.; Mandžuka, S. and Šoštarić, M., eds.: *Data Quality within National Access Point for Provision of Multimodal Travel Information within European Union.* 2019 International Symposium ELMAR, Zadar, 2019, http://dx.doi.org/10.1109/ELMAR.2019.8918828,

[13] Linking Danube – Interreg project: *Linking transnational, multimodal traveller information and journey planners for environmentally-friendly mobility in the Danube Region.* https://www.interreg-danube.eu/approved-projects/linking-danube, accessed 20th December 2021,

[14] OJP4Danube – Interreg project: *Coordination mechanisms for multimodal cross-border traveller information network based on OJP for Danube Region.* https://www.interreg-danube.eu/approved-projects/ojp4danube, accessed 20th December 2021,

[15] Twinning Open Data Operational – TODO project: *Project "TODO – Twinning Open Data Operational", approved by Horizon2020 programme – Twinning.* https://todo-project.eu/en/about-project, accessed 27th December, 2021,

[16] European Commission: *Delegated Acts of Directive 2010/40/EU.* https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32010L0040&from=EN, accessed 28th December, 2021,

[17] Open data Institute: *Personal data in transport: exploring a framework for the future.* https://theodi.org/wp-content/uploads/2018/06/OPEN-Personal-data-in-transport-.pdf, accessed 28th December, 2021,

[18] Jonuschat, H.; Stephan, K. and Schelewsky, M.: *Understanding Multimodal and Intermodal Mobility.* Sustainable Urban Transport (Transport and Sustainability) **7**. Emerald Group Publishing Ltd., Bingley, pp.149-176, 2015, http://dx.doi.org/10.1108/S2044-994120150000007018,

[19] Clauss, T. and Döppe, S.: *Why do urban travelers select multimodal travel options: A repertory grid analysis.* Transportation Research Part A: Policy and Practice **93**, 93–116, 2016, http://dx.doi.org/10.1016/j.tra.2016.08.021,

[20] Guo, H. et al.: *An Innovative Framework for Customised Intermodal and Multiobjective Trip Planner.* International Journal of Intelligent Transportation Systems Research **15**(2), 117-125, 2017, http://dx.doi.org/10.1007/s13177-016-0126-y,

[21] Solar, A. and Marqués, A.: *ENHANCED WISETRIP: Wide Scale Multimodal and Intelligent Journey Planning.* Procedia – Social and Behavioral Sciences **48**, 2940-2949, 2012, http://dx.doi.org/10.1016/j.sbspro.2012.06.1262,

[22] Boero, M., et al.: MyWay Personal Mobility: *From Journey Planners to Mobility Resource Management.* Transportation Research Procedia **14**, 1154-1163, 2016, http://dx.doi.org/10.1016/j.trpro.2016.05.186,

[23] Cheung, E. and Sengupta, U.: *Analysis of Journey Planner Apps and Best Practice Features.* http://www.complexurban.com/wp-content/uploads/2016/11/WP12.6.1-MMU-Report.pdf, accessed 20th December 2021,

[24] Arias-Molinares, D. and García-Palomares, J.C.: *The Ws of MaaS: Understanding mobility as a service from a literature review.* IATSS Research **44**(3), 253-263, 2020, http://dx.doi.org/10.1016/j.iatssr.2020.02.001,

[25] Hensher, D.A.; Mulley, C. and Nelson, J.D.: *Mobility as a Service (MaaS) – Going Somewhere or Nowhere?* Transport Policy **111**, 153-156, 2021, http://dx.doi.org/10.1016/j.tranpol.2021.07.021,

[26] Wong, Y.Z.; Hensher, D.A. and Mulley, C.: *Mobility as a Service (MaaS): Charting a Future Context.*
Transportation Research Part A: Policy and Practice **131**, 5-19, 2020,
http://dx.doi.org/10.1016/j.tra.2019.09.030,

[27] Vujic, M.; Cavar, I. and Bosnjak, I.: *Distribution of travel information and driver's perception.*
In: Proceedings of 14th International Symposium on Electronics in Traffic – ISEP, Ljubljana, 2006,

[28] Kem, O.; Balbo, F. and Zimmermann, A.: *Traveler-Oriented Advanced Traveler Information System Based on Dynamic Discovery of Resources: Potentials and Challenges.*
Transportation Research Procedia **22**, 635-644, 2017,
http://dx.doi.org/10.1016/j.trpro.2017.03.059,

[29] Carrese, F., et al.: *A Framework for Dynamic Advanced Traveler Information Systems.*
Future Transportation **1**(3), 590-600, 2021,
http://dx.doi.org/10.3390/futuretransp1030031,

[30] Gentile, G. and Nökel K., eds. *Modelling Public ransport Passenger Flows in the Era of Intelligent Transport Systems.*
Springer International Publishing, London, 2016,

[31] EuTravel: *D1.3 Technology Knowledge Base and Observatory.*
http://www.eutravelproject.eu/uploadfiles/D1.3_Technology%20Knowledge%20Base%20and%20Observatory_final.pdf, accessed 9th January 2022,

[32] Google Commerce Ltd. *Google Maps App.*
https://www.google.com/maps/place/Be%C4%8D,+Austrija/@48.2206635,16.3098491,12z/data=!3m1!4b1!4m5!3m4!1s0x476d079e5136ca9f:0xfdc2e58a51a25b46!8m2!3d48.2081743!4d16.3738189, accessed 9th January 2022,

[33] Mandžuka, B.; Brčić, D. and Škorput, P.: *Application of multimodal travel guides for urban and suburban travel.* In Croatian.
In: Šakić, Ž., ed.: 34th Conference on Transportation Systems with International Participation – Automation in Transportation 2014. KoREMA, Zagreb, 2014,

[34] CityMapper Ltd.: *All your Transport.*
https://apps.apple.com/us/app/citymapper-all-your-transport/id469463298, accessed 9th January 2022,

[35] Esztergár-Kiss, D.: *Framework of aspects for the evaluation of multimodal journey planners.*
Sustainability **11**(18), No. 4960, 2019,
http://dx.doi.org/10.3390/su11184960,

[36] Zhang, J.; Arentze, T. and Timmermans, H.: *Making our mobility more intelligent: A framework of a personalized multimodal traveller information system.*
In: Proceedings of the 12th World Conference on Transport Research. Lisbon, 2010,

[37] Arentze, T.A.: *Adaptive Personalized Travel Information Systems: A Bayesian Method to Learn Users' Personal Preferences in Multimodal Transport Networks.*
IEEE Transactions on Intelligent Transportation Systems **14**(4), 1957-1966, 2013,
http://dx.doi.org/10.1109/tits.2013.2270358,

[38] Welle Donker, F. and van Loenen, B.: *How to assess the success of the open data ecosystem?*
International Journal of Digital Earth **10**(3), 284-306, 2017,
http://dx.doi.org/10.1080/17538947.2016.1224938,

[39] Mensah, E. and Goderre, J.L.: *Data Sources and Data Tools: Preparing for the Open Data Ecosystem*.
In: Magnuson, J. and Dixon, B., eds.: *Public Health Informatics and Information Systems*. Health Informatics. Springer, Cham, pp.105-127, 2020,
http://dx.doi.org/10.1007/978-3-030-41215-9_7,

[40] European Data Portal: *Open Data Maturity Report.*
https://data.europa.eu/sites/default/files/edp_landscaping_insight_report_n6_2020.pdf, accessed 15th January 2022,
[41] Zuiderwijk, A.; Janssen, M. and Davis, C.: *Innovation with open data: Essential elements of open data ecosystems.*
Information Polity **19**(1-2), 17-33, 2014,
http://dx.doi.org/10.3233/IP-140329,
[42] Davies, T.; Perini, F. and Alonso, J.M.: *Researching the emerging impacts of open data. ODDC conceptual framework.*
http://www.opendataresearch.org/sites/default/files/posts/Researching%20the%20emerging%20impacts%20of%20open%20data.pdf, accessed 15th January 2022,
[43] Charalabidis, Y.; Alexopoulos, C. and Loukis, E.: *A taxonomy of open government data research areas and topics.*
Journal of Organizational Computing and Electronic Commerce **26**(1-2), 41-63, 2016,
http://dx.doi.org/10.1080/10919392.2015.1124720,
[44] Jetzek, T.; Avital, M. and Bjørn-Andersen, N.: *Generating Value from Open Government Data.*
In: Baskerville, R. and Chau, M., eds.: Proceedings of the 34th International Conference on Information Systems – ICIS, 2013,
[45] Pfenninger, S., et al.: *The importance of open data and software: Is energy research lagging behind?*
Energy Policy **101**, 211-215, 2017,
http://dx.doi.org/10.1016/j.enpol.2016.11.046,
[46] Janssen, M., Charalabidis, Y. and Zuiderwijk, A.: *Benefits, Adoption Barriers and Myths of Open Data and Open Government.*
Information Systems Management **29**(4), 258-268., 2012,
http://dx.doi.org/10.1080/10580530.2012.716740,
[47] Stone, M. and Aravopoulou, E.: *Improving journeys by opening data: the case of Transport for London (TfL).*
The Bottom Line **31**(1), 2-15, 2018,
http://dx.doi.org/10.1108/BL-12-2017-0035,
[48] European Parliament and EC: D*directive 2010/40/EU of the European Parliament and of the Council of 7 July 2010 on the Framework for the Deployment of Intelligent Transport Systems in the Field of Road Transport and for Interfaces with Other Modes of Transport.*
https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32010L0040, accessed 11th January 2022,
[49] European Commission: *Regulations – Commission Delegated Regulation (EU) 2017/1926.*
https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32017R1926&from=EN, accessed 15th January 2022.

Interdisciplinary Description of Complex Systems 20(2), 149-161, 2022

OPEN ELECTION DATA: EVIDENCE FROM CROATIA IN A COMPARATIVE PERSPECTIVE*

Petra Đurman[1, **], Dario Nikić Čakar[2] and Davor Boban[2]

[1]University of Zagreb, Faculty of Law
Zagreb, Croatia

[2]University of Zagreb, Faculty of Political Science
Zagreb, Croatia

DOI: 10.7906/indecs.20.2.7
Regular article

Received: 24 January 2022.
Accepted: 25 April 2022.

ABSTRACT

This article explores the concept of open election data, as a specific type of institutional open data. Transparency of electoral procedures, as the most fundamental democratic process, is crucial for the legitimacy of democratic political systems. By providing detailed information on electoral processes in open formats for the re-use of the general public, open election data provide an additional democratic dimension for contemporary democracies. The aim of this article is to assess the state of open election data comparatively and in Croatia. The analytical findings suggest that the availability of open election data in most of the countries included is rather limited in scope, with significant cross-national and within-country variations. Numerous countries make only election results, political party and candidate registration lists and polling station information available in open formats, while other types of election data cannot be accessed in machine-readable forms.

KEY WORDS

open data, open election data, transparency, electoral process, Croatia

CLASSIFICATION

JEL: C89, D72

*This is the extended version of the abstract published in: Vujić, M. and Šalamon, D., eds.: Book of abstracts of the National Open Data Conference. University of Zagreb, Faculty of Traffic and Transport Sciences, Zagreb, 2021.

**Corresponding author, η: pdurman@pravo.hr; -;
University of Zagreb, Faculty of Law, Trg Republike Hrvatske 14, HR – 10 000 Zagreb, Croatia

INTRODUCTION

Open data are one of the most salient developments in e-government and e-participation. The availability of data to everyone in an open and machine-readable form, free of charge, represents a specific mechanism for achieving government transparency, which goes much further in accomplishing the values of open government than traditional transparency, which refers to the accessibility of information, regardless of its form. The re-use of open data for commercial or non-commercial purposes also promotes participatory government, because users (i.e. the public – individuals, NGOs, private businesses, media, academia, etc.) constitute a critical element in generating the final outcome of the data (re)use, including different applications, sophisticated business product based on open data (e.g. legal information portals, business portals) and scientific research and analyses.

The potential benefits of different categories of open data have already been well documented in the literature [1, 2]. Different types of institutional and political data – such as data on state organisations and public sector authorities, their functioning, election data and similar – are particularly important for the democratic legitimacy of politico-administrative system. Because the transparency principle represents a *conditio sine qua non* for the democratic electoral process, information on different aspects of electoral organisations and processes – such as data on election and referendum results, campaign financing, electoral management bodies or voter lists – constitute a crucial element for providing the legitimacy and citizens' trust in politico-democratic processes. Accurate, complete and good quality open election data (OED) can enhance electoral integrity and accountability by providing detailed information on electoral processes not only to selected stakeholders, but also to the public in general, enabling them to make informed decisions[1] [3-5]. In addition, primary users of OED such as journalists and scientists can reuse the data for scientific and professional analyses, predictions, interpretations and similar.

This article represents an exploratory study with a purpose of assessing the state of OED in Croatia, from a comparative and national-specific perspective. To do so, we first elaborate OED as a specific type of institutional open data, after which we consider the theoretical relevance of election data for contemporary political systems. In the methodological part of the article, Croatia is compared to other EU member states and the UK with respect to the main OED indicators, followed by in-depth analysis of OED ecosystem in Croatia, including the regulatory framework, types of OED available, features of the portal/website, data provider and users. The applied research method included desk research and content analyses of Internet documents, portals and official websites.

OED AS A TYPE OF INSTITUTIONAL OPEN DATA

Although there is no single, unanimously accepted categorisation of open data types, they do not significantly differ. As identified in one of the earlier categorisations, the main types of open data include business, geographic, legal, meteorological, transport and social data[2] [7; p.14]. Within the category of social data – which includes different statistical data, such as economic, employment, health and population – specific types of institutional and political/public administration data can be extracted. These encompass data on different organisational and functional aspects of politico-administrative organisations and other public sector authorities at different levels of government (state, local, regional). This includes data on electoral processes, public officials, public sector bodies (e.g. lists of public sector authorities, register of national minorities' councils) and their functioning (e.g. schedule of government meetings). The openness of this type of data is particularly relevant for strengthening the procedural legitimacy of politico-administrative systems.

In general, electoral process in a democracy refers to all procedures and activities related to legally defined appointment of public officials and public bodies by voters. Electoral process, therefore, consists of procedures conducted before, during and after election day. The basic sources of election information are legal acts and other official documents regulating elections[3]. They primarily include information on the type of elections (e.g. parliamentary, presidential, local, European), the type of electoral system and its characteristics, constituencies, candidates, electoral lists, judicial and constitutional court decisions on elections and provisional and final results. Although a normative framework that contains electoral law as well as judicial practice does not represent electoral data in a narrower sense, they can be publicly available in open formats and in an easy searchable way via specialised portals or official websites. Therefore, the electoral framework is considered a component of electoral data [3, 8]. Election data include re-usable information on the pre-election process (campaign financing, voter and candidate registration, polling stations), the election process itself (e-voting and counting, voter lists) and the results of the election process and post-election actions (publishing results, complaints). We refer to OED only when they are published in an open, machine-readable format[4].

Election data are, in general, collected (and provided) by the central (state) organisations which organise and conduct the elections (Electoral Management Boards – EMBs, Central Election Commission or similar management bodies). In some countries, civil society associations have taken the role of data providers, in addition to their role as data users and mediators[5] [3; p.210]. Election data can be published on the official websites of EMBs, open data portals or third-party websites/portals, as in the case of election databases published by international organisations/associations. Primary users include scientists, media, journalists, electoral observers and agencies, who produce electoral predictions, analyses, explanations and interpretations of electoral processes and results based on OED. A secondary user of OED is the general public, whose use of the data is not in-depth, but rather related to information, education and socialisation, as such data facilitate familiarisation with the organisation, implementation and results of democratic procedures. The role of the public is, however, particularly emphasised in the context of OED. Namely, alongside being open data users in the ‘outcome’ dimension of the electoral process, citizens are involved as participants – active (candidates) and/or passive (voters) – during the elections as the most fundamental democratic process. Privacy issues are, in general, not very problematic in the case of election results, due to secret ballot for voters and public political candidacy. However, it is an issue of considerable concern in the case of voter registration and, especially, the e-voting process[6].

Principles for OED do not differ from the standards of open data in general. According to the Open Election Data Initiative, election data are open when they are: (i) timely (available as quickly as necessary for it to be useful); (ii) granular (primary, raw data which are not in an aggregate or modified form); (iii) available for free on the Internet (available without any monetary restrictions and easy to locate); (iv) complete and in bulk (all data are contained in a file so that the entire dataset can be obtained in one download); (v) analysable (available in digital, machine-readable form); (vi) non-proprietary (open, non-restrictive formats over which no entity has exclusive control – e.g. CSV, XML and JSON)[7]; (vii) non-discriminatory (available to any individual or organisation without limitations based on user identity; anonymous access to the data); (viii) license-free (open for re-use and redistribution for any purpose)[8]; and (ix) permanently available (permanent URL, portal or online archive) [13].

In practice, some serious limitations can be found when it comes to the type, availability and quality of OED. First, OED are often reduced to election results. Data related to voter registration and election results are the most common type of election data published, while information on political/electoral financing is more rarely available. Second, election results, as well as other types of election data, are not always published in open formats. Publishing election results as images instead of open file formats impedes their re-use and diminishes their democratic potential. Varying formats for official results also represent a barrier for their usability [3; p.210, & p.213, 4; p.8]. Third, whether published in machine-readable formats or not, the comprehensiveness and consistency of election data are always an important issue. Inconsistent retention of records is therefore an important obstacle for the openness of election data. With respect to all mentioned aspects, the existing literature points at significant variations, not only between different countries, but also within them.

THE RELEVANCE OF OED FOR POLITICAL SYSTEMS

Over the last two decades, the principles of government transparency and openness have become inherent concepts of contemporary governance and decision-making processes. They represent political values, referring to the availability of different government information to the public (transparency) and the possibility for the public to provide feedback information to the government (openness) [14]. Growing requests for government ‘opening’ towards the public can be attributed to recent developments and doctrines in public administration, such as good governance, which has been strongly advocated by international organisations (e.g. OECD, United Nations, European Union). However, government secrecy is always perceived by the public as suspicious [15], even if the government functions regularly and efficiently. Proactive provision of different government data to the public, on the other hand, implies there is nothing to hide. It enables the public to hold government officials accountable, thereby promoting citizens’ trust and the legitimacy of politico-administrative institutions and actors. Transparency is primarily achieved via right-to-know regulation and its instruments, such as open meetings, media reports, publishing documents, registers and databases [14, 16, 17].

The importance of election data, as a type of institutional data, stems from the elemental importance of the election process in a democratic political system. The transparency of election results, the data on financing political campaign, the composition of electoral management bodies and similar information represents a necessary precondition for the public to be motivated to participate in election processes. Availability of such information in open formats, containing comprehensive and accurate data, provides additional democratic value as well as practical benefits for the government and the public. On the one hand, it widens the circle of potential open data users, because detailed election information is not reserved for selected stakeholders only (such as important media), but is available to the general public, including civil society organisations, journalists, election observers, scientists and individuals in general [3; p.210]. As a result, the general public can better understand the election process, which has effects on procedural legitimacy, and make informed decisions, which concerns the outcome legitimacy. On the other hand, media, journalists and scientists – who represent the primary users of OED – are enabled to reuse the data for scientific and professional analyses, predictions and interpretations, which can produce very practical benefits. For instance, analyses based on OED can reveal interesting findings, factor interrelations or shortcomings of the election process (such as the quality across districts regarding population size, manipulation of electoral district boundaries and the like), which can then be useful input for policymakers to improve existing regulation and/or practices.

COMPARATIVE ANALYSIS

Although discussion of the benefits of OED has been gaining salience, ‘relatively little election data is published according to open-data principles’ [3; p.213]. Regarding the type of data, election results are the most commonly published OED, with other types of election data being less commonly publicly available, especially data on political and electoral financing. Some technical deficiencies of OED include the duration of data availability, the granularity of available data, restrictive licensing, non-machine-readable data formats and registration and authentication requirements for data users [3; p.213]. In methodological terms, cross-national comparisons of different OED are burdened by language barriers, because the name of the dataset is usually officially available in the native language only. Different databases, mostly published by academia and scientific organisations, contain comparative OED and are very useful, although most often reduced to election results.

The Open Election Data Initiative lists 16 categories of election data that can be published as open data, covering all activities of the electoral process in the pre-election period, moving to the election day and ending in the post-election period [3; pp.212-213]. In this part, we selected seven key categories which represent the core of the democratic election process and applied them to the analysis of availability of OED in 27 EU member states plus the UK[9] (see Table 1). We opted for the EU case-selection framework mostly because it represents the most comprehensive open data legislative initiative and also because of the high democratic standards to which its member states must adhere.

To analyse the availability of OED in the EU context, we focused our investigation primarily on two sources – the open data portals of EU member states and the official websites of national electoral management bodies. By doing so, we aimed to explore not only the impact of 'external' factors (EU legislation) on opening election data in an individual member state, but also the 'internal', country-specific state of the art when it comes to the question of availability of election data in open formats. Election data classified as open is published in formats such as CSV, JSON and XML, while data available in PDF, JPG and similar formats was not categorized as open. We also limited the scope of the analysis to only parliamentary elections in each country, mostly because the state-wide general elections to the representative bodies are seen as first-order elections, while other levels of election are seen as of less importance (local, regional, European).

Table 1 shows that, in general, the availability of OED in most of the countries included in the analysis is rather limited in scope. A large number of countries make only election results, political party and candidate registration lists and polling station information available in open formats. The availability of other categories of election data in open format is rather scarce, especially those relating to voter lists (access to detailed information about eligible voters), election campaigns (availability of timetables for campaigns) and electoral complaints (number of complaints and the outcomes of conflict resolution). In eight countries, election data are not published in any open format, but are rather available in other online forms which are not machine-readable.

There is also no consistency with regard to platforms where OED are published. In most cases, OED can be accessed on the official websites of electoral management bodies, while a smaller portion of election data, particularly election results, is downloadable from national open data portals. It is also possible to find OED on other websites as well, which implies that OED are scattered around the Internet instead of being kept available in timely and permanently manner at one central spot. In some instances, it took great effort to trace the final location where OED are published.

When turning attention away from general cross-national observations to the country-level perspective, we can report several cases of good practice, among which Lithuania, Romania and United Kingdom stand out. This is particularly true for Lithuania, with its Central Electoral Commission publishing a large variety of election data in machine-readable formats.

Table 1. Open election data in 27 EU member states and in the UK.

	Election results	Pol. party/ candidate registration	Campaign finance	Election campaigns	Voter lists	Polling stations	Electoral complaints
Austria	*						
Belgium							
Bulgaria	*/**					*/**	
Croatia	*/**		**			**	
Cyprus							
Czech Republic	*	*				*	
Denmark	**						
Estonia	*/**					*/**	
Finland	**	**					
France							
Germany	*/**	**				*	
Greece							
Hungary	*						
Ireland	*	*					
Italy	**						
Latvia	*/**	*/**				*	
Lithuania	**	**	**			**	**
Luxembourg							
Malta							
Netherlands	*/**	*					
Poland	**	**				**	
Portugal	*/**						
Romania	**	**	**			**	
Slovakia	*/**	**					
Slovenia	**						
Spain							
Sweden	*/**						
United Kingdom	**	**	**			**	

Remark: empty cells refer to no open data.
*Open Data Portal
**Electoral Administration

Data on election results, voter registration, candidates and the financing of political campaigns are systematically organised, easily searchable and accessible and cover the whole period since the introduction of Lithuanian democracy in 1990. On the other hand, there are also several cases of bad practice in publishing OED. For instance, Ireland and Austria have published results of some parliamentary elections which are not complete and in bulk. It is possible to find the results of the Austrian 2019 parliamentary elections at the open data portal, but only for the region of Upper Austria and not for the rest of the country. Furthermore, there are also cases like Spain and Portugal, which have a significant amount of election data available at the official websites of central election management bodies, but it can be downloaded only as PDF documents.

From a comparative perspective, it is important to notice that new democracies in Central and Eastern Europe (CEE) are apparently doing much better in terms of the 'openness' of election data than their 'older' counterparts in Western Europe. Table 1 shows that democratic latecomers in CEE such as Lithuania, Romania, Latvia, Poland and Croatia have OED available on a much larger scale than well-established democracies like France, Belgium and Luxembourg, without any OED published, or Austria, Denmark, and Sweden, with only election results published in open formats. This observation is somewhat puzzling when taking into consideration the differences in the level of politico-economic and democratic development between these two groups of countries, so further research should be conducted to provide a plausible explanation for these differences.

Finally, cross-national comparison of OED is indeed overburdened by language barriers, because native language versions of websites and published data are the norm. There were only few cases in our research for which fully functional English version of websites and data are available, which makes comparison difficult. On the other hand, there are several election databases containing different election results in open formats from numerous countries and these are a valuable source of OED for cross-national comparisons. For instance, The European Election and Referendum Database [18] provides election results on a regional level for European countries and publishes the results of parliamentary elections, EP elections, presidential elections and EU-related referendums for 35 European countries. The Constituency-Level Elections Archive (CLEA) [19] offers a dataset with detailed election results at the constituency level for lower and upper chamber legislative elections from around 170 countries. The Global Elections Database [20] provides data on the results of national and subnational elections around the world, with data available in various open formats. The ParlGov project [21] covers 37 EU and OECD democracies, offering data on about 1700 parties, 1000 elections and 1600 cabinets, with election results available in machine-readable files. The Political Data Yearbook [22] is published on behalf of the European Consortium of Political Research and covers 'election results, national referendum, changes in government, and institutional reforms for a range of countries, within and beyond the EU'. Data are available in CSV and XLSX formats.

OED IN CROATIA

The previous comparative analysis shows that Croatia is doing very well in comparison to other European countries, regardless of whether they are new democracies in CEE or well-established democracies in the West. In this section, we take a bird-eye snapshot of OED in Croatia, exploring other components of OED alongside indicators compared in the previous section. We provide more detailed description of available types of OED in Croatia, including their quality and providers, as well as legal framework. In addition to these 'provision' elements, we assess the 'outcome' dimension of OED as well, i.e. the users of OED[10]. These elements are commonly referred as wider environment of open data, i.e. open data ecosystem [1] and represent areas or sub-areas of indicators within different assessment frameworks (e.g. in Open Data Maturity Report, policy dimension – encompassing countries' open data policies and strategies, impact dimension - referring to open data re-use, government policies and government action within readiness in Open Data Barometer). For the purpose of this analysis, we rely on Open Data Maturity Report results as a general referential benchmark, although these findings encompass open data in general, not a specific category such as OED[11].

In Croatia, systemic regulation for open data is in place within the Law on the Right to Access Information [24], which transposes the PSI Directive (as it is the case with other EU member states who had to transpose the Open Data Directive into their national laws) (see [25]), postulating that each public body must ensure that the data are published on the internet and that is easily findable and machine-readable. In addition, a formal Open Data Policy (Politika otvorenih podataka) was adopted in 2018 by the Croatian Government as a strategic direction for further development of public administration openness, although without adopting a strategy or action plan for implementing the policy. According to the Open Data Maturity Report in policy dimension, Croatian score is slightly above the EU average – 87 %. Specific regulations referring to OED can also be found – the reports on campaign financing have to be published in open and machine-readable formats on the official website of Croatian EMB, in accordance with the Law on financing political activities, electoral campaigns and referendum [26]. Other electoral regulations do not refer to the openness of electoral data.

The types of available OED include election and referendum results, financing of political activities and campaign financing and the list of polling stations. The most extensive category is certainly election and referendum results, which encompass open data on presidential elections, parliamentary elections, elections for the European Parliament (EP), local elections, elections for national minorities' councils and representatives and data on the referenda (national, local and consultative). National election data comprise data from 2000 onwards (presidential elections are held every five years and parliamentary elections every four years); local elections data are available from 2013 (local elections take place every four years), as well as are European elections data (first elections for the European MPs held in 2013). Referendum data include data on two national referendums held in 2012 and 2013, the first one on Croatian EU membership and second one on the constitutional definition of the marriage.

The owner and provider of OED is the State Electoral Commission (Državno izborno povjerenstvo – DIP), which publishes data on election results on its official website, data on election campaign financing and regular financing of political activities and the list of polling stations [27]. Election results in open formats can be found on the national Open Data Portal [28] as well. All election data available on the Portal as well as on the DIP website are accessible without registration and free of charge. As in the case of many other types of open data in Croatia, a nation-specific open license is applied, which is substantially equivalent to the CC-BY license. OED are easily findable through a general Google search and orderly structured on the DIP website, while the Open Data Portal is easy searchable by filtering the type of data/publisher.

Regarding data quality, election results available on the DIP official website and Open Data Portal are available in CSV and excel file formats, financial reports on political activities and campaign financing are available in PDF and JSON, while the list of polling stations can be downloaded in an excel file format. Regarding content, general elections results are consistent since the 2000, albeit with some deficiencies related to the count of spoiled votes and individual vote counts for the representatives of national minorities. The data on constituencies are not available in open file format, while the data on polling stations are in open format but not integrated (available by constituency). Voter registration information is not publicly available. Because election datasets are static, uploading is not very frequent. The quality of metadata is one of the weak points of available OED, with metadata missing on the official DIP website and very scarce description of the election results datasets on the Open Data Portal. Feedback options include sending an email to the data provider via official website and an option to indicate an error and provide suggestion via the Open Data Portal.

There are three main types of primary users of OED in Croatia. First, there are journalists, media and public opinion agencies who interpret and present the data to the wider public via different media channels and whose activities are mostly concentrated within the timeframe of a specific election. Second, there are scientists and researchers (mostly political scientists), who permanently use election data for scientific research and analyses. Third, different types of NGOs use OED for their own analysis and policy papers, which are further used for policy advocacy purposes or the education of the general public. For instance, the NGO GONG regularly publishes various guides and manuals in the field of electoral politics, but is also focused on civic education with regard to citizens' electoral rights and understanding of electoral processes [29]. For European elections in 2019 and local elections in 2021, in cooperation with agency which provides IT support for the Croatian public sector, DIP launched applications for tracing election results and for monitoring the functioning of polling station committees.

Although Croatia scores high on Portal usage according to the latest Open Data Maturity Report (130/160), OED do not seem to be attractive type of open data for individual users. Statistics on the use of datasets is not published on the OD Portal nor on the DIP official website, but we can assume that such usage is rather low. The results of a survey conducted within the TODO project at the Faculty of Law, University of Zagreb, revealed a very low level of faculty employees' familiarisation with the concept and benefits of different types of open data and, considering the rather small academic community, we can assume that the same applies for political scientists as well.

CONCLUSION

Despite numerous social advantages and the positive impact on citizens' trust, education and overall legitimacy of the politico-administrative system, the potential of OED has not yet been accomplished in most of the countries [3]. The observation of Yang et al. [30], that OED is largely an emerging area, remains valid. On the one hand, some of the front-running countries in open data in general, such as Austria and Spain, are lagging behind when it comes to the 'opening' of election data, while open data 'followers', such as Romania and Croatia, are doing much better in providing OED. On the other hand, there are countries like Spain and Portugal that publish very extensive amounts of election data, but not in an open format. For researchers, this implies the necessity for further research on such discrepancies between the countries, while for practitioners (providers), it calls for more systemic opening of election data. However, the problem may be in the 'demand' side of the sub-ecosystem; OED appear not to be as attractive for individual users as geospatial, meteorological and some other institutional data or data concerning current policy issues/problems (such as data on the COVID-19 pandemic). The level and impetuses of election data re-use (and open data re-use in general) in Croatia should be investigated in further research.

From a comparative perspective, Croatia is among the countries that have made significant strides in making election data available in machine-readable formats. A good practice in publishing comprehensive electoral results on behalf of DIP is recognised as an example of increasing transparency and accountability in the 2021 Open Data Maturity Report. Still, more types of election data could be published in an open format (e.g. candidate registrations, election campaigns, electoral complaints) and the quality of existing metadata could be improved. However, in a number of OED categories, Croatia scored better than a significant number of other countries usually identified as front-runners in general open data initiatives or are seen as 'old' democracies that are much more inclined to transparency practices (e.g. France, Spain).

Finally, we also argue that the significant differences observed between new democracies in CEE and old democracies in western and southern Europe could also pave the way for further research on OED. For instance, one could explore the factors behind the higher levels of availability of OED in emerging democracies in the post-communist world. Can these differences be explained by the process of accession of these countries to the EU and the fostering of the EU conditionality policy? Are there any region-specific factors that have made CEE countries the frontrunners in opening election data? One possible explanation might be that this is a result of infrastructure development (i.e. older democracies already had long-term structures in place for handling election data, while newer democracies did not, which made it possible for them to start from a greater level of openness). These are puzzling research questions that require special attention in future research.

ACKNOWLEDGMENT

This research is part of TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592.

REMARKS

[1]The Open Data Maturity Report for 2021 stated that 'an increase in the impact of open data on transparency and accountability was observed, where 74% of the Member States define the impact as high, 11 % as medium, and 7 % as a low' [6].

[2]This categorisation is in accordance with the list of priority areas for open data as identified by the European Commission (Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information; repealed in 2019 by the Directive 2019/1024 on open data and the re-use of public sector information). The list of thematic categories for high-value datasets, as referred to in Article 13(1) of the Directive, includes geospatial, earth observation and environmental, meteorological, statistics, companies and company ownership and mobility data.

[3]The most important sources of electoral law are the constitution and electoral laws. The constitution of a country generally states only the most important electoral principles, such as that suffrage is universal and equal and that elections are secret and mostly direct, the age at which active and passive suffrage is acquired for a particular type of election and elected state or supranational authorities. Sometimes the electoral principle according to which elections must be conducted can be included in the constitution (e.g. the proportional principle in the Czech constitution), and in some cases a specific type of electoral system can be constitutionalised as well (e.g. individual transferable voting in the Irish constitution). However, most electoral matters are left to the legislator to regulate by individual laws, which often have the status of organic laws (i.e. a qualified majority of votes is needed for their adoption in the parliament). Electoral law is usually not regulated by single, but rather by several acts (as is the case in Croatia).

[4]This, in general, applies to democracies where elections are free and fair. For non-democratic countries – that is, those with totalitarian, authoritarian, hybrid and other undemocratic regimes where elections are not free and fair, or at least unfair – election data probably do not reflect the actual will of the citizens expressed in the elections, but are often fabricated in favour of regime candidates or electoral lists. Therefore, to analyse the election system and election data in a particular country, it is necessary to consider the type of political regime as well as the history of elections, before drawing conclusions on the credibility of election data.

[5]An example is the non-profit project *OpenElections* which, during the 2018 general elections in the United States, converted and published official precinct-level election results in an open format. Until then, great variations between the states existed and under half of the states had election results in usable formats [4].

[6]On e-voting see [9, 10].

[7]XLS and DOC file formats are, for example, proprietary formats owned by Microsoft. PDF was previously also a proprietary format, until Adobe released PDF as an open, non-proprietary standard in 2008 [10].

[8]Regarding licenses, there is considerable difference between the United States, where the licensing for election data is not seen as necessary nor desirable, in accordance with the understanding that government data are free as they are produced within the public domain, and European countries, where licenses are commonly used by the government to make the data available for everyone [12].

[9]We decided to include the UK as well, since it only recently left the EU.

[10]With regard to the reuse of OED, we rely on basic insights based on desk research and available benchmarks, since more systemic research should be conducted in that respect.

[11]According to the Open Data Maturity Report for 2021, Croatia has been placed within the category of 'followers' (scoring 74–86 %), alongside with Finland, Sweden, Greece, Bulgaria, Latvia, Romania and Czech Republic. In relation to the previous year, this represents a decrease in open data maturity level, when Croatia's score was ranked within the category of 'fast-trackers' [6]. More on open data in Croatia in [23].

REFERENCES

[1] Charalabidis, Y.; Zuiderwijk, A.; Alexopoulos, C.; Janssen, M.; Lampoltshammer, T. and Ferro, E.: *The World of Open Data: Concepts, Methods, Tools and Experiences*.
Springer, Cham, 2018,

[2] van Loenen, B., Vancauwenberghe, G., Crompvoets, J. and Dalla Corte, L.: *Open data exposed*.
In: van Loenen, B.; Vancauwenberghe, G. and Crompvoets, J., eds.: *Open Data Exposed*, T.M.C. Asser Press, pp.1-11, 2018,
http://dx.doi.org/10.1007/978-94-6265-261-3_1,

[3] Wolf, P.: *Open Data in Electoral Administration*.
TUT Press Proceedings of the Second International Conference on Electronic Voting, TUT Press, Tallin, pp.209-219, 2017,

[4] Willis, D.; Merivaki, T. and Ziogas I.: *Election Data Transparency: Obtaining Precinct--Level Election Returns*.
Public Integrity **24**(2), 1-17, 2021,
http://dx.doi.org/10.1080/10999922.2021.1883854,

[5] Chen, C.-C.; Chen, W.-S. and Lin, Y.-J.: *The Current Development of Open Election Data in Taiwan*.
Proceedings of the 3rd Multidisciplinary International Social Networks Conference on Social Informatics,

[6] Publications Office of the European Union: *Open Data Maturity Report for 2021*.
https://data.europa.eu/sites/default/files/landscaping_insight_report_n7_2021.pdf, accessed 30th December 2021,

[7] Dekkers, M.; Polman, F.; Velde, R. and de Vries, M.: *Measuring European Public Sector Information Resources MEPSIR. Final Report of Study on Exploitation of public sector information benchmarking of EU framework conditions*.
http://opendatatales.com/wp-content/uploads/2019/02/Dekkers-et-al.-2006-Measuring-European-Pub
lic-Sector-Information-Resou.pdf, accessed 15th September 2021,

[8] Open Election Data Initiative: *Section 3: Key Election Process Categories*.
https://openelectiondata.net/en/guide/key-categories, accessed 15th December 2021,

[9] Krimmer, R.: *The evolution of E-voting: Why voting technology is used and how it affects democracy.* Ph.D. Thesis.
Tallinn University of Technology, Tallinn, 2012,

[10] Manojlović Toman, R.: *Internet voting as a mechanism for strengthening citizen participation.* In Croatian.
In: Koprić, I. and Staničić, F., eds.: *Referendum and direct democracy in Croatia.* In Croatian. Institut za javnu upravu, Zagreb, pp.175-198, 2021,

[11] Open Election Data Initiative: *Section 2: Open Election Data Principles. Principle 6: Election data is open when it is in a non-proprietary format*.
https://openelectiondata.net/en/guide/principles/non-proprietary, accessed 15th December 2021,

[12] Open Election Data Initiative: *Section 2: Open Election Data Principles. Principle 8: Election data is open when it is license-free*.
https://openelectiondata.net/en/guide/principles/license-free/, accessed on 15th December 2021,

[13] Open Election Data Initiative: *Section 2: Open Election Data Principles*.
https://openelectiondata.net/en/guide/principles/, accessed on 15th December 2021,

[14] Musa, A.: *Information for citizens: transparent and open public administration towards better governance and trust of citizens.* In Croatian.
In: Koprić, I.; Musa, A. and Giljević, T., eds.: *Citizens, public administration and local self-government: trust, cooperation, support.* In Croatian. Institut za javnu upravu, Zagreb, pp.29-72, 2017,

[15] Pusić, E.: *The relationship between citizens and government.* In Croatian.
In: Public administration – teaching materials. Društveno veleučilište u Zagrebu & Pravni fakultet, Zagreb, pp.25-33, 2006,

[16] Piotrowski, S.J. and Van Ryzin, G.G.: *Citizen attitudes toward transparency in local government.*
The American Review of Public Administration **37**(3), 306-323, 2007,
http://dx.doi.org/10.1177/0275074006296777,

[17] Halberstam, M.: *Beyond Transparency: Rethinking Election Reform from an Open Government Perspective*.
Seattle University Law Review **38**, 1077-1067, 2015,

[18] Norwegian Centre for Research Data: *European Election Database*.
https://o.nsd.no/european_election_database, accessed 15th December 2021,

[19] –: *The Constituency-Level Elections Archive*.
https://electiondataarchive.org, accessed 15th January 2022,

[20] –: *Global Elections Database*.
http://www.globalelectionsdatabase.com, accessed 15th January 2022,

[21] –: *ParlGov project*.
https://www.parlgov.org, accessed 15th January 2022,

[22] –: *Political Data Yearbook*.
https://politicaldatayearbook.com, accessed on 15th January 2022,

[23] Musa, A.; Đurman, P. and Hadaš, P.: *Open data of public administration and local self-government*. In Croatian.
In: Koprić, I. and Staničić, F., eds.: *Referendum and direct democracy in Croatia.* In Croatian. Institut za javnu upravu, Zagreb, pp.273-313, 2021,

[24] Law on the Right to Access Information. Consolidated text of the law NN 25/13, 85/15. In Croatian.
https://www.zakon.hr/z/126/Zakon-o-pravu-na-pristup-informacijama, accessed 10th November 2021,

[25] Musa, A.: *Open data and information re-use: What does the Open Data Directive bring to citizens and the public sector?* In Croatian.
Informator (6604), 2-12, 2019,

[26] –: *Law on Financing Political Activities, Election Campaigns and Referendums NN 29/19, 98/19*. In Croatian.
https://zakon.hr/z/1957/Zakon-o-financiranju-politi %C4 %8Dkih-aktivnosti %2C-izborne-promid %C5 %BEbe-i-referenduma, accessed 10th November 2021,

[27] –: *State Election Commission of the Republic of Croatia*. In Croatian.
https://www.izbori.hr/site, accessed on 10th November 2021,

[28] –: *Open data portal*. In Croatian.
https://data.gov.hr/ckan/dataset?organization=drzavno-izborno-povjerenstvo&_organization_limit=0, accessed 10th November 2021,

[29] GONG: *GONG*.
https://gong.hr, accessed on 8th November 2021,

[30] Yang, H.-C., et al.: *Toward Construction of Open Election Data Model: A Case Study of Taiwan's Election Data*.
Proceedings of the 3rd Multidisciplinary International Social Networks Conference on Social Informatics, 2016.

Interdisciplinary Description of Complex Systems 20(2), 162-178, 2022

FRAMEWORK FOR FEDERATED LEARNING OPEN MODELS IN E-GOVERNMENT APPLICATIONS*

Emanuel Guberović[1,] **, Charalampos Alexopoulos[2], Ivana Bosnić[1] and Igor Čavrak[1]

[1]University of Zagreb, Faculty of Electrical Engineering and Computing
Zagreb, Croatia

[2]University of the Aegean
Athens, Greece

DOI: 10.7906/indecs.20.2.8
Regular article

Received: 16 January 2022.
Accepted: 25 April 2022.

ABSTRACT

Using open data and artificial intelligence in providing innovative public services is the focus of the third generation of e-Government and supporting Internet and Communication Technologies systems. However, developing applications and offering open services based on (open) machine learning models requires large volumes of private, open, or a combination of both open and private data for model training to achieve sufficient model quality. Therefore, it would be beneficial to use both open and private data simultaneously to fully use the potential that machine learning could grant to the public and private sectors.

Federated learning, as a machine learning technique, enables collaborative learning among different parties and their data, being private or open, creating shared knowledge by training models on such partitioned data without sharing it between parties in any step of the training or inference process. This paper provides a practical layout for developing and sharing machine learning models in a federative and open manner called Federated Learning Open Model. The definition of the Federated Learning Open Model concept is followed by a description of two potential use cases and services achieved with its usage, one being from the agricultural sector with the horizontal dataset partitioning and the latter being from the financial sector with a dataset partitioned vertically.

KEY WORDS

e-Government, open data, machine learning, federated learning open model

CLASSIFICATION

JEL: D80

*This is the extended version of the abstract published in: Vujić, M. and Šalamon, D., eds.: Book of abstracts of the National Open Data Conference. University of Zagreb, Faculty of Traffic and Transport Sciences, Zagreb, 2021.

**Corresponding author, η: emanuel.guberovic@gmail.com; -; -

INTRODUCTION

Utilization of Internet and Communication Technologies (ICT) by various governments worldwide to supply its citizens and other interested parties a whole new plethora of capabilities centered around the data and services that fall within its domain is known as electronic government (or shorthand e-Government). Six distinct governance properties experience improvements by using e-Government activities, including quality of public services, administrative efficiency, open Government (OG) capabilities, ethical behavior and professionalism, trust and confidence in government, and social value and well-being [1]. Three different generations of e-Government [2], differ in their final goals and essential ICT tools used to achieve them. The first generation focuses on informational and transnational services through ICTs and web technologies. The second generation focuses on improving openness and interoperability through web 2.0 concepts. Finally, the third generation aims to achieve innovative governance by exploiting disruptive technologies such as artificial intelligence.

Different governments utilize a plethora of varying open data policies [4], with a good potential found in all of them embracing further openness in increasing the participation and interaction of open data consumers and producers, resulting in positive results such as stemming corrupt behavior [5]. Simultaneously, there is a pronounced sense of privacy in personal data sharing resulting in numerous data protection regulations and acts appearing in recent years. The United Nations Conference on Trade and Development's (UNCTAD) publication on Data Protection regulations and international data flows [6] analyzed data protection laws that were current in the year 2016 (in e.g. GDPR [7]). It concluded a recognized set of core data protection principles in binding international and regional agreements and guidelines, including a limited and secure collection of personal data. Their enforcement poses a challenge to artificial intelligence usage because many of its applications owe their successful implementation to personal data used in training and inference of the models. Adherence to their requirements is a logical next step in the evolution of the implementation of machine learning in the cohabitation of ethical computing and intelligent services, as privacy is found to be one of the ethical guidelines for artificial intelligence [8].

In recent years, a new machine learning technique called federated learning (FL) has helped the field of artificial intelligence to abide by data privacy regulations. Standard machine learning aggregates data from different sources on a central server, the model training process takes part. The central learning principle partakes with different dataset instances firstly being aggregated on a single central point; in this way, the central dataset can be perceived as a per data source partitioned data shards database. On the other hand, FL is based on the distributed learning principle. Each data owner partakes in the training process with their local data shard. This process emphasizes transferring model parameters between respective data owners instead of sharing their data. Because data never leaves the data source, this method is private by design. FL is a machine learning method that elevates knowledge derived from one instance by aggregating individual latent values extracted through the training process of the crowd or multiple instances.

As an ICT e-Governance tool of the third generation, it allows using new technologies for accomplishing crowd intelligence that supports data-wise and evidence-based public services.

RELATED WORK

THIRD GENERATION E-GOVERNMENT

The primary objective of the research and practice in the domain of Digital Government (DG) is the exploitation of ICT in government and the provision of ICT-based services to their potential users: citizens, private and public companies, as well as public servants. However, the change in needs (and expectations) of citizens and societies also mandates the evolution in capabilities offered by ICT - not merely restricted to increase in performance and the number of services offered, but by shifting the focus of DG - thus driving the evolution of the digital government domain itself. Two major factors influence the evolution of the DG domain; the first one is defined by the wider external environment (economic, social, and political), and the second one by its technological environment. Nevertheless, a common pattern can be identified when observing evolutions in the DG domain; the first step preserves the existing practices, processes, and services and merely automates/supports them through existing or innovative ICT. Only in the second step the existing practices and processes are incrementally transformed and/or completely new practices adopted, usually through incremental ICT-based improvements introduced by the government [9].

Big Data generated by the Internet of Things (IoT) and Open Government Data (OGD) movement, Blockchain Technologies (BCT), Artificial Intelligence (AI), and particularly Machine Learning (ML) algorithms are some of the technologies used for modernizing the previous services provided by all of the governments around the world [10]. As Scholl [11] argues, future trends in DG that include "*smart approaches, many of which are Data Science-based, rely on the use of Artificial Intelligence (AI) and Machine Learning (ML) in combination with big structured and unstructured data to identify patterns and predictive models, which inform and evaluate decisions of human actors or non-human actors in real-time*". The latest generation in the digital government domain, namely, e-Government 3.0, is described exactly like that: "*e-Government 3.0 refers to the use of new disruptive ICTs (such as big data, IoT, analytics, machine learning, AI), in combination with established ICTs (such as distributed technologies for data storage and service delivery), and taking advantage of the wisdom of the crowd (crowd/citizen-sourcing and value co-creation), for supporting data-driven and evidence-based decision and policy making*" [3]. Vast amounts of data collected and aggregated in government agencies represent a massive potential for employing machine learning and other artificial intelligence techniques, thus unlocking the potential of that data by constructing descriptive and predictive models invaluable in supporting and enhancing government decisions and policymaking.

Considering AI, it is a broader concept that could be described by smaller and specific concepts: big data, machine learning, and decision-making. Castro and New [12] argue that "*AI is a field of computer science devoted to creating computing machines and systems that perform operations analogous to human learning and decision-making*". So, it needs the final concept of "automated decision making" in order for an application to be described as an AI one (i.e. face detection, voice recognition, and autonomous vehicles). The rest of the applications could be characterized as ML ones. As Abbod et al. [13] mentioned, "*Learning can be used to train a machine, so that it optimizes its rule base in a model and then new parameters may be tested in that model*", so the machines can learn with no use of explicit programming. Machine learning is a set of techniques that provides knowledge to any user or machine based on probabilistic algorithms applied to specific data. The most common techniques are classification and regression trees; Neural Network (Multilayer Perceptor); Bayesian Neural Network; Support Vector Regression (SVR); K-nearest neighbor model (KNN) and Gaussian Processes.

In recent years, governments have increasingly outlined ML as a research priority for a better understanding of government's data and implementing more efficient government solutions [14]. When it comes to a government, ML algorithms can help in the identification of significant factors and not yet defined interrelations. As such, they can be used to decrease the complexity of social phenomena that are related to policy problems.

In the literature, ML is applied to a plethora of sectors and fields regarding also the nature of data. In the legal and policy sector, the research focuses more on the analysis of the text. It deals with

Natural Language Processing and text mining, which includes techniques like arguments, topics and rules extraction, clustering, similarity check, and sentiment analysis. This could be further applied to comments or whole texts in several domains like legal texts [15], consultation platforms [16], and social media [17] enhancing the democratic process through participation and better interpretation of the results or finding contradictions in a specific legal system. Furthermore, they are used to classify news [18] or detect fake news [19, 20].

Other fields include cybersecurity and in terms of finding the related research of a domain as well as in multiple business domains [21]. For example, the topic modeling and the collaborative filtering algorithms (ML algorithms) are often used for the improvement of users' experience and for revenue increasing [22, 23]. ML is also used for information extraction from raw data and it can be used for a variety of purposes (e.g. prediction, understanding) [24]. Predictive modeling is defined as the analysis of large data sets to make inferences or identify meaningful relationships that can be used to predict future events [25, 26]. ML techniques in predictive modeling are used for the analysis of both current and historical facts for predictions making either for future or unknown events. Furthermore, ML is applied in the concept of smart cities dealing with traffic prediction and transportation. Accurate traffic prediction based on machine and deep learning modeling can help to minimize the issue [27, 28] of the tremendous rise in traffic volume causing a series of serious problems in modern society's quality of life, such as traffic congestion, delays, increased CO pollution, higher fuel prices, accidents [29], etc.

The list is continuously growing as more applications are included in the healthcare, environment, food, education, and agricultural domains. However, a series of challenges exist in the utilization of ML in the DG domain. As it is highlighted in [30] there is a list of barriers towards the full exploitation of the ML power with two of them being the most important ones. The first one is the combination of various ML techniques towards the production of proper results. Different ML techniques need to be tested to check their performance [31]. The second one is the availability of data. In many cases, the collection of personal data, the ownership of personal data, are subject to General Data Protection Regulation preventing the realization of the benefits from their processing. Policies like GDPR protect the corresponding entities regarding personal or even sensitive data. The publication of such data entails the risk of leading to privacy and ethical issues [21]. Furthermore, ML also depends upon collecting and processing data from society. This data may be explicitly sensitive (e.g., racial origin, religion, health data, ethnic origin) [32]. There are ways of preventing this phenomenon by applying anonymisation techniques before data publishing. But data anonymization in itself is not a fool proof system, being prone to de-anonymization attacks [33]. Even more, with the exponential growth of open data, de-anonymization techniques could work better maximizing the privacy and ethical risks. Based on the lack of the availability of proper data, quality issues occur that in turn, decrease the quality and quantity of the whole ML system [34]. Thus, in many cases, equilibrium should be achieved between these two major barriers. In addition, there can be difficulties of gaining regulatory approval of accessing data (for instance in healthcare), or even lack of data (geographical data) in order for an ML system to be properly trained for quality results. One of the challenges in producing e-Government services built on FL is in ensuring fairness and reproducibility, which is well emphasized in a paper on an analysis framework suitable for governmental scenarios in FL applications [35].

OGD could partially tackle the data availability issue since in most cases the usage of private data knowledge could increase the ML performance. A new solution is needed in order to safeguard legal and ethical issues regarding access to specific data while in parallel increasing the performance of ML algorithms. Federated ML and the proposed framework is moving towards this direction and by proposing a proper solution handling these barriers. This study describes and applies the framework at hand in two separate cases. The first use case revolves around a horizontally partitioned environment, with a goal of agricultural commodity price prediction by combining data from the EUROSTAT price index [36] and FAO product import/export dataset [37]. This data is partitioned on a country level, with each one being a distinct data unit. Using FLOM in this example allows individual producers to gain better information about the cost-effectiveness of producing each commodity. This new knowledge can be discovered without the need for producers to exchange their production cost data, often confidential. The second use case relies on the constructed dataset from the anonymized private data created for a loan approval task containing credit record data and some client-specific private data. By vertically separating the dataset into credit balance data and private data, we compare the gains achieved using FL with the knowledge extracted from the complete dataset versus using only the credit balance data.

OPEN MACHINE LEARNING MODEL INITIATIVES

Machine learning (ML) training data sets are stored in well-known data formats that include unstructured text formats, tabular text-based file formats, columnar data file formats, nested text file formats, binary text file formats, array-based formats, hierarchical data formats, language-specific formats, and various image, video, and document file formats [38].

When it comes to defining data models themselves, different ML frameworks use different formatting: TensorFlow uses protocol buffers [39], Keras models are stored as .h5 files [40] and both PyTorch and Scikit-Learn store models as pickled file formats [41].

By using language, framework and environment agnostic formats for defining ML models, they can be made more easily interoperable, facilitating adherence to open data attributes [42], thus making models open themselves. Formats for open models include common formats successfully implemented and used in previous years. Data Mining Group (DMG) pioneered the search for a common format for defining an open standard for defining ML model exchange types with their design of Predictive Model Markup Language (PMLL) [43] and newer Portable Format for Analytics (PFA) [44].

More recently an extensive work by different industry partners has been done in defining formats for language-agnostic neural network models exchange that include two distinct projects: Neural Network Exchange Format (NNEF) by Khronos Group [45] and Open Neural Network Exchange Format (ONNX) [46] originally authored by Facebook and Microsoft, now a Linux Foundation project.

PMML

PMML is an XML-based open standard for model interchange first developed by DMG in 1997, with the newest release, as of writing this paper, being version 4.4 released in November 2019.

PMML files are described within well-defined parts that include [47]:

- header: general information about the PMML document, including its description, copyright, and timestamp,
- data dictionary: definitions for all the possible fields used by the model, including a description of valid, invalid, and missing data,
- transformation dictionaries: definitions of user data mapping that include: normalization, discretization, value mapping, aggregation, and functions mapping,
- model(s): contains the definition of the models themselves that includes mining schema (per data dictionary), local transformations, targets, outputs as well as model-specific contents.

PMML currently supports 16 different model types combined into more complex ensembles. Furthermore, PMML models are fully interchangeable between different PMML-compliant systems, of which some of the most notable are the pmml package for R language [48] and jpmml [49] for SParkML.

PFA

PFA is a JSON-based open standard for model interchange also built by DMG, with the most current release dating to November 2015.

It is based on AVRO schemas for defining data types and encoding custom functions (actions) applied to inputs. The actions are built using a set of inbuilt functions and language constructs (such as control flow), essentially making PFA a mini functional math language with schema specification. On the other hand, PMML allows building model functionality using only a set of predefined models.

Open Data Group projects spearhead PFA implementation for full implementation for Java Virtual Machine (Hadrian), Python (Titus), and R (Aurelius). Unfortunately, both PFA and PMML currently lack support for standardized operators for describing deep learning models.

More recently, PFA models have been used in the Medical Informatics platform of the Human Brain Project [50] to achieve models built using medical data that are shared in an FL manner.

ONNX

ONNX was initially released as Toffee by Facebook (an interchange format between PyTorch and Caffe), with development later joined by Microsoft and now completely maintained as an open-source project. It uses protobuf as a data structure format and is built using the principle of computational acyclic graphs with built-in operators and standard data types. Each computational node has one or more inputs and outputs and a call to an operator. Definitions of the different operators are implemented externally to ensure that every framework supporting ONNX provides implementations of built-in operators.

Although relatively new, with its first release as ONNX in September of 2017, the project is actively developed, with the latest release being 1.9.0 dated to April 2021. The active development of ONNX is incremental to its success and adoption as it stays current with changes in the deep learning ecosystem of frameworks that support its format, of which some of the most notable include: TensorFlow, Keras, PyTorch, Caffe, and ScikitLearn.

NNEF

Kronos Group developed NNEF, initially released in December 2017, with the latest release date to July 2019. Although a similar project to ONNX, its main focus is on inference interchange, especially with a focus on edge devices. NNEF standard is by definition less frequently evolving with its governance done by a multi-company group.

Technically the main differences between the two standards include using structure definition in a text-based procedural format, the capability of defining compound operators, and avoiding references to machine representations by describing quantization on a more conceptual level, thus allowing for machine-specific inference optimizations favorable for usage in edge devices.

FEDERATED LEARNING

Federated learning is privacy by design and collaborative machine learning technique. In its essence, it allows machine learning to comply with the recently emerging data privacy regulations, incidentally creating a new possibility of using machine learning collaboratively without the need for a central data silo during the training process.

To achieve a collaborative learning process, a data-parallel distributed learning model uses one of the iterative model aggregation mechanisms as a center of the iterative learning process federation down to the data producers themselves.

In FL, every data owner *N* is the training process participant, updating global model weights by training purely on his local data D_l. Central aggregation server utilizes one of the aggregating algorithms to these new unique data owner model weights $w(M^l)$ on a central server, resulting in the new weights $w(M^F)$. This process is displayed in Figure 1. On the other hand, in standard central ML, the central server first aggregates all of the data owner's data shards before starting the training process to generate new model weights. Furthermore, since the central server only needs model weights for global model calculation, the need for data owners to exchange the original training data, often private, is eliminated.

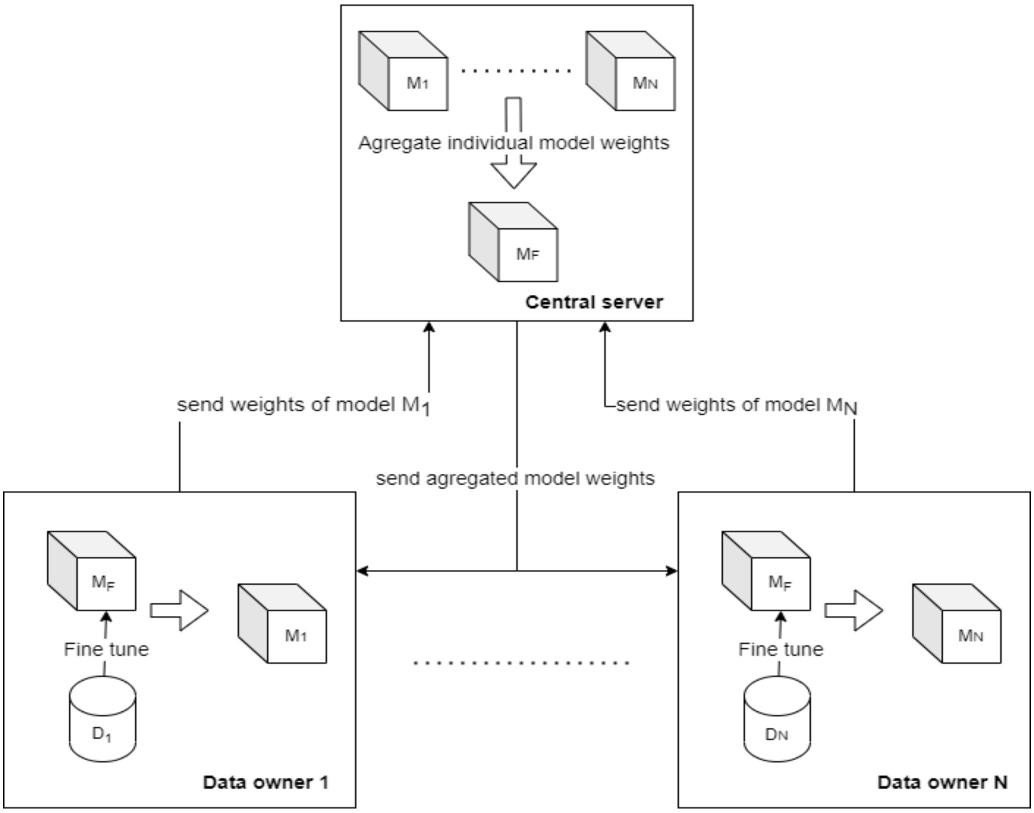


Figure 1. Federated learning process.

In general data-parallel machine learning, there are two ways the data shards (subsets of records in a dataset; physically stored in different locations but logically forming a complete dataset) can partition: horizontally and vertically. The main difference is sharing the same feature sample set in horizontally partitioned datasets (shown in Fig. 2) and contrastingly sharing the same sample set in vertically partitioned ones (shown in Fig. 3). E.g., if different hospitals had the same kind of data of different individual patients - the data is partitioned horizontally. However, if these hospitals had different data on the same patients, their datasets would be partitioned vertically.

Although the original FL model presented by McMahan et al. [51] is designed for horizontally partitioned datasets, several vertical FL models were designed in research that followed [52-54].

However, it is essential to note that the exchange of model weights and their storage on different data owner devices does pose a new possible vector of attacks, commonly known as model inversion attacks. If there is no control over the FL training process, there are vast possibilities of individual data owners tainting the global model by providing model weights trained on local datasets of low quality.

There are also a lot of technical challenges in achieving needed communication requirements for the training process and in eliminating potential problems that could emerge from a significant heterogeneity in data owner's device availability and data quality.

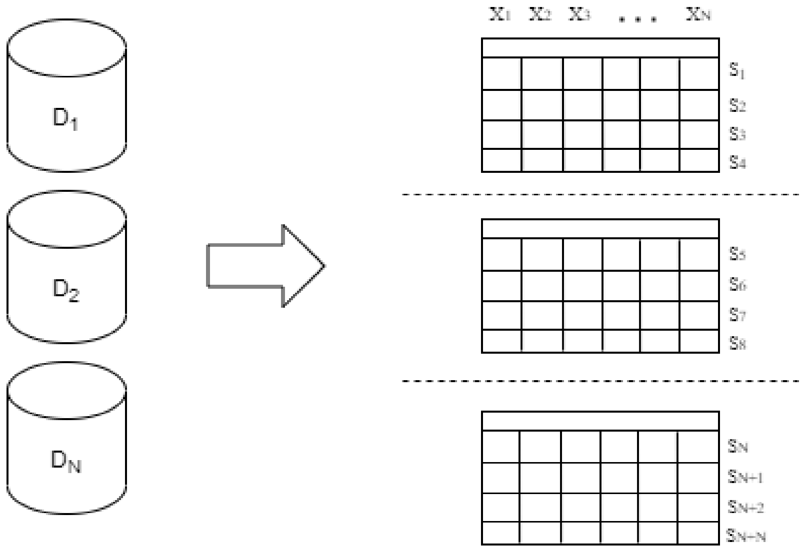


Figure 2. Horizontal data partitioning with each data shard sharing the same feature space.

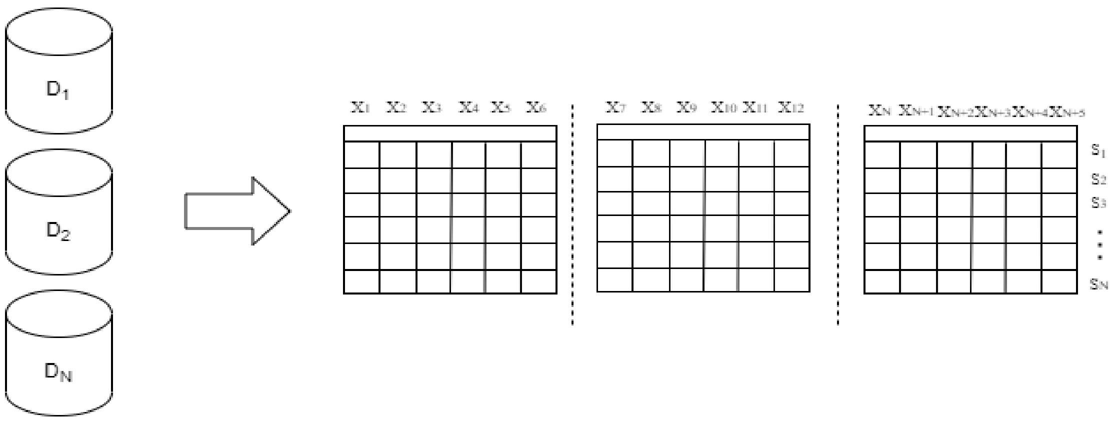


Figure 3. Vertical data partitioning with each data shard sharing the same sample space

Although there are many challenges in creating a real-world working use case, FL can be used as a tool for building cross-enterprise and cross-domain ecosystems for big data and artificial intelligence, where centralized machine learning and cloud-centric paradigms failed to overcome barriers for its inception. Authors in [55] emphasized the importance of coupling the practical usages with the evolution of business models that would accompany it by proposing the usage of FL in data alliances of enterprises.

FLOM FRAMEWORK

This artice introduces a new concept based on the symbiosis of the general federated learning process with that of open model specifications called Federated Learning Open Model

(FLOM). FLOM provides a layout for using FL in the creation of open models and federated training processes with the primary goal of overcoming the technical barriers to using FL. In essence, it allows an easier generation of business models built on the federation of the learning process and using global knowledge without sharing any private or confidential data.

FLOM is a framework for developing an open ML training process done in a federated manner, with model sharing being done by exchanging model definitions in an open standard.

FLOM is accompanied by a technical specification that consists of descriptions of:

- client (individual data owner) data and device requirements,
- a central (aggregation) server specifications and requirements,
- an inferable and trainable model shared with an open standard specification (e.g. PFA, PMML, ONNX),
- an Application Programming Interface (API) with the implemented endpoints for all of the necessary steps for achieving FL process.

The General FL process takes four specific steps that get iteratively repeated during the lifecycle:

1. clients send their model updates,
2. aggregation of model updates into new global model weights (learnable and non-learnable parameters of ML models),
3. disseminating the new global model weights to the client,
4. clients update their local models and start the new iteration.

From the client's side, the FLOM process has a few additional steps to acquire client and server specifications and register the client to the central server (steps 1-4 in Fig. 4).

In essence, our contribution by defining FLOM is in adding these extra steps available through an API endpoint enveloping a traditional FL process with model definitions in one of the open specification formats. By doing this, we hope to help facilitate the usage of ML models in an open and approachable manner that makes it easier to set up and use. On the more practical level, it allows for an easy integration of different prosumers to a collaborative ML process that FL made possible, and FLOM made more accessible.

CLIENT DATA AND DEVICE REQUIREMENTS

Client data and device requirements include a definition of the necessary minimum data quality metrics and optionally capabilities of the client needed to partake in the training process. These can include required not-null data attributes, data generation frequency, data quantity, and any possible additional metrics [56].

Critical endpoints accompany these requirements on the API side for receiving the human and machine-readable specifications and intrinsic procedures to check the quality of the newly generated model weights and the time that took the client to send the newly generated weights [57]. In addition, rules should be applied to drop out and late clients to ensure the model quality.

CENTRAL (AGGREGATION) SERVER SPECIFICATION AND REQUIREMENTS

Central server specifications and requirements include the description of the maximal number of clients and the estimated time it takes to conclude a single training iteration. Estimated training time is analogous to hardware and network capabilities of the server that is used for achieving aggregation and dissemination of model weights as well as serving client requests on the API endpoints.

ML MODEL

The description of the model that is used as the base for the service achieved by the FLOM process is distributed in one of the open formats that include PFA, PMML, and ONNX. The open format allows for training and inference across different software and hardware environments used to achieve the training and inference on the client side.

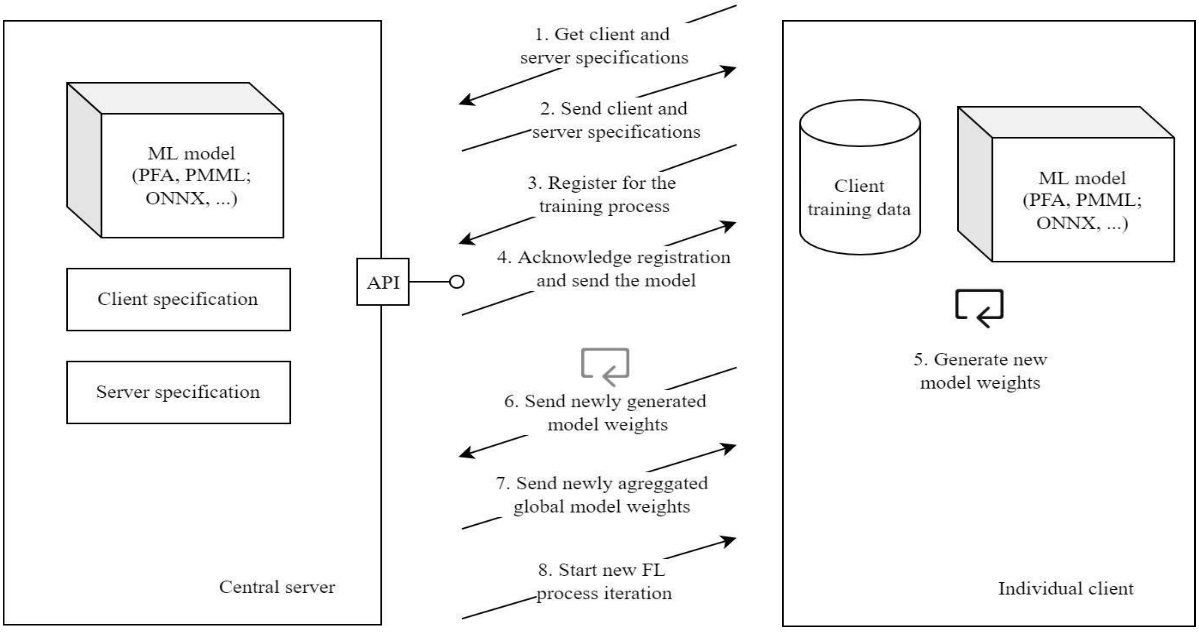


Figure 4. Steps of the FLOM process (steps 5-8 are iteratively repeated during the whole training lifecycle).

API INTERFACE

API interface is located on the central server and needs to support endpoints for registering clients, receiving client model updates, and sending the new global weights to all participating parties. From the client-side, these include actions to:

- receive client requirements description,
- receive central server specification,
- register for participation (generates a unique ID for internal client references),
- receive ML model in an open format,
- send client-specific model weights,
- receive a message with new model weights and a synchronization message to start of the new training iteration.

API interface is achieved using one of the many open-source web frameworks based on well-known standards for client-server communications.

APPLICATION

HORIZONTALLY PARTITIONED ENVIRONMENT

The first use case is created with a dataset containing horizontal data partitions. It aims to attain price prediction of agricultural commodities by incorporating data from the EUROSTAT price index [36] and FAO product import/export dataset [37], which are, in essence, horizontally sharded datasets with partitioning done on the per-country level. An excerpt from these datasets is displayed in Table 1.

The environment is partitioned horizontally on a market region level, where participants could be certain countries, regions, private companies and other organizations.

Although these datasets are open, the use of FLOM in this application allows new business models wherein individual organisations are incentivized to join the training process. The extra incentive is gained from better price forecasting by joining their privately built and historical knowledge on their market area with the latent knowledge located in the more globally distributed knowledge extraction.

Table 1. Data for countries Croatia and Greece found in the FAO and EUROSTAT datasets.

Country code	FAO Commodity ID	Description	Import (t)	Export (t)	Year
HR	882	Milk, whole fresh cow	189435	25849	2020
GR	882	Milk, whole fresh cow	91162	25849	2020

Country code	EUROSTAT Agricultural price ID	Description	Price index (% of 2015 price)
HR	121100	Cow's milk	102.9
EL (GR)	121100	Cow's milk	102.21

Country code	Description	EUROSTAT product price, € per100 kg
HR	Raw milk	34
EL	Raw milk	39

In this use case, the frequency of data generation is once per year, so aggregation server and client hardware specifications are not that stringent.

FLOM consists of:

- linear regression model distributed in the PFA format,
- server specifications that need computational capabilities to run the model aggregation on a yearly basis, with an estimation done by benchmarking using historical country data,
- API endpoints that are defined in the previous section with their locations referenced in OpenAPI format,
- client specifications that define the needed data frequency with yearly samples including organizational area extent in geoJSON format, historic price data in USD, and production and trade data in millions of tons.

By joining the FLOM trading process, the individual organizations would help build the global model by including more finely grained samples than the ones found in open datasets, that are generally per country level. This would further enhance the benefits that organizations would get from forecasting potential further prices, allowing them to compare the potential profits for the upcoming years, regarding changes to their own area of interest, production and trade data.

Using FLOM in this example allows individual producers to understand better the cost-effectiveness of producing each commodity. This new knowledge can be discovered without the need for producers to exchange their production cost data, often confidential.

VERTICALLY PARTITIONED ENVIRONMENT

The second example is built on a vertically partitioned data set that is artificially constructed from a loan ratification ML models analysis [58]. The loan approval prediction system has the goal of automatically calculating the weight of each attribute of the clients taking part in loan processing and ultimately making the decision whether a new applicant should be approved for the loan or not. Originally, these could be achieved using different ML models, including logistic regression, random forests classifiers, support vector machines, etc. Originally this dataset was a horizontally partitioned dataset with an individual sample being each client (person). The vertical partitioning is done on the client's attributes, and they are separated into two groups: private data and financial data. Private data being: gender (male or female), marital status, number of dependents, education qualification, whether the person is self-employed, and the financial data being: the person's income, co-applicant income, loan amount, loan amount term, credit history and property area (urban/suburban). An excerpt from this dataset can be seen in Table 2.

Table 2. An excerpt from the loan prediction task dataset, with private and financial partitions.

Loan ID	Gender	Married	Dependents	Education	Self employed
LP001032	Male	No	0	Graduate	No
LP001034	Male	No	1	Not Graduate	No
LP001036	Female	No	0	Graduate	No

Loan ID	Income	Co-applicant income	Loan amount	Loan amount term	Credit history	Property area	Loan status
LP001032	4950	0	125	360	1	Urban	Y
LP001034	3596	0	100	240		Urban	Y
LP001036	3510	0	76	360	0	Urban	N

Since people have become accustomed to safeguarding their personal data and becoming more and more unwilling to share it, this could hinder potential loan providers in using services of loan approval prediction systems. However, one could build a service-oriented around ML loan approval where training on private data is done on the client's side using FLOM with more financial data training done on the loan provider's side. This process could be implemented in mobile banking applications wherein, user's private data would stay on their own mobile device.

FLOM consists of:

- tree regression model distributed in the PFA format,
- server specifications that needed computational capabilities to run the model aggregation on a monthly basis, with an estimation done by benchmarking using historical data,
- API endpoints that are defined in the previous section with their locations referenced in OpenAPI format,
- client specifications define the needed computational capabilities to run the monthly training.

By joining the FLOM training process potential applicants could have the benefits of using automated loan approval prediction systems themselves, and better plan their financial future without needing to share their private data. Their presence in the training process would benefit the global model to build the weights that are applied to private data, unavailable to the loan provider.

Both use cases are displayed in Figure 5, which focuses on the process defined in Figure 4, with respect to the client and data partition specific to each use case.

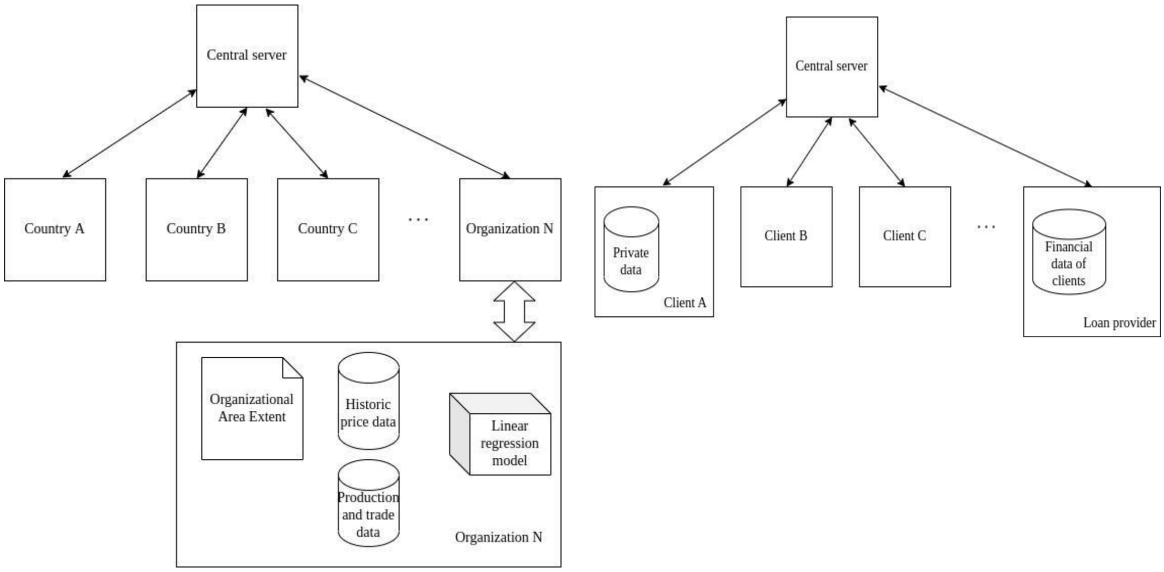


Figure 5. Diagrams of FLOM in the first (agricultural commodity price prediction) and second (loan approval prediction) use case.

CONCLUSION

In this article, we presented a framework for using open standard model formats in a federated machine learning manner. The FLOM framework represents a blueprint for defining open models and the requirements that support the federated learning processfor both the clients and the central server that are accompanied by a model defined in one of the currently available open standards.

The framework encourages the design of new tools, services, and applications for many previously not practically feasible domains. We see this framework as a tool for facilitating collaborative model training and sharing, allowing the combination of knowledge creation from both open datasets and datasets closed due to regulatory or confidentiality reasons. Its potential capabilities as an eGovernance tool are showcased using two potential use cases that leverage openly available and closed datasets attained through the collaborative FL. The use cases showcase the multitude of possible application domains and collaborations, with the first being private business-oriented and the second being private person-oriented.

Future work should be done in evaluating the use cases regarding central ML models that lack the knowledge extruded from private and confidential data. Further disseminating runnable FLOM examples that could easily be reused would encourage broader research on using FL in more general use cases. Using FLOM in the eGovernment context could enable many innovative services that could further citizen participation and incentivize private organizations to build and publicly provide intelligent services in collaboration with governmental and various public organizations.

ACKNOWLEDGMENTS

This research is part of TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592.

REFERENCES

[1] Twizeyimana, D. and Andersson, A.: *The public value of e-government – a literature review.* Government Information Quarterly **36**(2), 167-178, 2019, http://dx.doi.org/10.1016/j.giq.2019.01.001,

[2] Lachana, Z.; Alexopoulos, C.; Loukis, E. and Charalabidis, Y.: *Identifying the different generations of e-government: an analysis framework*.
The 12th Mediterranean Conference on Information Systems (MCIS), pp.1-13, 2018,

[3] Charalabidis, Y.; Loukis, E.; Alexopoulos, C. and Lachana, Z.: *The three generations of electronic government: From service provision to open data and to policy analytics*.
International Conference on Electronic Government. Springer, pp.3-17, 2019,

[4] Zuiderwijk, A. and Janssen, M.: *Open Data Policies, Their Implementation and Impact: A Framework for Comparison*.
Government Information Quarterly **31**(1), 17-29, 2013,
http://dx.doi.org/10.1016/j.giq.2013.04.003,

[5] Bertot, J.; Jaeger, P. and Grimes, J.: *Using ICTs to Create a Culture of Transparency: E-Government and Social Media as Openness and Anti-Corruption Tools for Societies*.
Government Information Quarterly **27**(3), 264-271, 2010,
http://dx.doi.org/10.1016/j.giq.2010.03.001,

[6] UNCTAD: *Data protection regulations and international data flows:Implications for trade and development data protection regulations and international data flows: Implications for trade and development*, 2016.

[7] EU General Data Protection Regulation (GDPR*): Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*, OJ 2016 L 119/1.

[8] Jobin, A.; Ienca, M. and Vayena, E.: *The global landscape of AI ethics guidelines*.
Nature Machine Intelligence **1**, 389-399, 2019,
http://dx.doi.org/10.1038/s42256-019-0088-2,

[9] Janowski, T.: *Digital government evolution: From transformation to contextualization*.
Government Information Quarterly **32**(3), 221-236, 2015,
http://dx.doi.org/10.1016/j.giq.2015.07.001,

[10] Lachana, Z.; Alexopoulos, C.; Loukis, E., and Charalabidis, Y.: *Identifying the different generations of Egovernment: an analysis framework*.
The 12th Mediterranean Conference on Information Systems (MCIS), pp.1-13, 2018,

[11] Scholl, H.J.: *Digital government: looking back and ahead on a fascinating domain of research and practice*.
Digital Government: Research and Practice **1**(1), pp.1-12, 2020,
http://dx.doi.org/10.1145/3352682,

[12] Castro, D. and New, J.: *The promise of artificial intelligence*.
Center for Data Innovation, 2016,
https://www2.datainnovation.org/2016-promise-of-ai.pdf,

[13] Abbod, M.F.; Catto, J.W.; Linkens, D.A. and Hamdy, F.C.: *Application of artificial intelligence to the management of urological cancer*.
The Journal of Urology **178**(4), 1150-1156, 2007,
http://dx.doi.org/10.1016/j.juro.2007.05.122,

[14] Leonard M.: *Government leans into machine learning*.
GCN |Technology, Tools and Tactics for Public Sector IT, 2018,
https://gcn.com/articles/2018/08/17/machine-learning.aspx,

[15] Avgerinos Loutsaris, M.; Lachana, Z.; Alexopoulos, C. and Charalabidis, Y.: *Legal Text Processing: Combining two legal ontological approaches through text mining*.
The 22nd Annual International Conference on Digital Government Research, pp.522-532, 2021

[16] Arana-Catania, M., et al.: *Citizen Participation and Machine Learning for a Better Democracy*.
Digital Government: Research and Practice **2**(3), 1-22, 2021,
http://dx.doi.org/10.1145/3452118,

[17] Androutsopoulou, A.; Charalabidis, Y. and Loukis, E.: *Policy Informatics in the Social Media Era: Analyzing Opinions for Policy Making.*
Proceedings EGOV-CeDEM-ePart 2018 Conference, 2018,

[18] Singh, R.; Chun, S.A. and Atluri, V.: *Developing Machine Learning Models to Automate News Classification*.
The 21st Annual International Conference on Digital Government Research, pp.354-355, 2020,

[19] Wani, A., et al.: *Evaluating deep learning approaches for covid19 fake news detection*.
International Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation. pp.153-163, 2021,

[20] Sahoo, S.R. and Gupta, B.B.: *Multiple features based approach for automatic fake news detection on social networks using deep learning*.
Applied Soft Computing **100**, No. 106983, 2021,
http://dx.doi.org/10.1016/j.asoc.2020.106983,

[21] Jordan, M.I. and Mitchell, T.M.: *Machine learning: Trends, perspectives, and prospects*.
Science **349**(6245), 255-260, 2015,
http://dx.doi.org/10.1126/science.aaa8415,

[22] Zhou, Y.; Wilkinson, D.; Schreiber, R. and Pan, R.: *Large-scale parallel collaborative filtering for the netflix prize*.
Algorithmic Aspects in Information and Management. pp.337-348, 2008,

[23] Ghoting, A., et al.: SystemML: *Declarative machine learning on MapReduce*.
IEEE International Conference on Data Engineering (ICDE), pp.231-242, 2011,

[24] Sarker, I.H.: *Machine learning: Algorithms, real-world applications and research directions*.
SN Computer Science **2**(3), 1-21, 2021,

[25] Guszcza, J.: *Analysing Analytics*.
Contingencies, American Academy of Actuaries, 2008,

[26] Biswas, P. and Bishnu, P.: *Application of data mining and CRM in banking sector medical insurance*.
International Journal of Innovative Research in Computer and Communication Engineering **3**(1), 38-46, 2015,
http://dx.doi.org/10.15680/ijircce.2015.0301007,

[27] Essien, A.; Petrounias, I.; Sampaio, P. and Sampaio, S.: *Improving urban traffic speed prediction using data source fusion and deep learning*.
IEEE International Conference on Big Data and Smart Computing (BigComp), pp.1-8, 2019,

[28] Gregurić, M.; Vujić, M.; Alexopoulos, C. and Miletić, M.: *Application of deep reinforcement learning in traffic signal control: An overview and impact of open traffic data*.
Applied Sciences **10**(11), 2020,

[29] Guerrero-Ibáñez, J.; Zeadally, S. and Contreras-Castillo, J.: *Sensor technologies for intelligent transportation systems*.
Sensors **18**(4), No. 1212, 2018,
http://dx.doi.org/10.3390/s18041212,

[30] Alexopoulos, C., et al.: *How machine learning is changing e-government*.
Proceedings of the 12th International Conference on Theory and Practice of Electronic Governance, pp.354-363, 2019,

[31] Singh, R.; Chun, S.A. and Atluri, V.: *Developing Machine Learning Models to Automate News Classification*.
The 21st Annual International Conference on Digital Government Research, pp.354-355, 2020.

[32] Goodman, B. and Flaxman, S.: *European Union regulations on algorithmic decision-making and a "right to explanation"*.
arXiv preprint:1606.08813, 2016,

[33] Ding, X.; Zhang, L.; Wan, Z. and Gu, M.: *A Brief Survey on De-anonymization Attacks in Online Social Networks*
International Conference on Computational Aspects of Social Networks, pp.611-615, 2010,
http://dx.doi.org/10.1109/CASoN.2010.139,

[34] Netten, N.; van den Braak, S.; Choenni, S. and van Someren, M.: *A Big Data Approach to Support Information Distribution inCrisis Response*.
Proceedings of the 9th International Conference on Theory and Practice of Electronic Governance, pp.266-275, 2016,
http://dx.doi.org/10.1145/2910019.2910033,

[35] Balta D., et al.: *Accountable Federated Machine Learning in Government: Engineering and Management Insights*.
In: *ePart 2021: Electronic Participation*. Lecture Notes in Computer Science **12849**, Springer, Cham, pp.125-138, 2021,
http://dx.doi.org/10.1007/978-3-030-82824-0_10,

[36] EUROSTAT: *Price index.*
https://ec.europa.eu/eurostat/cache/metadata/en/apri_pi_esms.htm,

[37] FAO: *Product import/export dataset.*
https://www.fao.org/faostat/en/#rankings/commodities_by_country_imports,

[38] Kleppmann, M.: *Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems*.
O'Reilly Media, Inc., 2017,

[39] –: *A Tool Developer's Guide to TensorFlow Model Files*.
https://chromium.googlesource.com/external/github.com/tensorflow/tensorflow/+/r0.10/tensorflow/g3doc/how_tos/tool_developers/index.md,

[40] Keras API Reference.
https://keras.io/api/models/model_saving_apis,

[41] –: *Python Object Serialization*.
https://docs.python.org/3/library/pickle.html,

[42] Open Knowledge Foundation: *What is Open Data?*
https://opendatahandbook.org/guide/en/what-is-open-data,

[43] Data Mining Group: *Predictive Model Markup Language Version 4.4.1.*
http://dmg.org/pmml/pmml-v4-4-1.html,

[44] Data Mining Group: *Portable Format for Analytics Version 0.8.1.*
http://dmg.org/pfa,

[45] –:*Neural Network Exchange Format: Version 1.0.4, Revision 1.*
The Khronos NNEF Working Group, 2021,

[46] –: *Open Neural Network Exchange*.
https://onnx.ai,

[47] Guazzelli, A.; Zeller, M.; Lin, W. and Williams, G.: *PMML: An Open Standard for Sharing Models*.
The R Journal **1**(1), pp. 60-65, 2009,
http://dx.doi.org/10.1002/j.1941-9635.2009.tb00398.x,

[48] –:*CRAN Predictive Model Markup Language*.
https://cran.r-project.org/web/packages/pmml/index.html,

[49] –: *Java PMML API.*
https://github.com/jpmml,

[50] Levitan, S. and Claude, L.: *Open standards for deployment, storage and sharing of predictive models. PMML / PFA / ONNX in action*, 2019.
http://dx.doi.org/10.13140/RG.2.2.31518.89920,

[51] McMahan, H.B., et al.: *Federated learning of deep networks using model averaging*.
ArXiv preprint:1602.05629, 2016,

[52] Wu, Y. et al.: *Privacy preserving vertical federated learning for tree-based models*.
Proceedings of the VLDB Endowment **13**(12), 2090-2103, 2020,
http://dx.doi.org/10.14778/3407790.3407811,

[53] Sun, J., et al. *Vertical Federated Learning without Revealing Intersection Membership*.
ArXiv preprint: abs/2106.05508, 2021,

[54] Romanini, D. et al.: *PyVertical: A Vertical Federated Learning Framework for Multi-headed SplitNN*.
ICLR 2021 Workshop on Distributed and Private Machine Learning, 2021,

[55] Yang, Q.; Liu. Y.; Chen, T. and Tong, Y.: *Federated machine learning: Concept and applications*.
ACM Transactions on Intelligent Systems and Technology **10**(2), 1-19, 2019,
http://dx.doi.org/10.1145/3298981,

[56] Sidi, F. et al.: *Data quality: A survey of data quality dimensions*.
International Conference on Information Retrieval & Knowledge Management, pp.300-304, 2012,
http://dx.doi.org/10.1109/InfRKM.2012.6204995,

[57] Shyn, S.K.; Kim, D. and Kim, K.: *FedCCEA : A Practical Approach of Client Contribution Evaluation for Federated Learning*.
ArXiv preprint:abs/2106.02310, 2021,

[58] Kumar, A.; Ishan, G. and Sanmeet, K.: *Loan approval prediction based on machine learning approach*.
IOSR Journal of Computer Engineering **18**(3), 79-81, 2016,
http://dx.doi.org/10.9790/0661-1803017981.

Interdisciplinary Description of Complex Systems 20(2), 179-189, 2022

SERIOUS GAMES FOR BUILDING DATA CAPACITY*

Davide Di Staso[1,] **, Ingrid Mulder[1],
Marijn Janssen[1] and Fernando Kleiman[2]

[1]TU Delft
Delft, The Netherlands

[2]NHL Stenden
Leeuwarden, The Netherlands

DOI: 10.7906/indecs.20.2.9
Regular article

Received: 24 January 2022.
Accepted: 22 April 2022.

ABSTRACT

Open data can support the creation of new services, facilitate research, and provide insights into everyday issues affecting citizens. Although public administrations are making efforts to create sustainable and inclusive open data systems, there is limited capacity to identify suitable datasets, clean, release, and reuse them. Serious games offer a possible solution for data capacity building and have already been used to train civil servants and citizens on the topic of open data. This research presents a review of serious games and discusses their potential for data capacity building. The games selected in the review are classified and described according to their different learning outcomes, formats, and type of media. Most serious games found in this review can be categorized as teaching games and are designed to raise data awareness, which is only a limited aspect of building data capacity. We found a lack of design games, research games, and policy games. Given their success for ideation in other fields, design games offer a particular opportunity to build data capacity by generating new ideas about how to reuse open datasets.

KEY WORDS

data capacity, serious games, open data

CLASSIFICATION

ACM: K.3.1, K.8.0
JEL: C18

*This is the extended version of the abstract published in: Vujić, M. and Šalamon, D., eds.: Book of abstracts of the National Open Data Conference. University of Zagreb, Faculty of Traffic and Transport Sciences, Zagreb, 2021.

**Corresponding author, η: D.DiStaso@tudelft.nl; -;
Faculty of Industrial Design Engineering, Landbergstraat 15, 2628 CE Delft, The Netherlands

INTRODUCTION

Open data is any data that is freely accessible and reusable by anyone for any purpose [1]. Open data can be reused to create or improve services, and to identify local issues and community needs more easily [2]. While public sector organizations play a significant role in releasing datasets to the public, the private sector may also open datasets to the public [3]. In this research, we will refer to the general concept of open data to include datasets released by both the public and the private sector.

The opening and reuse of datasets involves different actors and services, such as data providers, publishing organizations, infomediaries, tools for data storage and analysis, and researchers looking for data [3]. Opening data can effectively create a network of complex interdependencies and networks of interaction, an "ecosystem" [3].

Within the open data ecosystem, non-expert users (such as citizens and public administrators) have an important role in that they are aware of the issues and needs of their communities, which can be addressed using open data [4]. On the other hand, expert users, such as civic hackers and developers, own the skills required to implement practical solutions using open data [4]. Mulder, Jaskiewicz, and Morelli [5] explored recent paradigm shifts that have the potential to seed change within societal systems and look specifically at how open data can become a new type of "commons'" that can support digital citizenship. In the current work, we explore the use of serious games for building data capacity in problem-driven societies. Alongside the delivery of open data-driven solutions, open data can only become a new commons if a larger community and culture of working with data is created around it.

Serious games offer an important tool to bring together both expert and non-expert users and transfer the required knowledge and skills needed to work with open data. Serious games differentiate themselves from entertainment games in that their main purpose is not to amuse, but to educate [6] and they have been in use for over a decade to facilitate learning and ideation [7]. Some serious games adapt game mechanics from commercial video games to achieve educational objectives. For example, "Socrates Jones: Pro Philosopher" [8] takes inspiration from "Ace Attorney", a popular legal drama game which uses visual novel mechanics. The developers of Socrates Jones used Ace Attorney's mechanics but created dialogues and game content to teach philosophical thinking. In the public sector, serious games have been used in different scenarios, such as to ideate service delivery principles [9] and to train railway traffic controllers [10], among others.

In the remainder, we review serious games for open data and elaborate upon their potential contribution for building data capacity. We define building data capacity as the process that empowers citizens and civil servants to understand and reuse open data, thereby creating the needed practical and analytical skills.

This research will answer the following research questions:

1. Which games – or types of games – have the potential to build data capacity?
2. What kind of data capacity can these serious games build?

The review starts by looking at the list of games on the topic of open data compiled by Kleiman [11]. Entries are filtered according to four criteria, selecting interventions that: (1) are sufficiently documented, (2) fit the definition of a "game", (3) must also fit the definition of "serious game", and (4) have an educational purpose that is related to building data capacity. We analyze selected games using the classification by Grogan and Meijer [12], assigning them a type based on the kind of knowledge transferred or created by the game and its beneficiary.

CONCEPTUAL FRAMEWORK

To analyze the serious games selected in the review, we use the classification by Grogan and Meijer [12]. Starting from the type of knowledge that the game deals with and its beneficiary (see table 1), Grogan and Meijer [12] identify four broad categories of games. Policy games are based on real world scenarios so that the participant can experiment with different solutions and gather knowledge about the scenario represented in the game. Teaching games are based on a fictional setting, with the knowledge transferred by the game being generalizable and not based on a specific scenario. Design games "provide a participatory environment" [12, p.545] and can be used to ideate new artifacts and create new knowledge. Finally, research games are used to observe participants in an experimental setting and test hypotheses.

Table 1. Classification of games according to knowledge type and beneficiary [12].

	Knowledge beneficiary	
Knowledge type	**Participant**	**Principal**
Generalizable	**Teaching** Experiential learning Dangerous tasks	**Research** Hypothesis generation and testing Artifact assessment
Contextual	**Policy** Organizational learning Policy intervention	**Design** Interactive visualization Collaborative design

The paper is structured as follows: first, we describe the methodology used to compile a list of games for building data capacity. We then present our results, giving a brief description of each game and summary of their main characteristics and learning outcomes. We then discuss how serious games contribute to data capacity building and which specific aspects of this process they aim to tackle, followed by a summary of our conclusions.

METHODOLOGY

The list of gamified interventions related to data compiled by Kleiman [11] was used as a starting point to map games for data capacity. The list was screened using the following filters:

1) The intervention should have sufficient documentation to allow for the intervention and its educational content (if present) to be analyzed and categorized. This can include game manuals, scientific publications, or an actual playable copy of the game available online.
2) The intervention must be a game, meaning it must be an "attempt to achieve a specific state of affairs (prelusory goal)" while being limited by certain rules, which are accepted by the player(s) because they enable the game play [13, p. 41] as cited by [14].
3) The intervention must fit the definition of "serious game" by Abt [6] as cited in Djaouti et al. [15], meaning it should have an "explicit and carefully thought-out educational purpose" and the primary reason to play should not be entertainment.
4) The intervention's educational purpose must be related to the goal of "building data capacity", meaning it must be aimed at providing skills such as general knowledge about open data, data reuse, or operational and technical knowledge about how to use and visualize datasets [16].

The literature review on data-related gamified interventions by Kleiman [11] included a total of 23 entries. From these, two interventions were excluded as they didn't meet the definition of a "game" (filter 2). One intervention was excluded as it was not sufficiently documented.

Two interventions were excluded as they are not serious games, but rather entertainment games (filter 3). Ten interventions were excluded because, while they use open datasets to generate playable content, the educational purpose of the intervention is not directly related to building data capacity (filter 4). For example, Bar Chart Ball [17], generates bar charts from various datasets, such as the percentage of people who feel they can influence decisions in different cities in the UK. A ball is dropped on top of the bar chart and starts sliding around under the force of gravity. The aim of the game is "to control this ball, and make it go where they want" [17, p.1]. While this is an example of a data-related game and an interesting reuse of open datasets, its main educational outcome seems to be the memorization of the shapes of different bar charts, which is not directly related to building data capacity. For similar reasons, we filtered out the other games described by Gustafsson Friberger et al. [18] which reuse datasets to procedurally generate content but are not related to building data capacity.

To describe and categorize the serious games for data capacity building, we used similar variables to the ones suggested by Katsaliaki and Mustafee [19]. Variables to be captured were selected based on their relevance and scope of this research and to give a sufficient overview of the game's general characteristics. In a similar fashion to Katsaliaki and Mustafee [19], the data was collected by researching available materials about the game (cards, manuals, etc.), related publications, playing the games, or reading their descriptions on the respective websites. For each game, the general gameplay and rules are described, along with details about the game's platform, genre, learning objective, and learning purpose. In addition to this classification, hereafter, we describe each game, and its expected contribution to data capacity building.

CASE DESCRIPTIONS

Further in the text we introduce the twelve games selected, along with a short description of the rules and gameplay. The main characteristics of each game are summarized in Table 3.

Agenda 2030

Agenda 2030 is a discussion game for 6 to 31 players. A set of 50 cards representing 5 departments represent reports, maps and documents which are needed to monitor the Sustainable Development Goals within a local governmental context (Municipality of Teresina, in Brazil). One participant plays as the database for the teams, and the others are distributed through the 5 different departments of the local government. Each team has a negotiator which trades data with other teams. By trading cards, players need to find the specific datasets to complete their SDGs indicators. Completing indicators give teams another type of card, with random events, making the game more fun. The game ends when the full indicator checklist is completed.

Data Belt

Data Belt is a four-player online video game which shares some aspects with Winning Data (described later in this list), such as the four different player roles, the basic dynamic of answering citizen's demands for public services, generating datasets, and deciding whether or not to open. The game was tested in a pre-experimental setting and "participants were more inclined to believe that some public sector data can be shared" [20, p.162]. The game can be useful when played together my civil servants with different levels of experience in open data decision-making, as it can facilitate knowledge sharing among the players.

Data Dealer

Data Dealer is a single player online game about privacy issues related to data brokers and the resale of personal information [21]. The user fills the shoes of a corrupt data broker, trying to make as much profit as possible from shady deals with tycoons and corporations. The player owns a database connected to certain data sources (like dating sites and online personality tests). Money can be invested to upgrade these data sources, therefore capturing more data which can then be resold to corporations with dubious aims. Data Dealer is a management game, where the player needs to carefully balance resources to maximize profit. This game could be an important tool to understand the role of data brokers and how they manage to harvest (legally and illegally) data from different sources.

Digital Identity game (Data gedreven werken game)

The Digital Identity game is a board game where players need to reach the center of the board with remaining resources. Specific spots with discussion logo reduce the number of available resources from players - representing the loss of pieces of her digital identity. In some cases, disagreements between players need to be voted upon. The search engine DuckDuckGo is used to solve doubts about operating services. As defined by Zuboff's Surveillance Capitalism, when the players lose all their resources (a metaphor to giving away all her personal data), they are only the carcasses that remain when the data is plundered [22].

Datak

Datak [23] is a single player online game based on a journalistic investigation into the problematic aspects of big data [24]. In Datak, the player interprets the role of a new hire as the assistant to the mayor of DataVille. Part of the job is to make decisions that can affect the players and citizens, for example by deciding what kind of precautions to take when archiving voters' information or when a security breach occurs. Datak was developed after a journalistic investigation; its aim is to raise awareness about the implications of data collection and privacy violations. Datak could be useful in introducing a non-expert audience to the most common ways in which data privacy rules are violated and the basic terminology to describe these violations.

Datascape

Datascape is a board game in which the players are given research questions that can be answered using data [25]. The players are also given a stylized map, on which they need to point where to source the data from. Each section of the map possesses certain data types such as light, weather, wind, water level, etc. Datascape can play a role in introducing a non-expert audience into data collection and the different sources of datasets.

Dataspel

Dataspel is a board game in which a team leader is responsible to coordinate the team in making discoveries based on data. Each member of a team has a certain role, either being a content expert or a data expert. Each game round consists of three phases, from distributing the work to analyzing the available datasets. Specific problems and politically sensitive topics can influence the analysis and publications. Scores are defined based on the number of points each team leader archives by the end of the game for analyzing and publishing datasets.

Datopolis

Datopolis is a board game which can be played by two to five players [26]. Players are presented with datasets of three different types: open, closed, and private. Open datasets are public, and any player may use them to create new tools (services), whereas a closed dataset may or may not be opened by the player owning it. Private datasets can never be opened to other players. The game is designed so that players need to negotiate which datasets to open and combine in order to build services. There is a standard version and a short version which can be used during workshops. Datopolis could be useful in showing how, to create a useful service or application, developers need several entities to open datasets, which is sometimes challenging.

Jogo de Governo Aberto

The Open Government Game is a card game involving 4 to 6 players, each of them receiving a specific set of cards to be used in the gameplay. Each set contains actions related to specific actions on Transparency, Participation, and Accountability. These are considered as the main pillars to an open government, which the players must collaborate to achieve. The game has been adapted for remote play in tabletopia [27] though it is still only available in Portuguese.

Open Data Card Game

The Open Data Card game is an in-person game for multiple groups of three people [28], designed for ideation during workshops and hackdays. The game is aimed at getting participants excited about the possible uses and combinations of open datasets and generating new ideas. This game could be an effective way of facilitating brainstorming during hackathons, when participants need to think of ways to reuse datasets.

Run that town

Run that town is a single-player mobile game which uses real data from Australia's 2011 census [29]. The player can enter their postcode to customize the experience with data from their neighborhood. The player fits the shoes of a local politician, taking decisions about what kind of public works to initiate and where to spend money.

Winning Data

Winning Data is a four player in-person role-playing game [30] about open data. In Winning Data, players interpret the roles of civil servant, colleague, citizen, and boss and need to collaborate to answer citizens' demands for public services. Just like in a real-life public office, this activity leads to the creation of the datasets, which the team can either completely open to the public, partially share (removing some personal information), or completely close. In an experimental setting, after playing the game, civil servants had a "better understanding of the positive outcomes of data opening" [31, p. 18], thus showing potential for building data capacity among public sector employees. Similarly to Data Belt, this game can facilitate knowledge sharing about the risks and benefits of opening a given dataset, especially when a mix of more and less experienced decision-makers are playing.

The following two tables provide a summary of the selected case descriptions. Table 1 summarizes the cases (serious games) reviewed, their developer, availability (either in-person gameplay or digital), type of game (board game, role-playing game, etc.) and recommended number of players. Table 3 more specifically identifies each of the games' stated learning outcomes, their classification according to the categories identified by Grogan and Meijer [12], and how they each might contribute to building data capacity. As no specific classification system for serious games and data capacity exists, we broadly labeled each game as contributing to either *debate*, *data awareness* or *ideation*. Further research could investigate how to apply existing frameworks on data literacy, such as the ODI data skills framework [32], to serious games.

Table 2. Summary of the case descriptions.

Title	Developer	Availability	Type of game	Players
Agenda 2030	Teresina Municipality (Brazil/Piaui)	In-person	Card; discussion game	6-31
Data Belt	Independent [20]	Digital	RPG; collaborative game; quiz	4
Data Dealer	Independent [21]	Digital	RPG; resource management	1
Digital identity game	Provincie Zuid-Holland	In-person	Board game	2-6
Datak	dna studios for RTS	Digital	RPG; resource management	1
Datascape**	Independent [25]	Digital	Board game; Quiz	Unspecified
Dataspel	Provincie Zuid-Holland	In-person	Card game; discussion game	4
Datopolis	Open Data Institute [26]	In-person*	Board game	2-5
Jogo de Governo Aberto	IGA (Open Government Institute), Fast Food da Politica, and CGU (Comptroller General of the Union, Brazilian Federal Government)	In-person*	Card game	1-8
Open Data Card Game	Independent [28]	In-person*	Ideation game	Multiple groups of 3
Run that town	Millipede for the Australian Bureau of National Statistics [29]	Digital	Resource management	1
Winning Data	Independent [11]	In-person	RPG; collaborative game; quiz	4

*digital version available for online play
**not present in the original list by Kleiman [11]

Table 3. Case descriptions and their learning outcomes, categorization, and potential for data capacity building (continued on p.186).

Game Title	Stated learning outcome	Game category	Potential for data capacity building
Agenda 2030	Increase awareness on the importance of data sharing for the Sustainable Development Goals to be achieved	Teaching	**Debate** Participants are invited to discuss the need for datasets to be available in order to achieve the UN Sustainable Development Goals
Data Belt	Peer to peer knowledge transfer about the possible benefits and risks of opening certain governmental datasets	Teaching	**Debate** Civil servants can initiate discussions and share insights on the benefits (and consequences) of opening datasets, therefore building knowledge about opportunities to share data with the public

Table 3. Case descriptions and their learning outcomes, categorization, and potential for data capacity building (continuation from p.185).

Datopolis	Insight into the role played by open and closed datasets in order to build new services	Teaching	**Debate; Data awareness** Players have insight into the negotiations, collaboration and decision-making processes needed to open datasets
Jogo de Governo Aberto	What is open government? If you already know the topic and want to know more, or if you don't have the slightest idea what it's all about, but you're curious: this game is for you!	Teaching	**Data awareness** Players can understand the challenges to create open governments and the role of open data to it.
Open Data Card Game	"The aim is to make it easier for users to discuss and explore data, and generally to get people more excited about the potential of open data [...] The strength of this game comes from data-combining, which enabled participants to see the potential of this data in a new light." [28]	Design	**Ideation** Players can generate new ideas about how to combine and reuse datasets
Run that town	"[..] create awareness of the role of the census in shaping the direction of policy and its impacts on daily life" [29]	Policy	**Data awareness** Players can understand the role played by data in public policy and political decision-making
Winning Data	To influence civil servants' attitudes towards open data and nudge them towards opening more datasets while still considering privacy risks	Teaching	**Debate** In debating whether or not to open a certain dataset, civil servants share knowledge about the possible risks and benefits of opening data

DISCUSSION

We defined building data capacity as the process of empowering citizens and civil servants to reuse open data so that they can gain new insights about the world around them and create better services. With our two research questions, (1) we investigated which games – or types of games – have the potential to build data capacity and (2) what kind of capacity they can build. As shown in the case descriptions, serious games can play a significant role in building data capacity by raising data awareness, facilitating debate around open data and ideation for data reuse. However, from the review and analysis of existing games for building data capacity, it emerges that most games only focus on a limited aspect of this process, which is raising data awareness. In fact, most games only fit the teaching category identified by Grogan and Meijer [12]; meaning that they focus on transferring generalizable knowledge to the players or between the players.

Only one example of a design game was found through the literature review, the "Open Data Card Game". When using the game in a workshop, the facilitator can create card decks customized for the group that is about to play and insert datasets that the players might be already familiar with. The group can then use the custom cards to brainstorm together ideas for how to reuse these datasets, thereby generating new knowledge. The presence of only one design game suggests an interesting gap in games that can be used for ideation in the field of open data. Design games have been used to successfully facilitate idea generation in other fields. Brandt and Messeter [33] described several design games used for idea generation and found that games facilitate this process by creating artificial restrictions, which stimulate creativity. Agogué et al. [34] created a serious game for the employees of a company specialized in treatments for malnutrion. Each participant had to interpret a persona described by the game, for example "rural school director" or "deputy mayor of Jakarta East". Participants had to come up with new ideas that could create value for this persona. Game rules instructed players to divide in groups and change their composition at regular intervals. Finally, players could participate in a "marketplace of ideas" and work on the most promising proposals. Agogué et al. [34] found that serious games "play an effective role in supporting the management of heterogeneous and divergent knowledge during ideation" [34; p.423]. There is a need to explore the potential of serious games to play a similar role in ideation with data.

"Run that Town" is the only example of a policy game, which uses contextual knowledge to generate real-world scenarios. The game achieves this by looking at census data for the player's postcode, thus reflecting the real conditions of the neighborhood. The lack of policy games that make use of contextual knowledge is also an interesting gap. The review did not find any examples of research games, which are used to test or generate hypotheses or to assess other artifacts.

CONCLUSIONS

Our work presented a review of existing games that can contribute to building data capacity. To elaborate this review, we played several serious games and analyzed their content and game materials. We then categorized each game according to the type of knowledge it transfers and to which beneficiary. We also looked at the kind of capacity building that each game contributes to. The main finding that emerged through our review is that most games tend to build data capacity by raising data awareness. We found a lack of design games that can be used to generate new ideas about the reuse of open data. While this type of game has been successful in other fields, we only found one such example in the context of open data. Future research should explore the opportunities offered by different types of games, either by developing entirely new games or adapting existing ones from different fields.

AKNOWLEDGEMENT

This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 955569.

This research is part of TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592.

REFERENCES

[1] Open Knowledge Foundation: *Open Definition 2.1*. https://opendefinition.org/od/2.1/en, accessed 23rd April 2021,

[2] Kassen, M.: *A promising phenomenon of open data: A case study of the Chicago open data project*.
Government Information Quarterly **30**(4), 508-513, 2013,
http://dx.doi.org/10.1016/j.giq.2013.05.012,

[3] Zuiderwijk, A.; Janssen, M. and Davis, C.: *Innovation with open data: Essential elements of open data ecosystems.*
Information Polity **19**(1-2), 17-33, 2014,
http://dx.doi.org/10.3233/IP-140329,

[4] Van Loenen, B. et al: *Towards value-creating sustainable open data ecosystems: A comparative case study and a research agenda*.
eJournal of eDemocracy and Open Government (JeDEM) **13**(2), 1-17, 2021,
http://dx.doi.org/10.29379/jedem.v13i2.644,

[5] Mulder, I.: *On Digital Citizenship and Data as a New Commons: Can We Design a New Movement?*
Cuadernos del Centro de Estudios de Diseño y Comunicación **73**, 97-109, 2019,

[6] Abt, C.: *Serious Games.*
Viking Press, New York, 1970,

[7] Laamarti, F.; Eid, M. and El Saddik, A.: *An Overview of Serious Games*.
International Journal of Computer Games Technology **2014**, 2014,

[8] Fallon, C. and Reznitskaya, V.: *Socrates Jones.*
http://socratesjones.com,

[9] Bharosa, N.; Janssen, M.; Klievink, B; van Veenstra, A.-F. and Overbeek, S.: *Guiding Integrated Service Delivery: Synthesizing and Embedding Principles Using Role-Playing Games*.
Electronic Journal of e-Government **8**(2), 83-92, 2010,

[10] Van Lankveld, G.; Sehic, E.; Lo, J.C. and Meijer, S.A.: *Assessing Gaming Simulation Validity for Training Traffic Controllers*.
Simulation & Gaming **48**(2), 219-235, 2017,
http://dx.doi.org/10.1177/1046878116683578.

[11] Kleiman, F.: *Winning Data: Designing and testing a game to change civil servants' attitudes towards open governmental data provision*.

[12] Grogan, P.T. and Meijer, S.A., *Gaming Methods in Engineering Systems Research*.
Systems Engineering **20**(6), 542-552, 2017,
http://dx.doi.org/10.1002/sys.21409.

[13] Suits, B.: *Tricky Triad: Games, Play, and Sport*.
Journal of the Philosophy of Sport **15**(1), 1-9, 1988,
http://dx.doi.org/10.1080/00948705.1988.9714457,

[14] Sternon, J.: *The Game Definition Game: A Review*.
Games and Culture **12**(6), 499-520, 2017,
http://dx.doi.org/10.1177/1555412016655679,

[15] Djaouti, D.; Alvarez, J.P.; Jessel, J.-P. and Rampnoux, O.: *Origins of Serious Games*.
In: Ma, M.; Oikonomou, A. and Jain, L.C., eds: *Serious Games and Edutainment Applications*. Springer London, London, pp.25-43, 2011,
http://dx.doi.org/10.1007/978-1-4471-2161-9_3,

[16] Lee, D.: *Building an Open Data Ecosystem: An Irish Experience*.
Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance, Guimaraes, pp.351-360, 2014,
http://dx.doi.org/10.1145/2691195.2691258,

[17] Togelius, J. and Gustafsson Friberger, M.: *Bar chart ball, a data game*.
Foundations of Digital Games (FDG), Chania, pp.451-452, 2013,

[18] Gustafsson, M., et al.: *Data Games*.
Proceedings of the The 4th workshop on Procedural Content Generation in Games, pp.1-8, 2013,

[19] Katsaliaki, K. and Mustafee, N.: *Edutainment for Sustainable Development: A Survey of Games in the Field*.
Simulation & Gaming **46**(6), pp.647-672, 2015,
http://dx.doi.org/10.1177/1046878114552166,

[20] Di Staso, D.; Kleiman, F.; Crompvoets, J. and Janssen, M.: *A Digital Game to Learn About Open Data*.
20th IFIP WG 8.5 International Conference, EGOV 2021, Granada, pp.154-164, 2021,
http://dx.doi.org/10.1007/978-3-030-84789-0_11,

[21] Avertintsev, I.; Christl, W.; Osterwalder, P. and Traunsteiner, R.: *Data Dealer: Privacy? Screw that. Turn the tables!*
Data Dealer, 2012,
https://datadealer.com/about, accessed 2nd March, 2021,

[22] Zuboff, S.: *The Age of Surveillance Capitalism*.
PublicAffairs, 2019,

[23] RTS: *Datak: A game about the use of personal data*.
https://www.gamesforchange.org/game/datak, accessed 20th January, 2022,

[24] Jordan, P.: *Notes from the IAPP Europe Managing Director, 13 January 2017*.
Europe Data Protection Digest, 2017,
https://iapp.org/news/a/notes-from-the-iapp-europe-managing-director-13-january-2017, accessed 20th January, 2022,

[25] Wolff, A.; Barker, M. and Petre, M.: *Creating a Datascape: a game to support communities in using open data*.
8th international conference on Communities and Technologies, New York, pp.135-138, 2017,

[26] Broad, E. and Tennison, J.: *Datopolis board game*.
The Open Data Institute, 2015,
http://datopolis.theodi.org/about, accessed 14th January, 2021,

[27] –: *Jogo de Governo Aberto*.
https://tabletopia.com/games/jogo-de-governo-aberto, accessed 24th January, 2022,

[28] Goraya, J.: *Open Data Card Game*.
https://github.com/Atomland/open-data-card-game, accessed 14th April, 2022),

[29] Millipede: *Run That Town*.
https://millipede.com.au/work/run-that-town.html, accessed 14th April, 2022),

[30] Kleiman, F.; Janssen M. and Meijer, S.: *Evaluation of a pilot game to change civil servants' willingness towards open data policy making*.
Simulation and Gaming: through times and across disciplines, Warsaw, pp.286-297, 2019,

[31] Kleiman, F.; Janssen, M.; Meijer, S. and Jansen, S.J.: *Changing civil servants' behaviour concerning the opening of governmental data: evaluating the effect of a game by comparing civil servants' intentions before and after a game intervention*.
International Review of Administrative Sciences, 2020,
http://dx.doi.org/10.1177/0020852320962211,

[32] Open Data Institute: *Data Skills Framework*.
https://theodi.org/article/data-skills-framework, accessed 12th April, 2022),

[33] Brandt, E. and Messeter, J.: *Facilitating collaboration through design games*.
Proceedings of the 8th conference on Participatory design: Artful integration: interweaving media, materials and practices **1**, New York, pp.121-131, 2004,
http://dx.doi.org/10.1145/1011870.1011885.

[34] Agogué, M.; Levillain, K. and Hooge, S: *Gamification of Creativity: Exploring the Usefulness of Serious Games for Ideation: Gamification of Creativity*.
Creativity and Innovation Management **24**(3), pp.415-429, 2015,
http://dx.doi.org/10.1111/caim.12138.

Interdisciplinary Description of Complex Systems 20(2), 190-209, 2022

TOWARDS DIGITAL INNOVATION: STAKEHOLDER INTERACTIONS IN AGRICULTURAL DATA ECOSYSTEM IN CROATIA*

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

[1]University of Zagreb, Faculty of Organization and Informatics
Varaždin, Croatia

[2]University of Zagreb, Faculty of Agriculture
Zagreb, Croatia

[3]Centre of Excellence for Biodiversity and Molecular Plant Breeding
Zagreb, Croatia

[4]Delft University of Technology, Faculty of Architecture and the Built Environment
Delft, The Netherlands

DOI: 10.7906/indecs.20.2.10
Regular article

Received: 17 January 2022.
Accepted: 26 April 2022.

ABSTRACT

The successful adoption of digital innovations in agricultural production systems is based on the proactive participation of all stakeholders and represents an important step in establishing resilient agri-food chains and creating sustainable value. The key tool for the creation of sustainable value is integrating the nine aspects of the business process (cooperation; inclusion; financing; diversification; communication; policies; knowledge with entrepreneurship; and production) by re-using of open governmental and public endeavours data as well as by the contractual sharing. The objectives of this research are to identify stakeholders in the Croatian agricultural system, and to explore their roles and their potential for data supply and needs for data uptake. Open access repositories were queried to identify stakeholders. Direct observation methods and semi-structured conservational qualitative interviews were used for stakeholder characterisation and data flow detection. Stakeholder importance with respect to current data supply was analysed. Underdeveloped data flow relationships in the agricultural data ecosystem in Croatia could be built in a spontaneous process following the data opening of the Research and Consumer group of stakeholders and promoting data sharing initiatives of the early adopters in the Supplier group. In that way, data opening would be the driver of the effective cooperation creation required for sustainable value creation but also the adoption of the best management practices, sustainable solutions and digital development.

KEY WORDS

agriculture data ecosystem, open data, stakeholder analysis, sustainable value

CLASSIFICATION

JEL: O13, Q01, Q15, Q16

*This is the extended version of the abstract published in: Vujić, M. and Šalamon, D., eds.: Book of abstracts of the National Open Data Conference. University of Zagreb, Faculty of Traffic and Transport Sciences, Zagreb, 2021.

**Corresponding author, η: lhrustek@foi.unizg.hr; +385 42 390828;
Faculty of Organisation and Informatics, Pavlinska 2, HR – 42 000 Varaždin, Croatia

INTRODUCTION

The agricultural sector plays a significant role in global and regional development. However, despite its strategic importance, too little attention is still paid to building the key elements in creating value and establishing sustainable practices in business processes [1]. The key elements of strategic planning, decision-making, and management in complex systems such as agriculture are (i) using and integrating the emerging digital technologies; (ii) the circular approach and participation; (iii) transferring the knowledge and the appropriate policies into agricultural business process practices [1-6]. An approach that combines these key features improves the business performance [5] but also creates sustainable agricultural value [1].

The framework for sustainable value creation in the agricultural sector [1] focuses on activities creating value currently and in the future, and it includes maintaining the internal environment but also emphasizes simultaneous interaction with the external surrounding. Therefore, the integration of the following nine aspects: cooperation; inclusion; financing; diversification; communication; policies; knowledge with entrepreneurship; and production enables the creation of sustainable values. The key tool for the integration of these aspects is the data sharing (either as re-using of the open data or as contractual sharing), enabled by the effective data governance [7-14].

Sustainable data-based value creation is a common approach that allows different stakeholders to influence decisions at different governmental levels of a complex structure [2]. In this research, we focused the circular perspective of the agricultural data ecosystem on the cooperation of all stakeholders and investigated data sharing needs based on the key characteristics and role of individual stakeholders in the sector and their role in the supply and demand for data.

Cooperation and collaboration between the stakeholders in the agricultural sector through engagement, motivation and capacity to act together enables economic opportunities and ensures equal access to information and resources. It also contributes to the promotion of best management practices, the ones crucial for sustainable development and promoting effective resource management [3-5, 14, 15-17]. However, despite the focus on joint planning and management in agricultural operations, at the moment the effective cooperation remains limited [2]. The reasons are the complexity of the agricultural system, the large number of stakeholders with different perspectives, interests, values and concerns regarding business processes [2] which operate in an ecosystem where data and information exchange is lacking or is decentralized [6].

The open data ecosystem is a concept developed by emerging governments to encourage sharing and reuse of data, and as such includes key components which are policies for legal context, standards for interoperability, and an access network available for all stakeholders in the ecosystem [18, 19]. Open data ecosystem performance can be observed through three types of output indicators, namely data supply, data governance and user characteristics [18]. Open governmental data is an important part of the agricultural data ecosystem, however, creating value in this sector requires also a significant portion of different contractual data sharing [14].

This research contributes to the usage and value elucidation of open and shared data in the agricultural data ecosystem by defining and prioritizing the stakeholders through queries of open source databases, by discovering the data needs based on the stakeholder groups and alliances as well as the assigning the level of influence for the data supply in this data rich sector. The revealing of the potential for the development of data supply, taking into account stakeholder relations, will enable the development of a more resilient and sustainable ecosystem for agricultural data sharing in Croatia.

In the Croatian agricultural sector, data are often vague, scattered or not easily accessible [20], and in many cases when farmers (SMEs) are to obtain management, market or other information, they rely on informal channels such as a personal network of agricultural contacts (personal account and spoken-information). The farmers' associations, governmental and business advisory services do not yet recognize the valuable potential of the data, therefore, it is important to understand the groups of stakeholders and their relationships in the agricultural open data ecosystem in order to address the needs and problems of the data and information exchange and to formulate strategies and recommendations for further sector development.

The objectives of this research are as follows:

1) to identify all present and potential stakeholders in the agricultural open data ecosystem in Croatia and categorize them into key groups,
2) to define data sharing in this ecosystem, based on relationships between stakeholders and their role in the supply and demand for data,
3) to add to the knowledge of the key elements of the agricultural open data ecosystem in Croatia and prioritize further research.

METHODOLOGY AND MATERIALS

Stakeholder analysis (SA) is an approach used to define and understand a complex network of actors focusing on identifying key stakeholders, assessing their interests and needs, and clarifying how they can impact sustainability and improve processes in a particular research field [21, 22]. This research is based on the stakeholder methodology approach developed by Grimle and Chan [22, 23], where they consider the stakeholder analysis as a powerful tool for analysing the situation of the field, formulating policies, and developing programs based on an approach of understanding the observed system, changing it, identifying key actors or stakeholders and assessing their interests in the system. Lelea et al. conducted a transdisciplinary study in the field of agricultural and food systems where they developed a methodology for stakeholder analysis [24].

Figure 1 shows the four steps of the information collection, classification and validation analysis used for stakeholder analysis in the agricultural data sharing ecosystem in Croatia. In our research we have focused on the selecting of an activity system in agricultural sector and centering the issue to be addressed to data sharing. In the stakeholder’s analysis we have focused to identify and characterize the actors based on the initial categorization of the stakeholder groups (Figure 2): (i) Agricultural producers/ farmers; (ii) Management and support organisations; (iii) Consumer organisations/ consumers; (iv) Researchers and scientists and (v) Suppliers. These groups were the basis for selecting the participants for our research [24]. Identification of stakeholders in key groups was made based on the on-line queries. Stakeholder characterisation and the research of their roles in data supply and demand in the sector, as well as a proposal for integration and their potential contribution to the development of a collaborative network and data sharing model, were obtained from the queries, direct observations and interviews. Stakeholder interviews were conducted as an important source of qualitative information in this stakeholder analysis [25]. Validation of stakeholders as data providers was carried out by this research group based on information collected from interviews and direct observation.

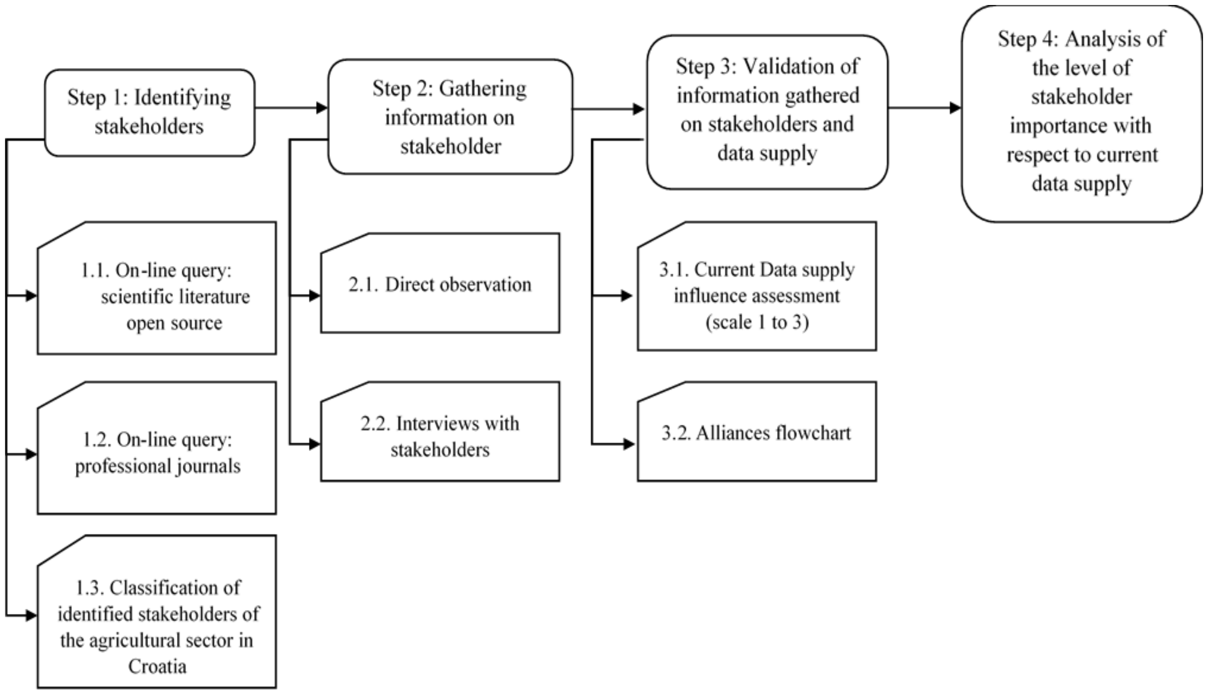


Figure 1. The four steps of agricultural stakeholder mapping in the context of data sharing ecosystem in Croatia.

Step 1: Identifying stakeholders

Open access repositories of scientific and professional publications were chosen for the initial stakeholder identification. Digital Academic Archives and Repositories (Dabar) and the Portal of Croatian scientific and professional journals – Hrčak were queried for scientific publications. Professional journals from Hrčak were searched by field, focusing on the journals in the field of agriculture (section: “agronomy”).

The complex query: "stakeholder" OR "persons" OR "actors" OR "agriculture" OR "agriculture business" OR "farms" OR "agriculture sector" OR "agriculture area" OR "agriculture field" AND "open data" was used for search of the national databases Hrčak and Dabar [26]. As shown in Table 1, a search based on a complex query in the Hrčak database resulted in a total of 63 literature sources. This query did not yield a result in the Dabar database, i.e. it resulted in 0 articles found. Therefore, a simplified query was used in Dabar with the keywords “agriculture” and “stakeholders” and resulted in 18 literature sources. In the Hrčak database, a search of professional journals resulted in 24 journals in the field of agriculture.

An important aspect of stakeholder analysis is based on differentiating and classifying stakeholders according to their roles in the ecosystem. This form of categorization enables the grouping of stakeholders according to similar characteristics in order to identify those of strategic importance [9]. Identification of relationships between all stakeholders can help determine how stakeholders could engage in an open data ecosystem for policy implementation, support the decision-making of sustainable development, and manage business processes. Building on the defined basic categorization of stakeholder groups in the agricultural sector in Croatia [26] the stakeholders identified in the queries were classified and the new grouping was used in further data supply influence assessment and alliance flow chart production encompassing also the data demand side, as well as for the final prioritization (Figure 2).

Table 1. Number of articles and journals in national databases.

Research goal	Research query/field	Database	Results
Additional categorization of stakeholders in agriculture open data ecosystem in Croatia	agriculture OR agriculture sector OR agronomy OR farming OR smart farming OR smart agriculture OR precision agriculture OR agriculture field AND open data	Hrčak	63
	agriculture AND stakeholders	Dabar	18
	Professional journals; field Agriculture (Agronomy)	Hrčak	24

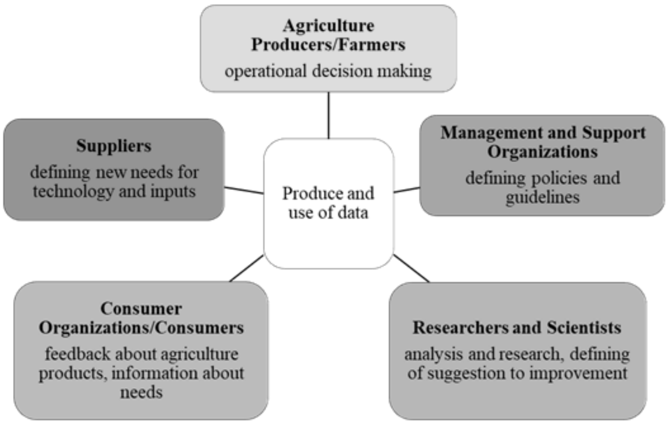


Figure 2. Preliminary key stakeholder groups of agricultural open data ecosystem in Croatia used for the classification of the open access sources query results [26].

Step 2: Gathering information on stakeholder connections and data supply influence

To identify the basic characteristics and roles of individual stakeholders in the agricultural data ecosystem, to investigate their data requirements, as well as the data that are the result of their work, field research was performed including direct observation and semi-structured interviews with the following questions (in Croatian):

1) What is your role, as a stakeholder, in the agricultural sector?
2) Which of the 5 stakeholder groups do you have the most cooperation and interaction with? Can you specify?
3) Do you share your management/business/product data with any of the 5 stakeholder groups? Can you specify to whom and how?
4) From the stakeholders identified in question 2, do you require/use any data sets and can you specify? Would it be useful to you to get some data for your business from some other entity you are currently not cooperating with?

A breakdown of the main stakeholder groups (Figure 2) provided the structure for interviewing the representatives of the five stakeholder groups. In total, 24 interviews were used for validation: 5 interviews per group of stakeholders were performed with representatives of Agriculture producers/Farmers, the group of Suppliers and the group of Customer Organization/Consumer; and 3 interviews per group of stakeholders were performed with representatives of Management and Support Organizations, as well as the Researchers/Scientists from the field of agricultural sciences. The additional 3 representatives were interviewed from the group of other stakeholders after the classification of the query results.

Characteristics and roles of stakeholders were additionally supplemented with 15 sets of direct observations for the 10 of the interviewed stakeholders (3 from Management and Support Organizations, 3 from Suppliers, 2 from Researchers/Scientists, and 2 from Agriculture producers/Farmers) and the randomly selected 5 additional representatives from each of the stakeholder groups. All available sources, including official web sites and on-line available documents, were reviewed for (i) the data needs, (ii) data sharing and (iii) data sharing possibilities.

Step 3: Validation of information gathered on stakeholders and data supply and demand

Qualitative data from the interviews were used to extract the information on data needs and data sharing of the stakeholders and stakeholder groups. Also for the data opening and sharing supply assessment and to formulate the alliance's flowchart. The alliances flowchart contains information on the stakeholder group relevance in the system, the data supply influence and the alliances derived from the interviews, direct observation and the query results.

Current data supply influence assessment was estimated on a scale of 1 to 3, where (i) 1 is considered as stakeholder data openness or sharing not known or exceptional; (ii) 2 is considered as stakeholder data known to share, some data known to be open; (iii) 3 is considered as stakeholder relevant open data supplier or data sharer. The interview results and the direct observation from the previous research step were used for this assessment according to this research group's judgment. Scoring results were included in the alliances flowchart to distinguish the groups assessed as relevant data suppliers, as well as to distinguish a specific stakeholder relevant in the stakeholder group, but scoring low as data supplier.

Query results provided the number of the references in common for specific stakeholders and stakeholder groups. This information was used as an indicator of the stakeholder interactions and supplemented the interview and observation information for the final alliances flowchart. The level of importance of each stakeholder and stakeholder group used for the alliances flowchart was derived from the frequencies of the queried references shown in Table 2. Based on the collected data, the relations between stakeholders were constructed [27, 28, 29].

Step 4: Analysis of the level of the stakeholder importance with respect to current data supply

Finally, to discover the relevant priority groups for further focus in open data ecosystem maturation, we have imposed the level of the stakeholder importance (derived from the number of the query references) to the level of the stakeholder influence to the data supply [24].

RESULTS

Key stakeholders in Croatia

The content of the found articles was analysed, and special emphasis has been placed on identifying specific stakeholders operating in or in relation to the agriculture sector. All stakeholders identified in the articles are classified to the key stakeholder groups as shown in Table 2. Additional group (Others) was formed containing the stakeholders that could not be classified to the one of the initial five stakeholder groups at all, or exclusively.

Table 2. Identified stakeholders in the agricultural data ecosystem in Croatia (continued on p.197).

Agriculture producers/Farmers	owners of agricultural land [30, 31] milk producers [32] animal breeders, cattle breeders [33] meat producers [34, 35] vegetable producers [36] forestry entrepreneurs [37] fish farmers, fishermen [38] producers of organic agriculture products [39] flower producers [40] wine and winery producers [41] energy plantations [42] family farms, local farms, rural holdings [39, 43] processors of agricultural products (milk, meat, fruits, vegetables) [44] mushroom growing [45]
Suppliers	manufacturers of technological solutions and mechanization in agriculture [46] hatcheries, rearing parent stock [47] seed growers [45, 47] seedling growers [48] manufacturers of plant protection products [45, 48] agriculture machinery market [49] agricultural cooperatives [49] forest owners [31] landowners [31]
Management and Support Organization	Ministry of Agriculture; Croatian Agency for Agriculture and Food; Agency for Payments in Agriculture, Fisheries and Rural Development [50, 51] Croatian Agricultural Advisory Service [52] national training providers in the agricultural sector [37] agriculture local action groups [41] developed agencies in rural development and agriculture [41] independent consultants in agriculture [41] agriculture producers cooperatives and local partnerships [39, 41] creators of agricultural programs [53] issuers of certificates in agriculture [54] Croatian Meteorological and Hydrological Service [55] meteorological stations [56] the institutional Animal Care and Ethics Committee [50] State Geodetic Administration [57] Croatian Veterinary Institute, veterinary institute [58] Croatian Agricultural Society [48, 59], Croatian Society of Plant Sciences [45]
Consumer Organizations/Consumers	business entities in tourism [60] private accommodation, hotels [61] organized gastronomic events [62] local population, household [39] buyers of agricultural products [49]

Table 2. Identified stakeholders in the agricultural data ecosystem in Croatia (continuation from p.196).

Researches and Scientists	educational institutions [34, 39]: Faculty of Agriculture [45, 49], Faculty of Agrobiotechnical Sciences [45] editorial boards of professional journals students of agriculture and agronomy [46] laboratory centres [58] researchers [51], research organizations [63], multidisciplinary research teams [64] forestry experts [37] organizers of educational programs [39]
Others	owners of organic gardens and ecovillages [39] hunting and hunting tourist centres [65] botanical gardens [66, 67] school gardens [48] media: agricultural portals, web pages [39] employees in agriculture [54]

The group of Agricultural Producers and Farmers includes all stakeholders who are primarily engaged in the production of any agricultural product, processing of agricultural lands, or animal husbandry. Only those producers who offer final agricultural product for further processing or sale are included in this group of stakeholders. Some of the agricultural producers in the agricultural data ecosystem are producers of milk and dairy products, meat, fruits, vegetables, flowers, and agricultural organic products. This group includes animal breeders who resell their breeding, but also grain producers, owners of agricultural land who grow various crops for food purposes. Also included are sustainable users of wild populations (e.g. fishermen and forest owners who sell their timber resources, and mushroom pickers), fish farmers and wineries. Apart from the type of agricultural activity they perform, Agricultural Producers and Farmers also differ in size, so in this division in Croatia, there are family agriculture farms, as well as small, medium and large agriculture businesses.

Suppliers are all those stakeholders who supply agricultural producers and farmers with all the necessary resources for work such as machinery and other technological solutions, plant protection products, re-selling seeds and seedlings, feed, flocks of animals, and more. These include producers of agricultural machinery and technological solutions for agriculture, hatcheries, producers of seeds and seedlings, producers of plant protection products, owners of land and forests who rent out their land.

Management and Support Organizations include all stakeholders who provide any form of support to farmers and enable the functioning and operations of all stakeholders in the ecosystem. Identified stakeholders can be divided by levels. The highest level representative is the Ministry of Agriculture and it oversees some of the lower level stakeholders. Agricultural advisory services operate at the local level but under the authority of the Ministry. National training providers in the agricultural sector and issuers of certificates in agriculture who provide support to farmers in terms of education, also operate under the authority of the Ministry. Furthermore, various agencies and consultants with advisory capacity operate at the local and regional levels. In addition to all the above, support in agriculture is provided by some organizations from related other disciplines such as the State Geodetic Administration, the Croatian Veterinary Institute, the Croatian Meteorological, and Hydrological Service, and others.

Consumer Organizations and Consumers are all stakeholders who buy and use agricultural products, produced exclusively by agricultural producers or farmers. These are households, private and business entities, organized events and companies that process agricultural products, and direct consumers.

The group of Researches and Scientists includes faculties and all educational institutions in the field of agriculture in Croatia. Educational institutions usually include research groups and laboratories, multidisciplinary teams, project teams, and experts in the field of agriculture. Students of agriculture are included in this group.

Finally, to the five basic stakeholder groups, the category of other stakeholders was added. This group includes stakeholders who cannot be included in any of the previously described groups. This group includes owners of organic gardens and ecovillages, botanical gardens, school gardens, and hunting and hunting tourist centres. Employees in agriculture and households that produce food for their own needs belong to this group. In addition to them, there are also media that cover agricultural activity and events related to it.

The role of stakeholders in the agriculture data ecosystem

From the agricultural producers and farmers group, the interview was attended by family agriculture farms and small agriculture businesses, who have less than 10 employees (Figure 3). They unanimously agreed that their role in the ecosystem is to produce and supply food and drink to the community. For the producers, the most important group of stakeholders are the suppliers who supply them with production resources. The management and support organizations are in charge of controlling food production and enabling support for the agriculture production. All of the agricultural producers mentioned that they have specific certificates from the field in which they operate and that they have acquired most of them through organized training of the Ministry of Agriculture and the Advisory Service. Some of the producers mentioned that they cooperate with consultants, i.e. organizations that prepare projects for them co-financed from European Union funds. Also, all of them are in communication with the Agency for Payments in Agriculture, Fisheries, and Rural Development, which takes care of the operational implementation of direct support measures and rural development measures. All farmers stated that they have no contact with researchers, scientists, and educational institutions.

The data needs of agricultural producers are diverse and depend on the type of activity they are engaged in. For example, the vegetable producer pointed out that the most important data for him are data about current protection products and substitute protection products, and data about demand and prices on the market. Meat producers stated that they need data on the quantities and prices of all food resources they use to prepare food on their farms, followed by data on the quantities and prices of final products on the market, data on available production incentives, and investment opportunities. All farmers agreed, that the data they can generate is related to the parameters of agricultural production.

From the supplier group of stakeholders, a regional representative of a company engaged in the sale of seeds, plant protection products, and the purchase of mercantile cultivation, an employee in an agricultural market, and a landowner who rents out his land were interviewed (Figure 4). The regional representative of the agricultural corporation stated that their business is based on cooperation with large and small agricultural producers who buy from them seeds, plant protection products, and mineral fertilizers, as well as redeem mercantile goods and lend for further production. An employee of the agricultural shop stated that they supply the local population and households with resources for agricultural production, and cooperate with larger farmers as agreed. The owner of the land pointed out that he does not cultivate the land, but rents it to an agricultural farm.

The regional representative emphasized the wide range of stakeholder partners: large and small agricultural producers, buyers of mercantile goods, large agricultural corporations such as feed mills, competitors, and others. An employee of an agricultural market mentioned households cultivating their gardens and small farmers as key stakeholders. The landowner, since he rents out his land, cooperates exclusively with the agricultural producer to whom he rents the land, the data flow is shown in Figure 4.

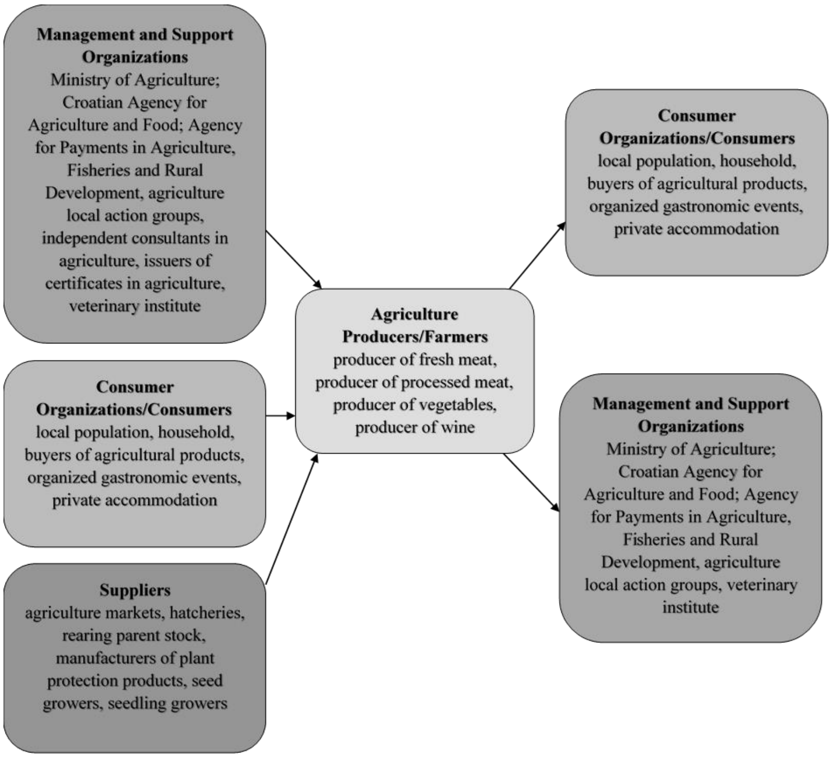


Figure 3. Primary needs of data flow as recognized in interviews with the agriculture producers.

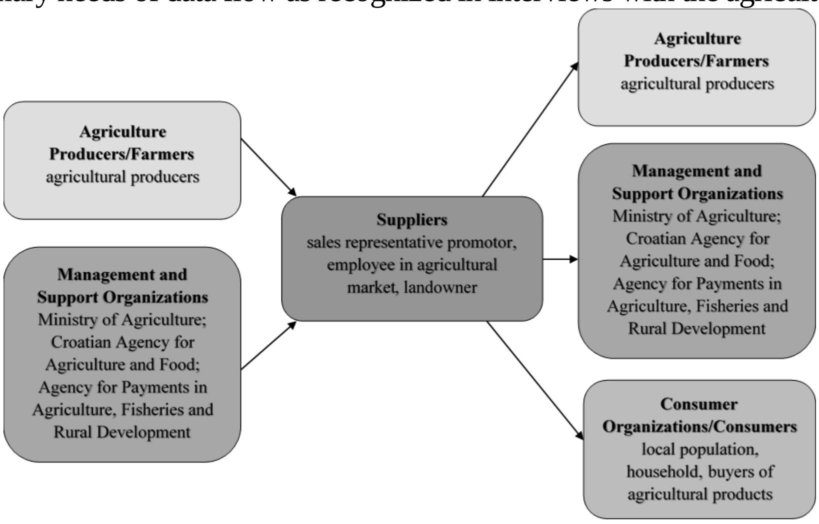


Figure 4. Primary needs of data flow as recognized in interviews with the suppliers.

The most important stakeholder representative from the group of the Management and Support Organization is the Ministry of Agriculture. The role of the Ministry is clear and refers to administrative and other affairs in the field of agriculture. From this group, an informal interview was conducted with an employee of the Advisory Service. They cooperate with farmers in the implementation of various experiments, where, based on the monitoring of agricultural production, they receive information on agricultural practices as well as the pros and cons of agricultural production. In addition, the role of the Advisory Service is to provide assistance to farmers through advice on agricultural production and to apply for various measures and incentives. They cooperate with suppliers in conducting experiments in different agriculture fields. For example, a regional representative and a representative of the Advisory Service, together with an agricultural producer, described their cooperation in testing new seeds.

In the group of consumer organizations, a local restaurant was selected for the interview, which procures its food exclusively from agricultural producers and farmers. The owner of the restaurant stated that he has big problems when procuring fresh groceries, especially fresh meat. Mostly information about producers who offer fresh and quality meat comes by word of mouth from acquaintances. Therefore, the lack of data on agricultural producers and the lack a network of producers to whom they can turn, are a problem for him in performing his business and sometimes he cannot satisfy the entire offer in his restaurant.

The scientists and researchers group uses the data provided by different stakeholder groups, or generates the data with their research. Importance of data provided from the private sector is crucial in order to improve their scientific and research work, develop new projects or suggest better solutions to certain problems. Moreover, based on the interviews, generated data of different granularity from research is provided to the farmers/source of the data often only upon the specific request and there is no obligation or channel for the direct feedback to source of the data.

Data availability is also important for the group of Other stakeholders. The media inform the wider community about the current state of the sector and individual activities, other stakeholders and also about agricultural products. School and kinder gardens serve to teach children from an early age about farming, especially vegetables. The Botanical Garden serves as a centre where it is possible to be educated and introduced of the wider community to the cultivation of plants of different species.

Alliances of the stakeholders in the agricultural data sharing ecosystem in Croatia are shown in the Figure 5. The most striking result is the perceived lack of data and information feedback of the Research group to the group of Agriculture producers / Farmers, which is in this research limited to smallholders. Additional lack of data sharing alliances is evidenced between the groups of the Suppliers and the Researchers. The group of Consumer Organisations and consumers does not have any detected data sharing relations with the Management and support group using our methods. The group of Agriculture producers / Farmers shares data with the Other stakeholders as was recognised from the interviews and direct observations.

Stakeholder importance with respect to current data supply

The stakeholder group matrix in Figure 6 emphasizes the two stakeholder groups distinguished from the others in the quadrant of the estimated high level of importance in the agricultural data sharing ecosystem in Croatia and the low estimated data supply for that ecosystem. Surprisingly, one of the groups is the Research and scientists and the other are the Suppliers.

Management and Support Organizations are recognized as the group of the significant importance and influence in data supply in the agricultural data sharing ecosystem in Croatia. The other stakeholder group in the same high influence and high data supply sector are the Agriculture producers and Farmers.

Consumer Organizations/Consumers and others, as stakeholders not directly involved in agricultural activities, have less importance and influence in the agriculture data ecosystem.

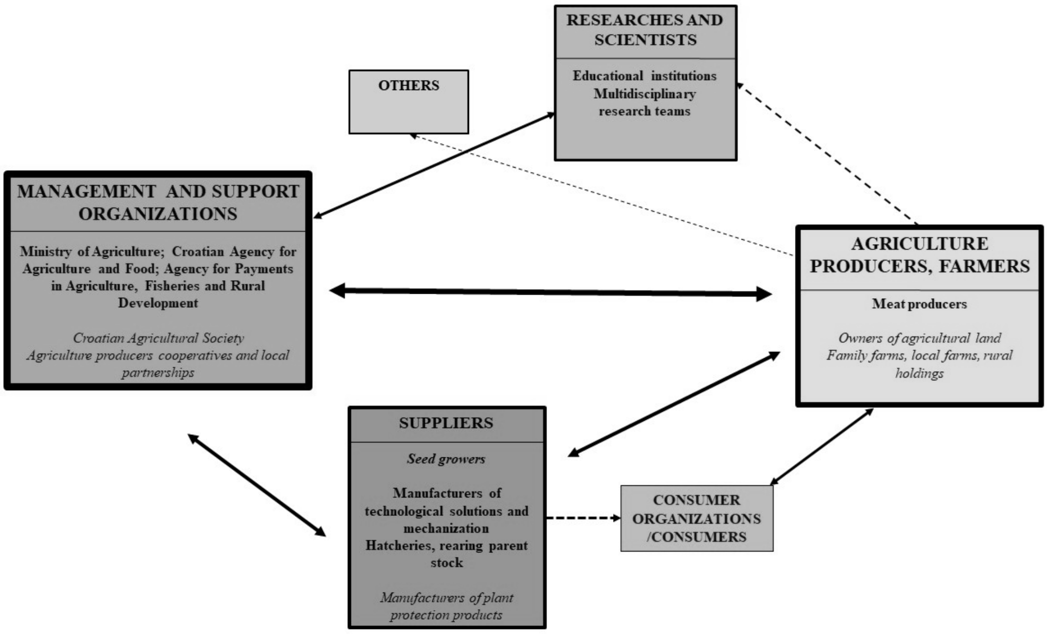


Figure 5. Alliances of the stakeholders in the agricultural data sharing ecosystem in Croatia.

However, the low level of data from that stakeholder was found to be concerning in the interviews where the group of Agriculture Producers and Farmers emphasized their need to know the market prices and market demand.

The shape size represents the assessed level of importance of the stakeholder group. The thickness of the shape border indicates relevant data suppliers. Two-headed arrows are data sharing alliances derived from the queried references and one-headed arrows are derived from the interviews and direct observations. The thickness of the arrow indicates the estimated relevance of the communication. Stakeholders with estimated high importance and low influence in data sharing are shown in italic and the stakeholders with high importance and influence are shown in bold font.

Distinguished stakeholders, when superimposing importance in the agricultural data sharing ecosystem in Croatia to the data sharing contribution, were examined from each group of the

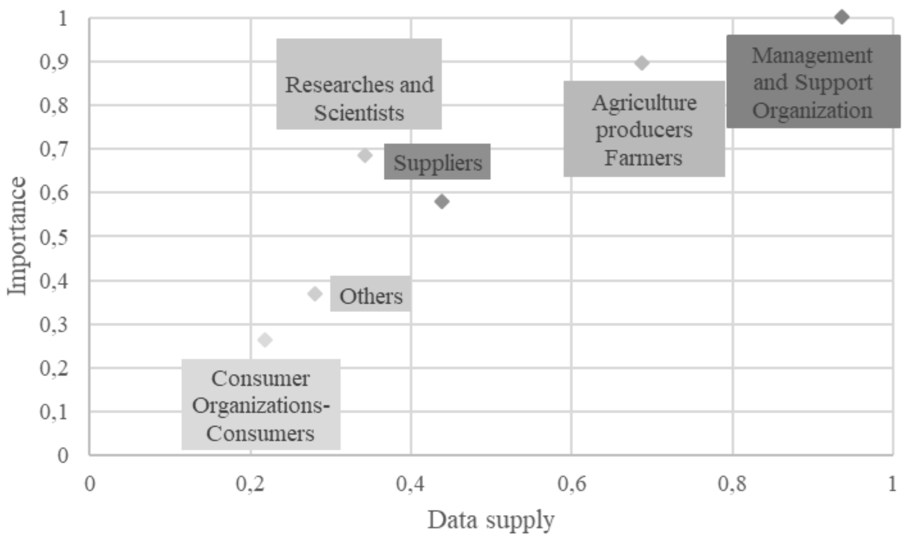


Figure 6. A stakeholder map matrix in the agriculture data ecosystem.

stakeholders (shown in the Figure 5). In the Agriculture Producers/Farmers group none of the stakeholders assessed as important have an average share in the data sharing. A high level of importance, but a low impact on the data supply was recorded for the owners of agricultural land, meat producers and family farms, local farms and rural holdings. Milk producers, producers of organic agriculture products and processors of agricultural products (milk, meat, fruits, vegetables) scored high in the group for data supply, but are assessed as less important.

In the group of Suppliers, the stakeholders that stand out with high assessed importance and low impact to data supply are the manufacturers of plant protection products. The seed growers are assessed as important and sharing their data. Manufacturers of technological solutions and mechanization in agriculture, hatcheries, seedling growers, and agricultural cooperatives scored high in the group for data supply, but are assessed as less important.

As expected, the Management and Support Organization records the highest level of importance, but also a high level of influence in the agriculture data ecosystem as well as data supply. The Ministry of Agriculture scored the highest together with the Croatian Agency for Agriculture, Croatian Agricultural Advisory Service and Food and Agency for Payments in Agriculture, Fisheries and Rural Development. Agriculture producer's cooperatives and local partnerships and the Croatian Agricultural Society were assessed as important stakeholders, but sharing less data then the other stakeholders in the group.

In the group of Researches and Scientists educational institutions (i.e. Faculty of Agriculture Faculty of Agrobiotechnical Sciences) and researchers (research organizations and multidisciplinary research teams) are distinguished as important and sharing more data than the other stakeholders in the group.

All stakeholders of high importance had high data influence in the Others group. In the Consumer Organizations/Consumers all stakeholders were assessed to be equally important and most of the data sharing in this group of stakeholders comes from the business entities in tourism and the organized gastronomic events.

DISSCUSION AND FUTURE WORK

The analysis of key stakeholder groups in the Croatian agriculture data ecosystem provided identification and characterisation of stakeholders and their relationships inside the agriculture business processes, with several relevant outside connections in the group of Management and support. Our analysis included different perspectives in gathering the data, and only further participatory approach in stakeholder analysis would add valuable qualitative and quantitative data and aspect contribution if more stakeholder details are required [24].

This stakeholder analysis results should prove valuable in developing collaboration, joint projects, or policies, but also in solving sectoral business problems where the participatory approach is required [24]. In the further maturation of the agricultural data ecosystem in Croatia, both for the open governmental data and the data of the public endeavours as well as with developing the contractual sharing and the effective data governance, the critical findings of the underdeveloped relationships, need for better data supply should be taken into account [7-14]. Despite numerous initiatives for cooperation and data sharing between stakeholders in public and private sector at different levels, limited impact to sustainable value creation has been achieved in industries including agriculture, and unsustainable practices persist [14, 68, 69]. Expanding on the initiatives of open data ecosystem readiness and the ability to identify the different ways in which stakeholders share data, build sustainable practices and systems is crucial to the successful adoption and implementation of innovation. Successful integration of sustainability aspects into innovation requires the collective participation of different stakeholders, matching objectives among stakeholders, and also their expertise as well as specialized roles for clear transfer of added value within the ecosystem [68, 70, 71].

In order to abandon unsustainable practices and to adapt to new digital agricultural practices it is important that the Research and scientist as well as the Supplier group of stakeholders promote their data governance and commit to data opening and sharing in the agriculture data sector in Croatia. It is concerning no data flow was indicated in this research results between the Research and Supplier groups. Additional emphasis to importance of opening up the stakeholders mapped to the Research and scientist group in the agricultural sector in Croatia are the results of the interviews showing limited data flow from those stakeholders to the group of Agriculture producers and Farmers, as seen from the perspective of the interviewed smallholders. Especially when considering smallholders are the ones actually feeding the population [72]. Aside from the food production, the smallholders are the source of a large amount of high-value data for all other stakeholders in the agricultural data ecosystem in Croatia. In addition, their data needs, according to the interviews, are increasing given the persistent disruptions in the agricultural sector due to environmental, economic, and social global challenges [72-74]. Market data considering the prices and the demand are the Consumer stakeholder data important to the Agriculture Producers and Farmers group that are not enough open and re-usable.

Also, group of Suppliers emphasized data needs, as they depend on market trends and the needs of agricultural producers, i.e. their survival depends on performance in the agricultural market. The data of the Croatian agricultural research sector is evidently not available to them and the contractual data sharing from the group of Agriculture producers and Farmers is evidently limited. Management and Support Organizations have/gather and open the most data according to the above results, but opening of this data is limited [20] and more care should be given to data governance research in this group in order to boost the usability of this data for solving relevant sectoral problems. The strong open data outreach responsibility of this stakeholder group is even more pronounced by the fact that the effective cooperation required for sustainable value creation is lacking in the complex business systems, such as agriculture, which operate in an data ecosystem where data and information exchange is lacking or is decentralized [6].

CONCLUSION

The agricultural sector in general is a mine of valuable data generated in agricultural processes, however, its’ potential to generate value is not yet exploited in the ecosystem of agricultural data in Croatia. The concept of smart and precise agriculture often referred to as the digital agriculture includes collecting and systematizing data from agricultural business processes, integrating, monitoring, analysing, and interpreting data, enabling the development of sustainable practices. To enable the transition to digital agriculture and the data to generate the value, it is necessary to strategically invest into maturation of agricultural data ecosystem (interaction of people, infrastructure, and processes) in Croatia working on the underdeveloped alliances by opening data of the identified key stakeholders.

The Management and Support Organization group of stakeholders is recognised as the most important and most data influential one with the most responsibility in promoting open governmental data and open data of the public endeavours. The Agriculture producers/Farmers are the second most important and data influential stakeholder group providing data to and trough other stakeholder groups, primarily the corresponding ministry. The Suppliers group of stakeholders was characterized in this research as the one not connected well enough through the data flow with the other stakeholders. The group of Researches and Scientists in the agricultural sector in Croatia were characterized as not contributing enough to the open data ecosystem. The group of consumers is not reaching the producers with the data they require. Underdeveloped data flow relationships in the agricultural data ecosystem in Croatia could be built in a spontaneous process following the data opening of the key stakeholders and promoting data sharing initiatives of the early adopters. In that way, data opening in the agricultural data ecosystem in Croatia would be the driver of the effective cooperation creation required for sustainable value creation but also the adoption of the best management practices, sustainable solutions and digital development.

Based on the emphasized requirements from the Research group for the Supply group, and all groups for the group of Agriculture producers and Farmers, as well as recorded importance of the word-of-mouth for the market information and food product availability, it is evident that the data supply of many stakeholders is underdeveloped. Future research focusing on data demands of the stakeholders could enable faster development and maturation of the agricultural data ecosystem in Croatia.

ACKNOWLEDGEMENT

This research is part of TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592.

REFERENCES

[1] Sadovska,V.; Ekelund Axelson, L. and Mark-Herbert, C.: *Reviewing Value Creation in Agriculture—A Conceptual Analysis and a New Framework.*
Sustainability **12**(12), 5021, 2020,
http://dx.doi.org/10.3390/su12125021,

[2] Nyam, Y.S.; Kotir, J.H.; Jordaan, A.J. and Ogundeji, A.A.: *Developing a Conceptual Model for Sustainable water Resource Management and Agricultural Development: the Case of the Breede River Catchment Area, South Africa,*
Environmental Management **67**(4), 632-647, 2021,
http://dx.doi.org/10.1007/s00267-020-01399-x,

[3] Ningrum, E.R., et al.: Stakeholder analysis for controlling land utilization in Parangtritis sand dunes.
IOP Conf. Ser.: Earth Environ. Sci **686**(1), 012019, 2021,
http://dx.doi.org/10.1088/1755-1315/686/1/012019,

[4] Cansino-Loeza, B. and Ponce-Ortega, J.M.: *Sustainable assessment of Water-Energy-Food Nexus at regional level through a multi-stakeholder optimization approach.*
Journal of Cleaner Production **290**, 125194, 2021,
http://dx.doi.org/10.1016/j.jclepro.2020.125194,

[5] Ahrisa, D.; Arhim, M. and M. Arsyad: *Collaboration practice in supply chain of unshu citrus in Muchacha-En, Ltd, Japan.*
IOP Conf. Ser.: Earth Environ. Sci. **681**(1), 012073, 2021,
http://dx.doi.org/110.1088/1755-1315/681/1/012073,

[6] Calliera, M., et al.: *Multi-actor approach and engagement strategy to promote the adoption of best management practices and a sustainable use of pesticides for groundwater quality improvement in hilly vineyards.*
Science of The Total Environment **752**, 142251, 2021,
http://dx.doi.org/10.1016/j.scitotenv.2020.142251,

[7] CEMA - European Agricultural Machinery - *EU Code of Conduct on agricultural data sharing.*
https://www.cema-agri.org/index.php?option=com_content&view=article&id=37&catid=19&Itemid=216, accessed 13th January 2022,

[8] –: *NZ Farm Data Code of Practice*.
http://www.farmdatacode.org.nz, accessed 13th January 2022,

[9] Australian Farm Data Code: *National Farmers' Federation.*
https://nff.org.au/programs/australian-farm-data-code/, accessed 13th January 2022,

[10] –: Creating value from data.
https://hoards.com/article-27985-creating-value-from-data.html, accessed 13th January 2022,

[11] Baldin, M., et al.: *Integrated Decision Support Systems (IDSS) for Dairy Farming: A Discussion on How to Improve Their Sustained Adoption.*
Animals **11**(7), 2021,
http://dx.doi.org/10.3390/ani11072025,

[12] Vimalajeewa, D.; Kulatunga, C.; Berry, D. and Balasubramaniam, S.: *A Service-based Joint Model Used for Distributed Learning: Application for Smart Agriculture.*
IEEE Transactions on Emerging Topics in Computing, 1–1, 2021,
http://dx.doi.org/10.1109/TETC.2020.3048671,

[13] Lovarelli, D.; Bacenetti, J. and Guarino, M.: *A review on dairy cattle farming: Is precision livestock farming the compromise for an environmental, economic and social sustainable production?*
Journal of Cleaner Production **262**, 121409, 2020,
http://dx.doi.org/10.1016/j.jclepro.2020.121409,

[14] Cue, R., et al.: *Data Governance in the Dairy Industry*.
Animals (Basel) **11**(10), 2981, 2021,
http://dx.doi.org/10.3390/ani11102981,

[15] Willenbrink, E.; North, L.A.; Nguyet, V.T.M.; Polk, J. and Graham, J.: *Communication Networks as a Catalyst for Holistic Sustainability on Karst Landscapes.*
Sustainability **13**(6), 3360, 2021,
http://dx.doi.org/10.3390/su13063360,

[16] Yurui, L., et al.: *Towards the progress of ecological restoration and economic development in China's Loess Plateau and strategy for more sustainable development.*
Science of The Total Environment **756**, 143676, 2021,
http://dx.doi.org/10.1016/j.scitotenv.2020.143676.

[17] Salmoral, G. et al.: *Water-related challenges in nexus governance for sustainable development: Insights from the city of Arequipa, Peru.*
Science of The Total Environment **747**, 141114, 2020,
http://dx.doi.org/10.1016/j.scitotenv.2020.141114,

[18] Welle Donker, F. and van Loenen, B.: How to assess the success of the open data ecosystem?
International Journal of Digital Earth **10**(3), 284-306, 2017,
http://dx.doi.org/10.1080/17538947.2016.1224938,

[19] Runeson, P.; Olsson, T. and Linåker, J.: *Open Data Ecosystems — An empirical investigation into an emerging industry collaboration concept.*
Journal of Systems and Software **182**, 111088, 2021,
http://dx.doi.org/10.1016/j.jss.2021.111088,

[20] Šalamon, D., et al.: *Promoting digital innovation: Identifying the key elements in agricultural open data ecosystem in Croatia.*
56. hrvatski i 16. međunarodni simpozij agronoma - zbornik radova, 305, 2021,

[21] Learning for Sustainability: *Stakeholder mapping and analysis*
https://learningforsustainability.net/stakeholder-analysis/, accessed 19th April, 2022,

[22] Grimble, R. and Wellard, K.: *Stakeholder methodologies in natural resource management: a review of principles, contexts, experiences and opportunities.*
Agricultural Systems **55**(2), 173-193, 1997,
http://dx.doi.org/10.1016/S0308-521X(97)00006-1,

[23] Grimble, R. and Chan, M.-K.: *Stakeholder analysis for natural resource management in developing countries.*
Natural Resources Forum **19**, 113-124, 2009,
http://dx.doi.org/10.1111/j.1477-8947.1995.tb00599.x.

[24] Lelea, M.A.; Roba, G.M; Christinck, A. and Kaufmann, B.: *Methodologies for stakeholder analysis: for application in transdisciplinary research projects focusing on actors in food supply chains: reload reducing losses adding value.*
DITSL, Witzenhausen, 2014,
[25] Allen, W. and Kilvington, M.: *Stakeholder analysis*.
In: Frame, B.; Gordon, R. and Mortimer, C., eds.: *The Capacity for Sustainable Development.* Landcare Research (ManaakiWhenua), Lincoln, 2010,
[26] Hrustek, L.: *Key stakeholder groups for digital innovation of agriculture in Croatia*.
Proceedings of 56th Croatian and 16th International symposium on Agriculture, 198, 2021,
[27] Ecommerceceo: *10 Types of Ecommerce Business Models That Work In 2022.*
https://www.ecommerceceo.com/types-of-ecommerce-business-models, accessed 13th January 2022,
[28] Scnsoft: *Types of Ecommerce: 3 Alternative Classifications with Examples and Success Tips.*
https://www.scnsoft.com/ecommerce/types-of-ecommerce, accessed 13th January 2022,
[29] USAID Learning Lab: Collaboration Mapping,
https://usaidlearninglab.org/library/collaboration-mapping, accessed 19th April, 2022,
[30] Dudáš, M. and Eliaš, P.: *Alien weed Xanthium spinosum in Slovakia I: distribution and habitats.*
Journal of Central European Agriculture **22**(2), 305-316, 2021,
http://dx.doi.org/10.5513/JCEA01/22.2.3083,
[31] Balenović, L.; Marjanović, H. and Benko, M: *Application of Aerial Photographs in Forest Management in Croatia.*
Šumarski list **134**(11-12), 623-630, 2010,
[32] Čobanović, K.; Kučević, D.; Plavšić, M. and Bogdanović, V.: *Impact of non nutritional factors on milk urea concentration and its relationship with production and fertility traits in Vojvodina dairy herds.*
Mljekarstvo : časopis za unaprjeđenje proizvodnje i prerade mlijeka **67**(4), 267-276, 2017,
http://dx.doi.org/10.15567/mljekarstvo.2017.0404,
[33] Degmečić, D.: *Support model for breeding the red deer (cervus elaphus, l.) In the open hunting grounds.*
Poljoprivreda **15**(2), 66-67, 2009,
http://dx.doi.org/10.1111/j.1468-2311.1976.tb00316.x,
[34] Franić, R.: *Rural development policy - a new opportunity for Croatia.*
Agronomski glasnik : Glasilo Hrvatskog agronomskog društva **68**(3), 221-235, 2006,
[35] Žimbrek, M. *The position of small and large agricultural producers in the framework of the support of the Common Agricultural Policy from 2021 to 2027*. Master thesis.
University of Zagreb, Faculty of Agriculture. Department of Agricultural Economics and Rural Development, 2019,
https://urn.nsk.hr/urn:nbn:hr:204:727416, accessed 15th January, 2022,
[36] Družić, J., et al.: *Influence of cultivation system on fruit quality of cultivar Elsanta,*
Pomologia Croatica : Glasilo Hrvatskog agronomskog društva **12**(4), 255-262, 2006,
[37] Krajter Ostoić, S., et al.: *Training Programmes in Sustainable Forest Management in Austria, Croatia and Slovenia.*
South-east European forestry **8**(2), 137-146, 2017,
http://dx.doi.org/10.15177/seefor.17-13,
[38] Primyastanto, M.; Lestariadi, R.A. and Haris, A.K.: *Sustainable operational analysis of the cultivation of Indonesian thunnus albacares by bioeconomic approach.*
Ribarstvo **79**(2), 61-70, 2021,
http://dx.doi.org/10.2478/cjf-2021-0007,
[39] Berisha, V.: *Development of organic agriculture in the Republic of Croatia.* B.Sc. Thesis. In Croatian.
https://zir.nsk.hr/islandora/object/unipu:1066/preview, accessed 15th January, 2022,

[40] Juretić, N.: *Leaf and Flower Deformations of Datura stramonium L. Infected With the Cucumber Mosaic Virus.*
Acta Botanica Croatica **26-27**(1), 117-144, 1967,

[41] Jakšić, L.: *Strategic planning in development programs of Croatian viticulture and winemaking.* Ph.D. Thesis. In Croatian.
University of Zagreb. Faculty of Agriculture, 2018,
https://urn.nsk.hr/urn:nbn:hr:204:104357, accessed 15th January, 2022,

[42] Danilović, M.; Stojnić, D.; Vasiljević, V. and Gačić, D.: *Biomass from Short Rotation Energy Plantations of Black Locust on Tailing Dump of "Field B" Open Pit in "Kolubara" Mining Basin.*
Nova mehanizacija šumarstva : Časopis za teoriju i praksu šumarskoga inženjerstva **34**(1), 11-19, 2013,

[43] Svirac, M.: *We Have Built the Out-Door Oven First Baker's or Out-Door Oven in the Western Slavonia.*
Studia Ethnologica Croatica **12/13**(1), 111-124, 2001,

[44] Croatian Dairy Association: *HMU.*
https://hmu.hr, accessed 15th January, 2022,

[45] Croatian Plant Protection Society: *HDBZ.*
https://hdbz.hr, accessed 15th January, 2022,

[46] Dimšić, J.: *The impact of technological innovation on agriculture and rural development.* M.Sc. Thesis. In Croatian.
University of Zagreb, Faculty of Science, Department of Geography, Zagreb, 2020,
https://urn.nsk.hr/urn:nbn:hr:217:848679, accessed 15th January, 2022,

[47] –: *Agriculturae Conspectus Scientificus.*
https://acs.agr.hr/acs/index.php/acs, accessed 15th January, 2022,

[48] –: *Agronomski glasnik HAD.*
http://www.agronomski-glasnik.agronomsko.hr, accessed 15th January, 2022,

[49] –: *Agroeconomia Croatica.*
http://haed.hr/aec, accessed 15th January, 2022,

[50] Tvrdeić, A. and Kočevski, D.: *Does behavioral response to novelty influence paw withdrawal latencies in repeated Hargreaves test?*
Periodicum biologorum **110**(1), 95-100, 2008,

[51] Varga, F., et al.: *Dalmatian pyrethrum spatial records quality from Croatian Open Databases.*
56th Croatian & 16th International Symposium on Agriculture Book of Abstracts, 134, 2021,

[52] Odeljan, I.: *International cooperation and institutional support for agricultural advisory services.* M.Sc. Thesis. In Croatian.
University of Zagreb. Faculty of Agriculture. Department of Agricultural Economics and Rural Development, 2016,
https://urn.nsk.hr/urn:nbn:hr:204:940900, accessed 15th January, 2022,

[53] Kovačićek, T: *Development and testing of models for evaluation of the process of creating agricultural policy in the segment of goal setting and selection of measures.*
https://www.bib.irb.hr/954972, accessed 15th January, 2022,

[54] Tomić, D.: *Socially responsible business in hospitality.* Master Thesis. In Croatian.
Sveučilište Jurja Dobrile u Puli, Fakultet ekonomije i turizma "Dr. Mijo Mirković", Pula, 2019,

[55] Miličević, T.; Ivić, D.; Cvjetković, B and Duralija, B.: *Possibilities of Strawberry Integrated Disease Management in Different Cultivation System.*
Agriculturae Conspectus Scientificus **71**(4), 129-134, 2006,

[56] Bonacci, O. and Ljubenkov, I.: *Different air temperature values and trends at two stations on a small island: the case of meteorological stations Korčula and Vela Luka on Korčula Island.*
Hrvatske vode **28**(113), 183-196, 2020,
http://dx.doi.org/10.2478/achi-2020-0013,

[57] Gašparović, M.; Simic Milas, A.; Seletković, A. and Balenović, I.: *A novel automated method for the improvement of photogrammetric DTM accuracy in forests.*
Šumarski list **142**(11-12), 567-576, 2018,
http://dx.doi.org/10.31298/sl.142.11-12.1.

[58] –: *Časopis Veterinarska stanica*.
http://www.veinst.hr/casopis-veterinarska-stanica, accessed 15th January, 2022,

[59] –: *Krmiva - HAD*.
http://www.krmiva.agronomsko.hr, accessed 15th January, 2022,

[60] Jerković, D.: *Increasing the competitiveness and interdependence of agriculture and tourism by using learning organizations.* Ph.D. Thesis.
Tourism and hospitality management **25**(1), 232-237, 2019,

[61] Brajdić, I.: *An analysis of decision-making performances in Croatian hotel enterprises from 1994 to 1998,*
Tourism and hospitality management **5**(1-2), 161-184, 1999,
http://dx.doi.org/10.20867/thm.5.1-2.12,

[62] Perković, Z.: *Enology and gastronomy of Baranja as a resource for the development of enogastronomic experiences.* M.Sc. Thesis. In Croatian.
Virovitica College, 2020,
https://urn.nsk.hr/urn:nbn:hr:165:573102, , accessed 15th January, 2022,

[63] Harding, K.: *Global Health Innovation Technology Models.*
Nanobiomedicine **3**, 3-7, 2016,
http://dx.doi.org/10.5772/62921,

[64] Zergollern Čupak, L.; Barišić, I; and Morožin Pohovski, L.: *Development of Genetics in the World and in Croatia – Forty Years of the Croatian Society of Human Genetics of the Croatian Medical Association.*
Collegium antropologicum **38**(3), 809-818, 2014,

[65] Tolušić, Z.: *Models of organization and management of hunting in the Republic of Croatia.* Ph.D. Thesis. In Croatian.
Josip Juraj Strossmayer University of Osijek, Faculty of Economics, 2019,
https://urn.nsk.hr/urn:nbn:hr:145:960439

[66] Perušić, M.: *Historical Architecture in the Botanical Garden of the Faculty of Science in Zagreb.*
Portal : godišnjak Hrvatskoga restauratorskog zavoda **9**, 93-112, 2018,
http://dx.doi.org/10.17018/portal.2018.7,

[67] Animasaun, D.A.; Oyedeji, S.; Onyegwukwu, F. and Azeez, M.A.: *Evaluation of growth and pollen viability in relation to fruit set among five varieties of tomato grown in Nigeria.*
Agronomski glasnik : Glasilo Hrvatskog agronomskog društva **76**(4-5), 203-218, 2014,

[68] Yamoah, F.A.; Kaba, J.S.; Amankwah-Amoah, J. and Acquaye, A.: *Stakeholder Collaboration in Climate-Smart Agricultural Production Innovations: Insights from the Cocoa Industry in Ghana.*
Environmental Management **66**(4), 600-613, 2020,
http://dx.doi.org/10.1007/s00267-020-01327-z,

[69] Daher, B.; Hannibal, B.; Portney, K. and Mohtar, H.: *Towards Creating an Environment of Cooperation between Water, Energy, and Food Stakeholders in San Antonio.*
Science of The Total Environment **651**(2), 2913-2926, 2019,
http://dx.doi.org/10.1016/j.scitotenv.2018.09.395,

[70] Agyekumhene, C., et al.: *Making Smallholder Value Chain Partnerships Inclusive: Exploring Digital Farm Monitoring through Farmer Friendly Smartphone Platforms*.
Sustainability **12**(11), 4580, 2020,
http://dx.doi.org/10.3390/su12114580,

[71] Zoll, F.; Diehl, K. and Siebert, R.: *Integrating Sustainability Goals in Innovation Processes: Applying a Decision Support Tool in a Dual-Purpose Chicken Case Study.* Sustainability **11**(14), No. 3761, 2019, http://dx.doi.org/10.3390/su11143761,
[72] Ricciardi, V., et al.: *How much of the world's food do smallholders produce?* Global Food Security **17**, 64-72, 2018, http://dx.doi.org/10.1016/j.gfs.2018.05.002,
[73] Gleeson, T.; Wada, Y; Bierkens, M.F.P and van Beek, L.P.H.: *Water balance of global aquifers revealed by groundwater footprint.* Nature **488**(7410), 2012, http://dx.doi.org/10.1038/nature11295,
[74] Hrustek, L.: *Sustainability Driven by Agriculture through Digital Transformation.* Sustainability **12**(20), 2020, http://dx.doi.org/10.3390/su12208596.

Interdisciplinary Description of Complex Systems 20(2), 210-221, 2022

OPEN ACCESS ON GNSS PERMANENT NETWORKS DATA IN CASE OF DISASTER*

Andrea Latinčić[1], Željko Bačić[2] and Zvonimir Nevistić[2, **]

[1]Granica d.o.o.
Kaštel Novi, Croatia

[2]University of Zagreb, Faculty of Geodesy
Zagreb, Croatia

DOI: 10.7906/indecs.20.2.11
Regular article

Received: 14 January 2022.
Accepted: 3 March 2022.

ABSTRACT

Earthquakes, as a natural phenomenon causing large physical and social destruction, are the subject of intensive research throughout the world. Spurred by the fact that in year 2020, two catastrophic earthquakes hit Croatia, in March with epicenter near Zagreb and December with epicenter near Petrinja, at the Faculty of Geodesy, University of Zagreb activities were initiated with the aim of strengthening the ability to react in these situations. Focus of those activities is on providing fast, adequate, and complete information on the disaster in the field of geodesy and geoinformatics. The research was focused on interpretation of kinematics of surface motion during the earthquake itself for what high rate permanent GNSS (Global Navigation Satellite System) network stations registrations are necessary. The Croatian earthquakes experience as well as the Mexico (June 2020) and Samosa earthquake (October 2020), pointed out, related to the use of high-rate registration GNSS data, that the primary problem in the use of this data is open access to the data itself. That is why this study has been launched - to gain a global picture of the availability of data from permanent GNSS networks around the world. The research included the collection and processing of information on open access policies for permanent GNSS networks data in the event of natural disasters with an emphasis on earthquakes. A global survey of institutions around the world responsible for managing GNSS permanent networks has been conducted. The survey contains three groups of questions that include general information on the type of permanent networks, models of access to network data and the readiness of countries to reach an international agreement on the opening data of the GNSS network in the event of a disaster. The results indicated that a high percentage of countries participating in the survey were ready to agree to open the data and introduce a common international portal through which scientists and researchers would be able to download GNSS permanent network data free of charge in the event of natural disasters.

KEY WORDS

earthquakes, global survey, GNSS permanent networks, open data

CLASSIFICATION

JEL: L86
PACS: 84.40.Ua, 91.30.Px

*This is the extended version of the abstract published in: Vujić, M. and Šalamon, D., eds.: Book of abstracts of the National Open Data Conference. University of Zagreb, Faculty of Traffic and Transport Sciences, Zagreb, 2021.

**Corresponding author, η: znevistic@geof.hr, + 385 1 4639 538
University of Zagreb, Faculty of Geodesy, Kačićeva 26, HR - 10 000 Zagreb, Croatia

INTRODUCTION

Natural or man-made disasters occur every day around the world, but only a few of them are destructive, resulting in large-scale material and social damage. These can include earthquakes (e.g. Lisbon 1775, Chile 1960, Sumatra 2004), floods (e.g. Gunja 2014, Belgium, Germany and the Netherlands 2021), fires (e.g. seasonal events in California, Australia 2020, Greece and Turkey 2021 or the Adriatic coast), volcanic eruptions (Vatnajokull whose eruption caused air traffic congestion in the world) and man-made disasters (e.g. silo explosion in Beirut, spill of the Exxon Valdez oil platform in the Gulf of Mexico, etc.). During 2020, Croatia was hit by two such catastrophes, earthquakes in Zagreb and surrounding (March 22, 2020, M 5,5.) and Sisak-Moslavina County (December 28 and 29, 2020, M 5,2 and M 6,4). A last earthquake of this magnitude and consequences hit Zagreb 140 years ago (November 9, 1880), and the last earthquake of similar strength in the area of Pokuplje and Banija occurred on October 8, 1909.

Catastrophic earthquakes, no matter how much material damage and human casualties they cause, are at the same time an opportunity to gain new knowledge about the seismicity and behavior of the earth’s crust during and after earthquakes. An example of this is the discovery of Andrija Mohorovičić, which was triggered by earthquakes in Zagreb (1880) and in the Kupa Valley (1909), which revealed the surface of the discontinuity of velocities separating the crust from the Earth’s mantle, which was called Mohorovičić`s discontinuity. The discovery of the Moho layer is the largest scientific discovery in history ever made by a scientist working in Croatia [1].

In the situation of a catastrophic earthquake, civil protection services and professions close to the issue (construction, geodesy, geophysics, etc.) give their contribution in order to reduce the consequences of the earthquake, ie to provide information about what happened, why and how. Among others, the geodetic profession contributes to rescue services, earthquake remediation and mitigation in many ways such as interpreting crustal deformations, crust dynamics during and after earthquakes, by creating various cartographic substrates and applications. An example of this are geodetic measurements conducted after a devastating earthquake in the Sisak-Moslavina County, which found that Petrinja and Sisak moved up to 86 cm, and in the area of Glina there was a height shift, i.e. lowering the soil by 10 cm [2]. After the earthquake, the Faculty of Geodesy in Zagreb, the Croatian OpenStreetMap (OSM) community and Open IT d.o.o. in cooperation with the Croatian Mountain Rescue Service, the Civil Protection Operations Center and the State Geodetic Administration (SGA) created and maintain an interactive digital map Earthquake 2020 (https://potres2020.openit.hr/views/map) to help victims in earthquake-affected areas. Also, shortly after the earthquake, the Faculty of Geodesy made, with the help of registered experts, a new Digital Orthophoto for the area of the city of Petrinja [3].

After the earthquake in Zagreb (2020), geodetic experts conducted a study in which the analysis of 24-hour Global Navigation Satellites Systems (GNSS) measurements of reference GNSS points of the Croatian Positioning System (CROPOS) was conducted. GNSS observations provide users with autonomous geospatial positionings globally, based on satellite signals from Earth’s orbit. Based on the transmitted satellite signals, the receivers on the ground accurately determine their location. Today, there are several operational GNSS systems, the most famous of which is the GPS (Global Positioning System). Conducted study from GNSS observation found compression in the Earth’s crust two days before the earthquake in Zagreb [4]. Such compressions have also occurred (determined by analysis of GNSS measurements between GNSS points of position systems) in four previous earthquakes around Kraljevo, Drežnice, Skopje and Zagreb. These results indicate the potential of GNSS permanent networks data such as CROPOS in seismology and the possibility, along with other geophysical measurement data and methods, to determine the seismic activities occurred in an earthquake.

At the global level, various research of GNSS methods were also conducted for the purpose of better understanding of seismic activities, i.e. earthquakes. Japan, the country located in the most tectonic active area, has developed the national GNSS network GEONET as a basis for the development of methods and other systems for disaster mitigation such as earthquakes [5]. An additional example is the earthquake in the state of Haiti (January 12, 2010) which took many lives and caused great material damage. After the earthquake, OSM volunteers created a detailed map of Haiti to support the operations of rescue services [6]. OSM has great potential in supporting humanitarian mapping and has provided essential information in this but also other major disasters.

Recent experiences with the use of GNSS technology in earthquakes in Zagreb and Petrinja have prompted a study to strengthen the ability to respond to such situations and to provide fast and reliable information on disasters in the field of geodesy and geoinformatics. Preparedness, management, and response, but also an understanding of events such as an earthquake, is unthinkable today without GNSS data and permanent GNSS networks. One of the main obstacles of using this data is the availability of open GNSS data. Most GNSS permanent networks, whether private or public, require registration or payment for services use. All the above makes it difficult to download and use data for the purpose of interpreting an earthquake and mitigating its consequences.

For this reason, as part of this thesis, a global survey of institutions responsible for GNSS networks was launched to gain insight into their disaster data access policies. The results of the survey will show whether countries use special models of access to data in the event of disasters and the readiness of countries for international cooperation, standardization, and data exchange of GNSS networks at the global level in the event of large-scale disasters.

Even though GNSS were created primarily for military purposes, over time, their benefits have been recognized in many other areas. With the development of technology and sensors and the cooperation and interoperability of global and regional navigation satellite systems, their application is expanding, and the characteristics and capabilities of these systems are being improved. To further increase the quality of products and data, GNSS permanent networks have been established. Today, GNSS data are the basis of many scientific and engineering studies both in the field of geodesy and in many other professions.

GNSS PERMANENT NETWORKS

GNSS permanent networks are established with the aim of improving and providing more precise and reliable positioning and navigation. They consist of GNSS continuously operating satellite monitoring stations, data processing and analysis centers and major systems for management and monitoring of the entire system. The places where stations are placed (GNSS receivers and antennas for permanent satellite observation) must be geologically stable areas to be protected from damage and displacement of the points.

GNSS Permanent networks can be global (International GNSS Service – IGS), continental (EPN, CORS, ARGN) or national (e.g., Japan, New Zealand, Italy, Germany, Austria, etc.). Also, the Republic of Croatia has established a National Network of GNSS reference stations called the Croatian Positioning System (CROPOS) [7].

IGS consists of more than 400 globally deployed satellite monitoring stations, data processing and analysis centers. In this way, they provide highly reliable GNSS data in near real time to meet the needs of diverse scientific and engineering research. IGS is an association of more than 200 agencies, universities, and research institutions from more than 100 countries whose cooperation contributes to defining the most accurate orbits of GPS satellites in the world, which forms the basis for further application in many areas [8].

The International Association of Geodesy's subcommittee on the Regional Reference Framework for Europe (EUREF) has developed a GNSS network called the European Permanent Network, made up of continuously operating GNSS stations on the European continent. The EUREF permanent network is an association of more than 100 self-funded agencies, universities, and research institutions from more than 30 countries in Europe. Under the auspices of EUREF, ETR89 coordinates and GNSS observation data from over 200 permanent GNSS observation stations across the European continent have been made public [9].

Geoscience Australia is a government organization in Australia that manages and maintains a network of approximately 100 continuously operating reference stations in the Australian region. The network includes the Australian Regional GNSS Network (ARGN), the AuScope Network, and the South Pacific Regional GNSS Network. The primary role of ARGN is to define a geodetic framework for the spatial data infrastructure of Australian territories. ARGN consists of a network of permanent GNSS receivers and antennas placed in geologically stable places in Australia. The AuScope Association in Australia provides research infrastructure to the geospatial and terrestrial scientific community. Their tools, data, services and analyses enable scientists to understand Earth's evolution over time and how Earth's resources can meet growing human needs [10].

The National Geodetic Survey (NGS), as part of the National Ocean Service, operates the Continuously Operating Reference Stations (CORS). The CORS network provides quality GNSS data containing measurements of carrier phases and code distances to support 3D positioning, meteorology, space weather, and geophysical applications for the United States, and several other countries. It operates based on cooperation between organizations that independently manage stations and share data with NGS, while NGS analyses and shares data free of charge. The CORS network consists of approximately 2000 stations owned by over 200 different organizations.

The Croatian Positioning System (CROPOS) is the National Network of Reference Stations of the Republic of Croatia consisting of 33 reference GNSS stations at 70 km from each other. They are arranged so that they cover the entire territory of the Republic of Croatia to collect satellite measurement data and calculate the correction parameters for positioning. The system is available 24/7, and observation data is exchanged with the permanent networks of neighboring countries in real time. Observation data and correction parameters are available to users in real time and can also be used for post-processing. CROPOS allows real-time positioning with an accuracy of 2 cm horizontally and 4 cm vertically. In 2011, the State Geodetic Administration made a proposal to include five CROPOS stations in the European Permanent Network. After controls and analysis of observation data, on 16 June 2013 the stations CAKO (Čakovec), DUB2 (Dubrovnik), PORE (Poreč), POZE (Požega) and ZADA (Zadar) were included in the EUREF [11].

GNSS permanent networks applications are widespread in many areas such as monitoring the rotation of the Earth, determining the parameters of reference systems, monitoring climate change and their impact on our planet, etc. Increasingly, data recorded by GNSS permanent network stations are being used to support seismic tectonics in tracking the deformations of the

Earth's crust after earthquakes. Various studies have shown that GNSS stations recorded terrain compressions before the earthquake, indicating the possibility of an earthquake to occur. Also, satellite methods can determine the horizontal and vertical displacements of stations before, during and after the tremor at great distances from the epicenter of the earthquake. This indicates that there is great potential for the application of GNSS permanent network data in earthquake impact assessments on natural and man-made structures on earth [7].

Nevertheless, it is known fact that permanent GNSS networks and services are providing GNSS stations data with registration rate of 15 seconds making them available via IGS or national organizations. This is usually not the case with high-rate data (1 second registration rate for example) which are necessary for kinematic interpretation of earthquake surface displacement behaviour during the earthquake itself. Beside this, in case of catastrophic earthquakes rescue and recovery services and society in general request such information promptly, meaning that access to those data is enabled immediately and data made open by the organization managing GNSS permanent network in affected area.

OPEN GNSS DATA

Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike. Open data is characterized by availability and accessibility, reuse and redistribution, and global inclusion. The data should be available in full extent, at a price not exceeding the amount of publication, which is why their sharing and downloading is recommended via the Internet. As far as global inclusion is concerned, it implies the right of everyone to use and redistribute data without discrimination between areas of use or people or communities [12].

There are several different initiatives, both private and government, for the purpose of standardization and openness of government data. Great efforts are being made to achieve this goal through organizations such as the Open-Source Geospatial Foundation (OSGeo), the Infrastructure for Spatial Information in Europe (INSPIRE Directive), e-government and others. Many countries have recognized the potential of open data, and one of them is the Republic of Croatia, which established the Open Data Portal in March 2018. The portal is a data node for the collection, categorization, and distribution of open public sector data. It is a unique place to access open data for reuse by all citizens. Through the portal it is possible to access many data sets such as spatial data, traffic, meteorological data, environmental data, and others [13].

Today, defining the term open data is extremely important because data openness allows interoperability. Interoperability is the joint work of different sciences, organizations and systems and the merging of different data sets. The ability to connect various components is the foundation of building the large, modern, and complex systems of today. Open data is an important factor in many research and increasingly essential for many professions, including geodesy. One example is the data of GNSS permanent networks, whose applications in geodesy are multiple, and in order to achieve their efficient use, it is important that access to this data is open.

The importance of GNSS network data is shown by the fact that reliable and precise navigation and positioning is one of the basic items of economic development, both at the European and global level. For this reason, wide availability of GNSS services is required to all users. In order to be accessible, the data must be open, which means that access to data should be simplified, free and user-friendly. Open access to data is provided on IGS and EPN web portals that distribute the observation data of GNSS reference stations of networks of different countries. The data is usually given in RINEX (Receiver Independent Exchange Format) format, which is commonly used data interchange format for raw satellite navigation system data. Data is provided with a registration rate of 15 seconds. For most of the needs such as positioning and navigation, they are of satisfactory quality, as well as for calculating the permanent displacement of the earth's crust in the event of catastrophic earthquakes.

The problem arises in case of need of more precise and reliable data of high registration rate. Access to such data is limited and requires the registration of users, and thus the payment of fees for the data use. High registration rate data are crucial for research on natural processes on Earth such as earthquakes. Earthquakes due to their destructive nature require special attention and detailed research of activities related to them is required. Surveyors who are using high quality GNSS data and high registration rates play a significant role in assisting seismologists in studying such catastrophic events. Through the already mentioned portals, it is possible to access high-quality data, but they have a lower registration rate than required for seismic research and contain observations only of national GNSS stations that do not create a sufficiently dense network as needed for seismology.

RESEARCH OF OPEN DATA POLICIES OF GNSS NETWORK ADMINISTRATORS

RESEARCH GOALS

Encouraged by the fact that earthquakes are one of the most destructive natural phenomena and the need to improve preparedness for such situations as well as to mitigate the consequences, research was launched on the openness and policies of access to GNSS permanent network data in the event of a catastrophe. An example of the catastrophic consequences that earthquakes leave behind are the events in Croatia in the cities of Zagreb and Petrinja. This events additionally stimulated the research on the possibilities of providing more complete and better information in the field of geodesy and geoinformatics to achieve a better and stronger response to such situations.

The availability of reliable high registration rate data of GNSS networks is essential for the observation and understanding of tectonic and seismological activities. Therefore, research has been developed with the aim of gaining a global picture of the availability of data from permanent GNSS networks. A global survey was created and then sent to the institutions responsible for managing GNSS networks to find out about their policies regarding access to data in the event of a disaster or earthquake.

GLOBAL SURVEY

As part of the research, a global survey was conducted, which included the institutions responsible for the managing of GNSS permanent networks in individual countries in Europe and the world. Data about networks such as number of GNSS stations, network name and contact were collected through the official websites of the institutions that monitor and maintain the GNSS permanent networks of individual countries or regions.

The global survey was created using Google Forms. The survey was divided into three groups of questions. The first group of questions refers to basic information about the network (name, coverage area, purpose, etc.) and the institution that manages the network. The second group of questions refers to data access to networks, i.e. which data access models are used by individual countries and whether they apply a different model of data access in the event of natural disasters such as earthquakes. The last, third group of questions is related to the topic of standardization of open access to data for scientists and researchers in case of disasters.

SURVEY RESULTS

The survey was sent by e-mail to total of 44 institutions for the management GNSS permanent networks. The feedback was received from 16 institutions in total, which is 39 %, 15 of them come from Europe, what makes this survey a representative sample for research in Europe, while the remaining 2 responses were received from the rest of the world (Mexico and Sri Lanka). In Table 1 list of the countries and names of the GNSS permanent networks for which responses have been received are shown. Since the presentation of the answers to all survey questions would be too extensive, representative answers of each group of questions were selected and explained. In addition, they are graphically described, to systematize and more clearly present the results of the survey.

Table 1. List of countries and GNSS Network covered by this research.

Country	GNSS Permanent Network
Slovakia	SKPOS - Slovak real-time positioning service
Latvia	Latvian permanent global positioning base station network
Portugal	ReNEP
Poland	ASG-EUPOS
Italy	Rete Inteigrata Nazionale GNSS (RING)
Serbia	AGROS- Active Geodetic Reference Network of Serbia
Mexico	TLALOCNet
Sri Lanka	CORSnet
Switzerland	AGNES
Great Britain	OS Net
Ireland and North Ireland	GNSS Network of Ireland
Spain	ERGNSS
Bosnia and Herzegovina	FBiHPOS
Slovenia	SIGNAL
Hungary	Hungarian GNSS Service (GNSSnet.hu)
Estonia	ESTPOS

The first group of questions provided information on the purpose and character of GNSS permanent networks. The left graph in Figure 1 shows that most countries use GNSS networks for positioning purposes. A slightly smaller number use networks for geodynamic purposes, then as support for transport systems, seismic surveys, but also for combined purposes. It is clearly shown in the right graph on the Figure 1 that in all countries, except for Sri Lanka, the GNSS network is managed by state or public bodies.

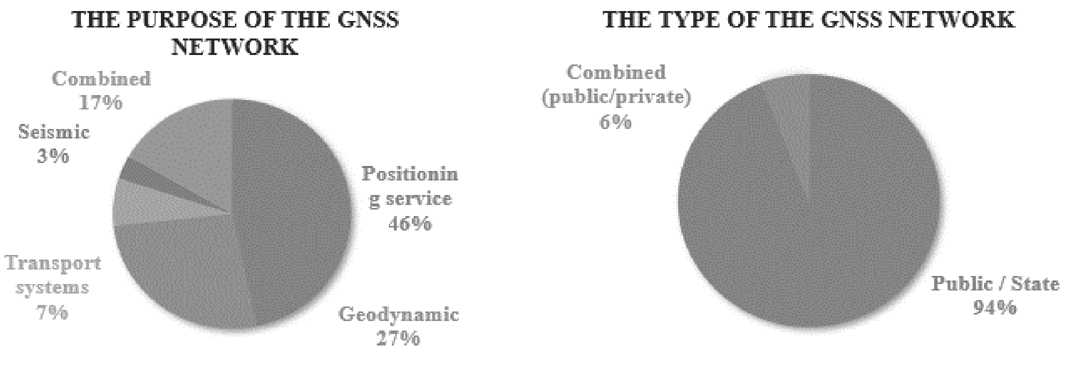


Figure 1. Answers to the first group of questions about purpose and type of GNSS networks.

The survey also sought to gain an idea of the model and method of accessing GNSS network data. The results indicate that the combined access model is more prevalent, which means that most users are charged for access to data, while the use of data is free for certain groups of users as it's shown on the left graph in Figure 2. Almost the same number, as shown on right graph in Figure 2, of institutions provide access to data exclusively through the official portal of network institutions and those that additionally provide access to data through another information service.

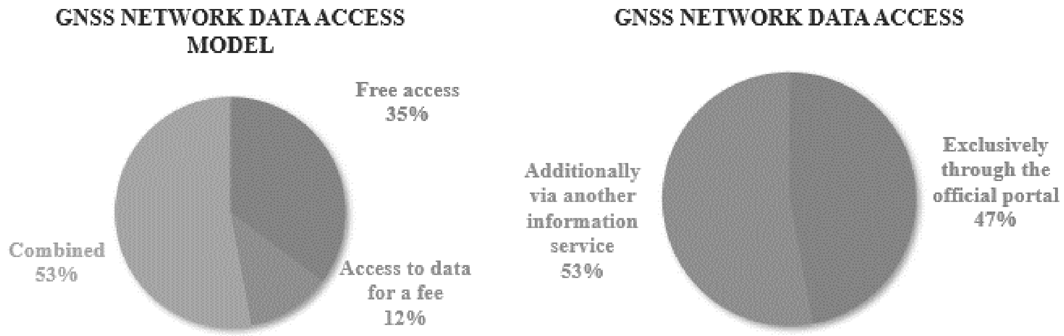


Figure 2. Answers about the model and how to access the data.

The last group of questions sought to achieve the main goal of the research, which is to gain insight into the open data policies of the institutions responsible for managing GNSS permanent networks in the event of a disaster. When it comes to sharing data after disaster with scientists and researchers, all institutions are willing to make their data available in such situations as shown on left graph in Figure 3. Some countries did not give a completely positive answer but stated that their decision to provide data to scientists depends on the type and amount of data. In survey, an international agreement on open data of GNSS networks for scientists and researchers in the event of disasters has been proposed. Right graph in Figure 3 shows that most institutions (71 %) would be ready to reach an international agreement. Hungary was the only one to give a completely negative answer, and this is described in more detail in the next chapter. Other institutions (23 %) stated that consent to the agreement depends on the type of agreement itself, registration and confirmation, i.e. the decision of the government (Serbia).

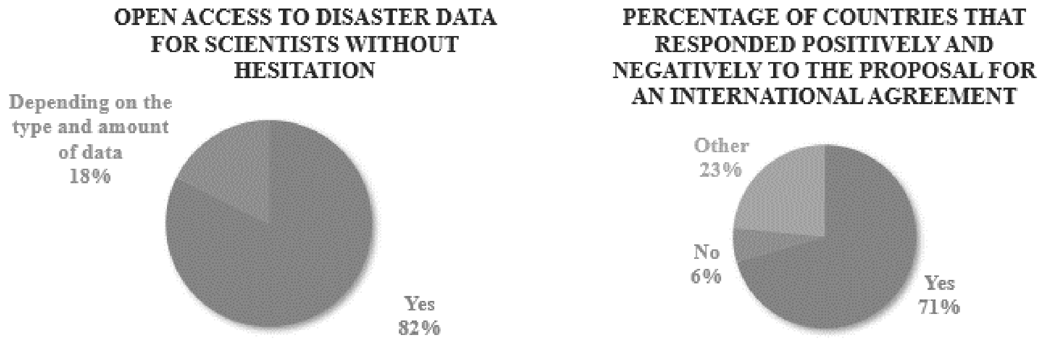
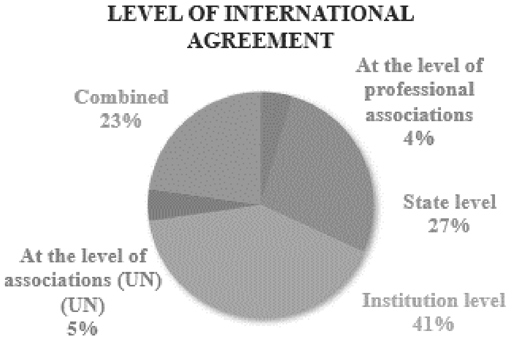


Figure 3. Answers to questions about data access in case of disasters and establishing an international agreement

In the survey was also examined which level of international agreement would be the most acceptable for individual countries. Left graph in Figure 4 shows that most countries would agree to an agreement at the level of institutions responsible for managing GNSS permanent networks. The readiness of countries to publish CORS station data for a certain period before and after the disaster through a common international portal was examined in the last question of the survey. Right graph in Figure 4 shows that a high percentage of countries are willing to agree to this type of agreement (70 %). The institutions which state that their decision depends on the registration, the amount of data or the content of the mentioned agreement, are classified under other answers (18 %).



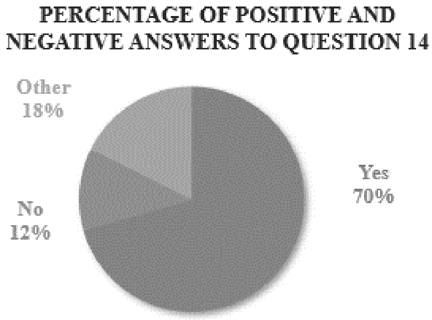


Figure 4. Graphic representation of the answers to the last questions of the survey.

DISCUSSION

This research provided insight on policies for access to GNSS data on permanent networks in the event of a disaster in individual countries that participated in the survey. The survey and web research provided reliable data on the institutions that manage individual GNSS networks, the type and purpose of networks and portals through which network data can be accessed. Research has shown that the networks are mostly public and are mostly managed by public, state institutions. In all countries covered by the survey purpose of GNSS permanent networks is to provide data for positioning services. Many of them also use networks for geodynamic purposes, and some to support transport systems, agriculture, and seismology.

The survey also includes official contacts for data users through which it is possible to contact the persons responsible for data management, as well as the types of network data access models describing whether the data usage service is charged or free. Only two countries (Hungary and Bosnia and Herzegovina) provide data only to registered users and their services of using data is possible only with payment of a fee. Some countries provide data to all registered users free of charge. An equal number of countries have stated that they provide data to registered users for a fee and allow certain groups of users to use the data free of charge.

Following the previous response, a request was made to list groups of users who are exceptionally allowed to use GNSS data for free. It is noticeable in the answers that these are most often state, public institutions and organizations and universities that use the data for research and scientific purposes. The question of whether access to data is available only through the official portals of institutions that manage networks or also through other information services is almost the same percentage of both answers. Some countries provided direct Internet addresses through which the mentioned services can be accessed, while others explained that the availability of the portal depends on the type of user because they are most often managed privately.

Almost all institutions answered that they do not apply a special model of access to data in case of disasters, except for Hungary and Bosnia and Herzegovina. Hungary shares data in case of natural disasters free of charge for research purposes with scientists from their country, and Bosnia and Herzegovina provide data exclusively to registered users.

Further in the research, GNSS network managers were asked about their GNSS data access policies in the event of large-scale disasters and their willingness to introduce an international agreement, in a form of standard on open access to GNSS network data. The first question in this section received mainly affirmative answers in which the institutions stated that they were ready to share the data of the GNSS networks, which they manage, with scientists and researchers free of charge. They also stated that they would agree to the establishment of an international agreement for scientists and researchers that would be related to the availability and distribution of GNSS network data. Latvia stated that it would agree to such agreement if there is a registration of users, Serbia if the agreement would be approved by their government, and Estonia's consent depends on the content and type of agreement.

When asked which level of international agreement would be the most acceptable, 9 countries would agree to an agreement at the level of agencies, institutions, or companies operating networks, 4 institutions replied at the state level, Latvia would agree to an agreement at the level of the United Nations and Switzerland would agree to an agreement at the level of professional associations (geophysical, geodetic, etc.). The last question in the research was about readiness of making the data of their CORS stations for a certain period before and after the disaster available through a common international portal for researchers and scientists in the event of large-scale disasters. Most countries agreed to the proposal, In Serbia and Estonia it depends on the amount of data and the agreement from previous question, and Bosnia and Herzegovina did not give a clear answer. The Hungarian institution answered the last group of questions by e-mail, and they explained that they are not ready for an international agreement in the field of open data of GNSS permanent networks. However, in the event of natural disasters, they are willing to share GNSS RINEX data free of charge for research purposes only.

CONCLUSION

The focus of the research is on defining the global picture of the availability of GNSS permanent network data, with an emphasis on the openness of high-frequency network data. This type of data is crucial for research on seismological activities in earthquakes events that have an impact on the economic and social aspect of human lives. It was found that the data, which are publicly and free of charge available through Internet portals, are not sufficient for the study of seismological activities. Limited access to high-quality data of GNSS networks prompted this research and interest in creating arrangements to open access to data in the event of large-scale disasters for researchers and scientists. More accessible data would mean further progress in understanding, preparedness and responding to situations such as devastating earthquakes.

To achieve our goal, a global survey was sent to the institutions responsible for managing GNSS permanent networks around the world. Given that out of 44 countries to which the survey was sent, 17 countries responded, and 15 of them are in Europe, we can draw conclusions more at the European level than at the global level, which was originally the goal.

The first two groups of questions resulted in general information about GNSS networks and the data access model. Of greatest interest are the data collected in the last group of questions within the survey and relating to standardization of open access to data for scientists and researchers in case of disasters. All countries would agree to open GNSS network data in the event of a disaster for scientists and researchers without delay and payment. If an international agreement were proposed in the field of open data on GNSS networks for scientists in the event of natural disasters, the institutions would mostly be ready for this type of cooperation and agreement. Finally, it was investigated whether countries are willing in the event of a disaster to allow data to be uploaded through a common international portal for scientists and researchers. Mostly positive responses were received, except for some countries that are either not in favour of such a form of cooperation at all, or it depends on the already mentioned agreement and the type of data it would cover.

Research results provides a basis, at European level, for a possible agreement to open high-frequency GNSS permanent network data for research purposes. Data availability is a key factor for all types of successful research as it greatly facilitates and speeds up the whole process. Access to data in case of natural disasters, such as earthquakes, for geodetic experts, but also other scientists, would mean further progress in understanding seismic activities. The development of an international portal through which scientists could download high-quality data would simplify research that might provide support in the future for better understanding of major earthquakes. The primary goal of the research has been met, a foundation has been created for future agreements with the countries that participated in the research, but also for all others that recognize the potential of open data of GNSS permanent networks.

ACKNOWLEDGEMENT

This research is part of TODO project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857592.

REFERENCES

[1] Državni hidrometeorološki zavod: *Andrija Mohorovičić – the most famous Croatian geophysicist.* In Croatian.
https://meteo.hr/objave_najave_natjecaji.php?section=onn&param=objave&el=zanimljivosti&daj=zn22012021, accessed 22st January 2021,

[2] Ros Kozarić, M.: *Earthquakes in Banovina.* In Croatian.
Geografski Horizont **LXVI**(2), 9-22, 2020,

[3] Bačić, Ž.; Šugar, D. and Nevistić, Z.: *Quick use of satellite and GI technologies after earthquake - Petrinja 29.12.2020,*
8th International Conference Contemporary Achievements in Civil Engineering, Subotica, 2021, https://www.bib.irb.hr/1126359, accessed 20th December 2021,

[4] Solarić, N., and Solarić, M.: *Compression of the Earth's crust two days before the earthquake on March 22, 2020 in Zagreb.* In Croatian.
Geodetski list **75(98)**(1), 1-8, 2021,

[5] Kato, T., et al.: *A new tsunami monitoring system using RTK-GPS.*
International Tsunami Symposium, Seattle, 2001,
https://www.researchgate.net/publication/228768016_A_new_tsunami_monitoring_system_using_RTK-GPS, accessed 12th December 2021,

[6] Frančula, N.: *Application of OpenStreetMap in humanitarian mapping.* In Croatian.
Geodetski list **74(97)**(3), 352-353, 2020,

[7] Šugar, D. and Bačić, Ž.: *Kinematic effects of M5.5 Zagreb earthquake assessed by GNSS method supported by Galileo Satellite System.*
1st Croatian Conference on Earthquake Engineering, Zagreb, 2021,
http://dx.doi.org/10.5592/CO/1CroCEE.2021.74,

[8] International GNSS Service (IGS): *About the Network.*
https://www.igs.org/about, accessed 10th December 2021,

[9] EUREF Permanent GNSS Network: *Components.*
Royal Observatory of Belgium, 2021,
http://www.epncb.oma.be/_organisation/components.php, accessed 10th December 2021,

[10] Australian Government. *Geoscience Australia: Global Navigation Satellite System Networks.*
https://www.ga.gov.au/scientific-topics/positioning-navigation/geodesy/gnss-networks, accessed 10th December 2021,

[11] Državna geodetska uprava: *CROPOS – Croatian Positioning System.* In Croatian.
https://www.cropos.hr/o-sustavu/cropos-drzavna-mreza-referentnih-stanica-republike-hrvatske, accessed 18th December 2021,

[12] Open data Handbook: *What is open data?* https://opendatahandbook.org/guide/hr/what-is-open-data, accessed 10th December 2021,
[13] Portal otvorenih podataka: *About the open data portal and what open data is.* In Croatian. http://data.gov.hr/o-portalu-otvorenih-podataka-i-sto-su-otvoreni-podaci, accessed 30th November 2021.

MANUSCRIPT PREPARATION GUIDELINES

Manuscript sent should contain these elements in the following order: title, name(s) and surname(s) of author(s), affiliation(s), summary, key words, classification, manuscript text, references. Sections acknowledgments and remarks are optional. If present, position them right before the references.

ABSTRACT Concisely and clearly written, approx. 250 words.

KEY WORDS Not more than 5 key words, as accurate and precise as possible.

CLASSIFICATION Suggest at least one classification using documented schemes, e.g., ACM, APA, JEL.

TEXT Write using UK spelling of English. Preferred file format is Microsoft Word. Provide manuscripts in grey tone. For online version, manuscripts with coloured textual and graphic material are admissible. Consult editors for details.

Use Arial font for titles: 14pt bold capital letters for titles of sections, 12pt bold capitals for titles of subsections and 12pt bold letters for those of sub-subsections. Include 12pt space before these titles.

Include figures and tables in the preferred position in text. Alternatively, put them in different locations, but state where a particular figure or table should be included. Enumerate them separately using Arabic numerals, strictly following the order they are introduced in the text. Reference figures and tables completely, e.g., "as is shown in Figure 1, *y* depends on x …", or in shortened form using parentheses, e.g., "the *y* dependence on *x* shows (Fig. 1) that…", or "… shows (Figs. 1-3) that …".

Enumerate formulas consecutively using Arabic numerals. In text, refer to a formula by noting its number in parentheses, e.g. expression (1). Use regular font to write names of functions, particular symbols and indices (i.e. sin and not *sin*, differential as d not as *d*, imaginary unit as i and not as *i*, base of natural logarithms as e and not as *e*, x_n and not x_n). Use italics for symbols introduced, e.g. $f(x)$. Use brackets and parentheses, e.g. {[( )]}. Use bold letters for vectors and matrices. Put 3pt of space above and below the formulas.

Symbols, abbreviations and other notation that requires explanation should be described in the text, close to the place of first use. Avoid separate lists for that purpose.

Denote footnotes in the text by using Arabic numerals as superscripts. Provide their description in separate section after the concluding section.

References are listed at the end of the article in order of appearance in the text, in formats described below. Data for printed and electronic references is required. Quote references using brackets, e.g. [1], and include multiple references in a single bracket, e.g. [1-3], or [1, 3]. If a part of the reference is used, separate it with semi-colon, e.g. [3; p.4], [3; pp.4-8], [3; p.4, 5; Ch.3]. Mention all authors if there are not more than five of them, starting with surname, and followed with initial(s), as shown below. In other cases mention only the first author and refer to others using et al. If there are two or more authors, separate the last one with the word "and"; for other separations use semicolon. Indicate the titles of all articles, books and other material in italics. Indicate if language is not English. For other data use 11pt font. If both printed version and the Internet source exist, mention them in separate lines. For printed journal articles include journal title, volume, issue (in parentheses), starting and ending page, and year of publication. For other materials include all data enabling one to locate the source. Use the following forms:

[1] Surname, Initial1.Initial2.; Surname, Initial1.Initial2. and Surname, Initial1.Initial2.: *Article title*.
Journal name **Vol**(issue), from-to, year,
http://www.address, accessed date,

[2] Surname, Initial1.Initial2. and Surname, Initial1.Initial2.: *Book title*.
Publisher, city, year,

[3] Surname, Initial1.Initial2.; Surname, Initial1.Initial2., eds.: *Title*.
In: editor(s) listed similarly as authors, ed(s).: *Proceedings title*. Publisher, city, year.

If possible, utilise the template available from the INDECS web page.

CORRESPONDENCE Write the corresponding author's **e**-mail address, **t**elephone and **a**ddress (i.e., η).

ISSN 1334-4684 (printed)
http://indecs.eu